



Corporate Responses to Hate Speech in the 2013 Kenya Presidential Elections

Case Study: Safaricom

Case Study Number 1
NOVEMBER 2013



Institute for
Human Rights and Business

Shaping Policy • Advancing Practice • Strengthening Accountability

Corporate Responses to Hate Speech in the 2013 Kenyan Presidential Elections

Case Study: Safaricom

About this paper

This is the first in a series of case studies in the Digital Dangers Project by the Institute for Human Rights and Business (IHRB).

“Digital Dangers: Identifying and Mitigating Threats in the Digital Realm” is an initiative developed by IHRB in collaboration with the School of Law at the University of Washington in Seattle. Over the next two years, the Digital Dangers Project will develop an online global database of cases from around the world where human rights are infringed in the digital world or in the real world as a result of using digital technology, alongside a series of ICT company case studies.

Case studies in this series provide independent analysis addressing recent events and other emerging issues based on the digital dangers identified. Researchers from IHRB will be embedded in companies that face digital dangers to observe the challenges and steps taken to prevent or mitigate risks to human rights.

Lucy Purdon, IHRB’s ICT Programme Support Manager, prepared this paper with input from Salil Tripathi, IHRB Director of Emerging Issues, and Margaret Wachenfeld, IHRB Director of Legal Affairs.

IHRB wishes to thank Safaricom for their help and co-operation in preparing this study, including facilitating access to Safaricom premises in Nairobi and arranging interviews with staff. Safaricom contributed no financial assistance. The study reflects the views of IHRB and not of any company quoted in the study.

IHRB wishes to thank the Ministry of Foreign Affairs of the Netherlands and the Ministry of Foreign Affairs of Sweden for initial funding for this work.

See www.ihrb.org/digitaldangers for more information.

All rights reserved. The IHRB permits free reproduction of extracts from any of its publications provided due acknowledgement is given and a copy of the publication carrying the extract is sent to its headquarters at the address below. Requests for permission to reproduce or translate the publication should be addressed to the IHRB.

Institute for Human Rights and Business
34b York Way
London N1 9AB
UK

Phone: (+44) 203-411-4333
Email: info@ihrb.org

ISBN: 978-1-908405-16-6

Contents

Executive Summary	2
Key Findings and Recommendations	4
1. Context: Mapping the Current ICT Landscape in Kenya	5
(i) Mobile industry and Safaricom	6
(ii) Internet use	7
2. Background: The 2007/2008 Kenya Presidential Elections	8
(i) The challenge of documenting examples of hate speech	10
3. Dealing with the Aftermath: Enacting New Laws in Kenya	11
4. Safaricom's Actions to Mitigate Hate Speech on its SMS Platform	14
(i) Starting point	14
(ii) Why political bulk SMS?	15
(iii) Bulk SMS originating from outside Kenya	16
(iv) CEO approval	16
(v) Developing internal guidelines to mitigate hate speech sent via bulk SMS	17
(vi) Reviewing existing laws	17
(vii) Guidelines for political mobile advertising on Safaricom's premium rate messaging network	18
5. Development of National Guidelines for the Prevention of Transmission of Undesirable Bulk Political Content via Electronic Communications Networks	19
6. The Elections	21
(i) Confusion over SMS blocking	22
7. The Impact of Social Media on Efforts to Curb Hate Speech During the Kenyan Election	23
(i) Facebook	24
8. Freedom of Expression and its Restrictions in International Law	28
9. The Importance of Context	32
10. Applying a Systematic Approach to Dealing with Hate Speech through the UN Guiding Principles on Business and Human Rights	34
(i) Developing a policy commitment and embedding respect for human rights	36
(ii) Embedding	36
(iii) Identifying risks and potential impacts	36
(iv) Prioritising impacts for action	37
(v) Integrating and Acting	37
(vi) Tracking performance	38
(vii) Communicating	38
(viii) Working with business relationships	39
(ix) Issues of context and remediation	39
(x) Continuing stakeholder engagement	40
Conclusion and Next steps	40

Executive Summary

This study has a three-fold aim. First, to assist ICT companies, which provide SMS services and services based on user-generated content in dealing with situations where accusations of content containing 'hate speech' arise. Second, to explore ways in which civil society groups can work with governments and companies to arrive at common approaches. And third, to reinforce the importance of the government and its role towards protecting rights – of those seeking protection from abuse, and those seeking to express themselves.

The study seeks to provide practical guidance for companies expected to block or remove content as well as act in a way that is consistent with the corporate responsibility to respect human rights, in particular the right to freedom of expression and privacy. It is hoped that government and civil society actors will also find this study useful in applying it to their own work defending freedom of expression and privacy.

To illustrate the challenges facing companies, Kenya was selected as a case study, with the link between the post election violence of 2007/2008 and the spread of hate speech via SMS and blogs as the focus of analysis. This example demonstrates a 'worst case scenario' for companies, where platforms intended to enable communication and the exchange of ideas are used to incite violence taking place 'offline'.

The disputed 2007 Presidential election in Kenya resulted in an outbreak of post-election violence that left over 1,000 people dead and over 600,000 people displaced. The post-election enquiries concluded that hate speech and incitement to violence was widespread on the campaign trail and in the mainstream media, exploiting tensions between ethnic communities (or 'tribes'). These enquiries noted that hate speech was also disseminated via SMS messages on mobile phones and online blogs. Wide dissemination of certain forms of restricted speech, commonly referred to as "hate speech", is seen as an increasingly urgent problem because Information and Communication Technologies (ICTs) make the means to disseminate such speech easily accessible, and provide the ability to circulate it widely and quickly.

The study examines the preparations by government, civil society and business to avoid a repeat of the violence in the 2013 Kenyan Presidential election and provides further context by focusing on the actions of the Kenyan mobile operator Safaricom and later the actions of the social networking company Facebook.

Freedom of expression does not only protect positive and affirming sentiments. It also protects views that may shock, offend, or disturb. However, the question of how to tackle certain forms of speech in the digital realm is an on-going controversy that affects every country in the world. While the Universal Declaration of Human Rights (UDHR) does not place restrictions on freedom of expression, the International Covenant on Civil and Political Rights (ICCPR), which gives legal force to the aspirations of the UDHR for those states that ratify the treaty, does place reasonable restrictions on free expression, including speech that incites violence. But the lack of an internationally agreed definition of "hate speech" has made it difficult to clarify how such acts should be dealt with in the real world, including in the digital realm.

'Hate speech' is not a term used in international law; rather the ICCPR talks of certain restrictions on the right to freedom of expression. But 'hate speech' has become a catch-all term that often encompasses both illegitimate expression that can be banned under international law, and legitimate, if offensive, expression. Mixing these two kinds of expression under the same heading is problematic. It can be problematic for companies when deciding whether content should be removed or blocked. Companies take these decisions based on their own terms of services, or are asked to do so by authorities such as courts or governments, and in some cases, interest groups. A common criticism leveled at companies is that either they do not remove enough content, or block too much. It is not an easy task to draw the line on where freedom of expression ends and legitimate restrictions begin.

In the aftermath of the post-election violence, the Kenyan government adopted a new Constitution¹ and enacted news laws, including the National Cohesion and Integration Act², which set out penalties for spreading hate speech. New guidelines for print and broadcast media were also agreed, placing obligations on them to avoid facilitating hate speech. But the definitions of 'hate speech' therein are thought to be problematic by some civil society groups as they are not aligned with the legitimate restrictions placed on freedom of expression in international law.

In the years since the 2007 elections, a few politicians were arrested for engaging in hate speech and inciting violence. In 2009, political candidates Uhuru Kenyatta and William Ruto, who now serve as President of Kenya and Deputy President respectively, were indicted by the International Criminal Court (ICC) for crimes against humanity for their alleged part in orchestrating the 2007 post-election violence.

Recognising the potentially negative role of SMSs sent by politicians during the last elections, Kenya's largest mobile operator, Safaricom, decided to develop its own code of conduct in vetting bulk SMS content that political parties, politicians and aspirants wished to send in the run up to the 2013 elections. In 2012, Safaricom developed *Guidelines for Political Mobile Advertising on Safaricom's Premium Rate Messaging Service*.³ Under these guidelines, anyone intending to send bulk SMS of a political nature would first have to submit an application to Safaricom, which would vet the content to ensure they did not contain instances of hate speech. In addition to this, Safaricom sought and received the support of the government and the Communications Commission of Kenya (CCK) to develop and release *Guidelines for the Prevention of Transmission of Undesirable Bulk Content/Messages Via Electronic Communications Networks*⁴ in October 2012, which then applied to all mobile network operators in the country.

In the run up to, during and after the 2013 Presidential elections, there was a palpable will on the part of the people, civil society, the government and business to ensure the elections passed peacefully. There were concerns of another outbreak of violence; new, untested laws; fears over the potential of SMS to simultaneously send messages that incite violence; a government that hadn't clarified how it intended to enforce the law; and an expectation from the public - at least civil society - that some action was needed to curb violence.

It seemed Kenyans were determined the elections would pass peacefully and they would not to be used by politicians who had previously exploited ethnic tensions for their own gain. This study is dedicated to those efforts.

1 *The Constitution of Kenya* (2010), The National Council for Law Reporting with the Authority of the Attorney General: www.kenyaembassy.com/pdfs/The%20Constitution%20of%20Kenya.pdf

2 *The National Cohesion and Integration Act*, 2008: www.twawezacommunications.org/docs/cohesion.pdf

3 bloggers.or.ke/safaricom-issues-tough-guidelines-for-political-messaging/

4 *Guidelines for the Prevention of Transmission of Undesirable Bulk Political Content via electronic Communications Networks*, Communications Commission of Kenya: www.cc.k.go.ke/regulations/downloads/Guidelines_for_the_prevention_of_transmission_of_undesirable_bulk_political_content_via_sms.pdf

Key Findings and Recommendations

- The ICT sector is developing at speed and increasingly converging. This convergence results in changes in the way networks are built and run, and may see telecoms companies venturing into providing extra features, such as platforms for user-generated content. Therefore, it is important for telecommunication companies to understand the responsibilities associated with respecting freedom of expression around user-generated content as this may be central to their business model in the near future.
- The emphasis on criminalising many forms of speech not only creates a “chilling effect” on freedom of expression, but also results in the public expectation that many people will be prosecuted. This puts pressure on both the government and companies to act, which may result in legitimate views being restricted and citizens being unduly arrested. It also distracts from efforts to tackle the root cause of such speech.
- The 2013 Kenyan elections were largely peaceful. But this does not mean that hate speech disappeared. Although the language used in campaigns by politicians, political parties and aspirants were scrutinised through various new laws and initiatives, hate speech did not disappear from public rhetoric. Instead, the method of disseminating such messages seems to have shifted. Hate speech appears to have largely left SMS and found a new home on the web, in social media.
- Lack of data from the 2007 elections regarding instances of hate speech led some civil society organisations to take the lead in monitoring social media for hate speech. They were concerned over the amount of content perceived to be hate speech on social media, in particular on Facebook, the most popular social media platform in Kenya by 2013.
- Social media companies worldwide face global pressure to act on the spread of hate speech on their platforms, which raises important questions about the limits of freedom of expression and the legitimate restrictions placed upon it in international law.
- Hate speech often reflects deep-rooted societal tensions and attitudes, a problem that blocking content alone will not solve, which is why a more nuanced approach is needed by companies that explores wider options. Companies could develop a ‘tiered response’ to requests that takes into account the severity and context of the content, based on the work of UN Special Rapporteur on Freedom of Expression, Frank La Rue, who distinguished between three different types of expression that demand different responses and action:⁵
 - Those that are an offence under international law and can be prosecuted criminally;
 - Expression that is not criminally punishable but may justify restriction and a civil suit; and,
 - Expression that does not give rise to criminal or civil sanctions but does raise concerns in terms of tolerance, civility, and respect for others.

The *UN Guiding Principles on Business and Human Rights*⁶ can help companies to apply a systematic approach to tackling hate speech. Following on from IHRB’s work in producing the European Commission *ICT Guide to Implementing the UN Guiding Principles on Business and Human Rights*⁷, the study applies the research and lessons learned in the development of this study to the framework of the Guiding Principles in this specific context.

5 UN General Assembly, *Promotion and Protection of the Right to Freedom of Opinion and Expression. Note by the Secretary General*. 10 August 2011, A/66/90, para 79: <http://www.ohchr.org/Documents/Issues/Opinion/A.66.290.pdf>

6 *The UN Guiding Principles on Business and Human Rights* www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

7 Written by IHRB and Shift, June 2013: www.ihrb.org/publications/reports/human-rights-guides.html

Methodology

This case-study report focuses on the experiences of Kenyan mobile operator Safaricom in the days leading up to the country's elections in early 2013. Between February 25 and March 8 2013, IHRB researcher Lucy Purdon spent seven days at Safaricom headquarters, to observe at close quarters the kind of issues the company faced, the way it took decisions, the framework it applied to these decisions, and the impact of those decisions – both on preserving peace and respecting rights. Safaricom co-operated with IHRB's request for visiting the company headquarters⁸ and made its staff available for interviews. IHRB agreed to share its draft report with Safaricom before its publication in order to identify any proprietary information or information of commercial significance the company did not want made public.

This report also includes desk based research and additional interviews with experts in Kenya and abroad, including content service providers, government and civil society actors. During a follow-up visit to Nairobi in July 2013, IHRB presented initial findings at the Kenya Internet Governance Forum (IGF) on hate speech and social media. Quotes from a panel discussion at this event are included in the study. IHRB also benefited from discussions with civil society organisations based in Nairobi. IHRB would like to thank all those who contributed to this study.

1. Context: Mapping the current ICT landscape in Kenya

The Kenyan government is clear in its ambitions for the country's ICT development. *The National ICT Masterplan* (2012), developed by the Ministry of Information and Communication, sets out the vision for Kenya to be Africa's ICT 'hub' by 2017⁹ and plans to build a huge tech-city on the outskirts of the capital Nairobi by 2030.¹⁰ Kenya is home to a dedicated blogging community and technologists, including the founders of Ushahidi¹¹ (which means "witness" or "testimony" in Swahili) an open source crowd-sourcing tool¹² that has gained international acclaim. Several tech 'hubs' have sprung up in and around Nairobi offering workspace and support to tech entrepreneurs. The government funded Kenya ICT Board also provides seed funding for web entrepreneurs with projects ranging from documenting museum artifacts to initiatives to increase connectivity, especially in rural areas.

8 This study was written with the co-operation of Safaricom, but with no financial assistance from the company. IHRB would like to thank Safaricom for their co-operation. The study reflects the views of IHRB and not necessarily those of Safaricom.

9 Kenya ICT Board, *National ICT Masterplan 2012-2017, Inclusion. Innovation. Beyond Broadband*: www.ict.go.ke/index.php/ict-masterplan/420-overview

10 BBC News, 23rd January 2013, *Kenya begins construction of 'silicon' city* Konza: www.bbc.co.uk/news/world-africa-21158928

11 Ushahidi began a website used to 'map' incidents of violence in the 2007 elections, enabling people to see where violence was happening and which areas were safe. People texted (using an SMS short code) or emailed in reports of the incident, geographic location which were then transferred to the online map and marking the report as verified in accordance with the verification guidelines or as unverified until the reports could be investigated further. It has since been used to monitor elections and disasters worldwide, such as the 2009 Indian elections, track violence in Gaza and gather reports globally regarding the swine flu outbreak.

12 Crowdsourcing refers to soliciting information from a large number of people, particularly the online community. Open source software means that anyone can employ their crowdsourcing tool and adapt for other projects.

(i) Mobile industry and Safaricom

The ICT sector in Kenya is dominated by mobile telephony. Out of a population of around 43 million, there are to date around 29.8 million mobile subscriptions in Kenya, compared with 12 million at the time of the 2007 Presidential elections. By 2016, business analysts predict there will be 39.5 million mobile subscriptions in Kenya.¹³

Kenya's mobile network is served by four operators: Safaricom, Airtel, Orange, and Yu. Safaricom is the market leader, holding 65.1% of the market share of mobile subscriptions.¹⁴ Safaricom was established in 2000 and is a joint venture between the state-run Telkom Kenya Ltd, which held a 60% stake, and Vodafone Kenya Ltd, a locally listed subsidiary of Vodafone plc., which has a 40% stake. In 2008, Safaricom conducted a successful Initial Public Offer where the public acquired 25% ownership leaving the Government stake at 35%, with Vodafone maintaining its 40% shareholding.

Pre-paid or 'pay as you go' subscriptions make up 98.7% of all mobile subscriptions¹⁵ and SIM (Subscriber Identity Module) cards cost 100 Kenyan shillings (around 0.77 GBP). To send an SMS costs around 1 Kenyan shilling. The communications regulatory authority, the Communications Commission of Kenya (CCK) reported that in the period January-March 2013, 4 billion SMSs were sent, up from 3.6 billion in the previous quarter.¹⁶ Safaricom dwarfs its rivals in SMS traffic, commanding 94.8% of the market in the period January-March 2013.¹⁷

Safaricom operates almost exclusively in Kenya and the company's CEO is able to make decisions independently. In comparison, Airtel is a multinational company operating in 17 African countries and major decisions are escalated to the company headquarters in India. There is a sense then, that Kenyans consider Safaricom as more of a 'local' company. In addition, the development of money transfer via mobile phones, pioneered by Vodafone with their branded system called M-PESA and first launched by Safaricom in 2007, has had an extremely positive effect on Kenyans' ability to transact, pay, and earn, transforming the ways of doing business and enhancing financial inclusion, particularly for the large number of Kenyans who do not have a bank account. As of May 2012, there were 19.3 million mobile money transfer subscriptions in Kenya and 14.91 million of those use Safaricom's M-PESA service.¹⁸ Users credit their account with cash at one of the thousands of local agent outlets (for a small fee). Users can then use their mobiles to pay money to other people, utility bills or buy commodities from shops. A regular sight in shops is a sign with a 5-digit number that identifies that business and is used to target the transfer of money. The receiver can then withdraw the cash from one of the local agent outlets or an ATM.

13 Business Monitor International, *Kenya Telecommunications Report*, March 2012: www.marketresearch.com/Business-Monitor-International-v304/Kenya-Telecommunications-Q2-6844273/

14 At March 2013. Communications Commission of Kenya, *Sector Statistics Report, January-March 2013*, p10: www.cck.go.ke/resc/downloads/Sector_Statistics_Report_for_3rd_Quarter_2012-2013.pdf

15 Ibid, p6

16 Ibid

17 Ibid, p15

18 Safaricom 2012 Annual Report, p31: www.safaricom.co.ke/safaricom_annual_report/pdfs/Safaricom_Annual_Report.pdf

(ii) Internet use

In 2007, at the time of the last elections, it was estimated there were three million Internet users in Kenya.¹⁹ In 2013, the figure rose to 16.4 million users.²⁰ High connection costs for the user and the expense of buying a computer or laptop are still a barrier to fixed-line Internet access and the country is still affected by regular and sometimes lengthy power cuts due to an erratic electricity supply, which is another barrier to fixed line Internet access. It is perhaps no surprise then that mobile data (i.e. Internet accessed via mobile phones) makes up 99% of Kenya's Internet subscriptions and this is set to increase rapidly, in line with global trends.²¹

In Safaricom's 2012 Annual Report, the company said that one of its strategic goals is to reduce dependence on revenue from voice calls and increase revenue from data. Safaricom has been investing in and improving its 3G²² infrastructure and negotiating with mobile phone manufacturers like Samsung and Nokia to roll out low cost smart-phones to enable more Kenyans to access the Internet via mobile phones. The company has also reduced tariffs for mobile Internet browsing, resulting in mobile data revenue increasing by 23% in 2012.²³ Safaricom has launched its own app store, with a focus on developing apps around issues particular to Kenya, such as apps to assist farmers, as well as users being able to download social media apps such as Facebook, Twitter and Skype. Social networking is not yet as widespread as in other parts of the world²⁴, but is thought to be driving the demand for data services accessed via mobile phones due to its popularity with young people.²⁵

In response to this, platforms like Facebook now recognise that their services may be encouraging people in developing countries to subscribe to mobile data packages and are working on projects to encourage this. Facebook Zero is one such project, where Facebook works with a mobile operator to offer free access to Facebook on a mobile for a limited period of time. Richard Allan, Facebook's Director of Policy for Europe, Middle East and Africa, explained at the 2013 Kenya Internet Governance Forum (IGF):

"From the Facebook point of view, we get a new user, from the Telco point of view, they get people using the Internet who will hopefully take up a subscription after the free period has ended and from a national broadband strategy point of view, someone is now connected to mobile broadband who might not have been connected otherwise."²⁶

This development in the ICT landscape shows the convergence of the sector, with mobile operators expanding into being Internet service providers and social media platforms like Facebook becoming more involved with mobile operators. Traditionally, mobile operators and web-based services have operated quite independently with different operating models and regulation. This convergence means that companies traditionally not involved with handling content may soon confront some of the challenges currently faced by web-based companies. Therefore, it is important to consider these challenges at the start of operations so that companies are prepared for situations that may arise quickly.

19 Internet user historical data graphs: www.indexmundi.com/g/g.aspx?c=ke&v=118

20 Communications Commission of Kenya, *Sector Statistics Report*, January-March 2013, p22: www.cck.go.ke/resc/downloads/Sector_Statistics_Report_for_3rd_Quarter_2012-2013.pdf

21 GSMA predict that data revenues will overtake voice revenues worldwide by 2018. GSMA, *The Rise of Connected Devices Will Drive Mobile Operator Data Revenues Past Voice Revenues Globally by 2018*, February 25th 2013: www.gsma.com/newsroom/gsma-the-rise-of-connected-devices-will-drive-mobile-operator-data-revenues-past-voice-revenues-globally-by-2018

22 3G stand refers to the 3rd generation of mobile technology that allows faster connections and mobile Internet access.

23 Safaricom Annual Report 2012, p3: www.safaricom.co.ke/images/Downloads/Annual_Reports/Annual_Report_2012_Final.pdf

24 As of December 2012, there were just over two million Facebook users and an estimated 300,000 Twitter users: digital.theedition.co.ke/i/121728/21 and www.internet-worldstats.com/africa.htm#ke

25 CCK, *Sector Statistics Report*, January-March 2013, p22: www.cck.go.ke/resc/downloads/Sector_Statistics_Report_for_3rd_Quarter_2012-2013.pdf

26 Comments made during panel on hate speech and social media, chaired by IHRB, Nairobi 26th July 2013: bit.ly/kigf13webcast

2. Background: The 2007/2008 Kenya Presidential Elections

On December 27 2007, Presidential elections took place in Kenya, involving two main candidates - Mwai Kibaki of the Party of National Unity (PNU) and Raila Odinga of the Orange Democratic Movement (ODM). Odinga was leading in the polls by over one million votes, but in the final hours of voting, Kibaki took the lead and was declared winner by a very narrow margin.²⁷ Amid accusations by Odinga of vote rigging, Kibaki was hurriedly sworn in as President, in private, three days later. Riots erupted in the Odinga stronghold of Kibera in Nairobi - Kenya's largest slum and Odinga's parliamentary constituency for 20 years - and then spread to the rest of the country with increasing violence. Atrocities committed against ethnic groups included murder, torture, rape and forced eviction.²⁸

On February 28, 2008, following failed mediation attempts, former UN Secretary General, Kofi Annan, brokered a power sharing agreement between Kibaki and Odinga to bring peace in which Kibaki remained President and the position of Prime Minister was created for Odinga. Uhuru Kenyatta, son of Kenya's first President Jomo Kenyatta, who had withdrawn from the election as the Kenya African National Union (KANU) presidential candidate to back Kibaki, was given one of the Deputy Prime Minister positions. The post-election violence left over 1,000 people dead and over 600,000 people displaced.²⁹

According to subsequent enquiries into the post-election violence³⁰, hate speech was rife during the electioneering period, from the promotion of negative ethnic stereotypes to outright incitement to violence between ethnic communities (commonly referred to as "tribes" in Kenya). The enquiries also acknowledged the role played by SMS and blogs. The Kenya National Commission on Human Rights report, *On the Brink of a Precipice*, summarised:

"The entire electioneering period was characterised by hate speech and incitement to violence³¹... One thing that is memorable about the 2007 elections is the role played by all manner of information from all manner of sources, unsolicited and solicited- SMS, blogs, emails, newsletter, leaflets, not to mention the newspapers, TV and radio sources."³²

The following are existing examples that are reportedly characteristic of typical SMS messages circulated during the 2007/2008 elections:

"Fellow Kenyans, the Kikuyus [Kenyan tribe] have stolen our children's future. Hope of removing them through the ballot has been stolen. We must deal with them the way they understand, violence. We must dominate them."

27 See Goldstein and Rotich, Digitally Networked Technology in Kenya's 2007-2008 Post Election Crisis, Berkman Centre for Internet and Society at Harvard University, 2008, p4: cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Goldstein&Rotich_Digitally_Networked_Technology_Kenyas_Crisis.pdf

28 See Human Rights Watch (2008) *From Ballots to Bullets*: www.hrw.org/reports/2008/03/16/ballots-bullets-0.

Amnesty International (2008) *Kenya's Victims of Political Violence Speak Out*:

www.amnesty.org/en/news-and-updates/feature-stories/kenyan-victims-political-violence-speak-out-20080226 .

Waki Commission (2008) *Waki Commission of Enquiry into Post-Election Violence Final Report*:

wikileaks.org/wiki/Kenya:_Waki_Commission_of_Inquiry_into_Post-Election_Violence_final_report,_15_Oct_2008.

Kriegler Commission (2008) *Kriegler Report on the 2007 Kenyan Elections*: kenyastockholm.files.wordpress.com/2008/09/the_kriegler_report.pdf

29 *Waki Commission* (2008) Ibid.

30 The Commission of Inquiry on Post Election Violence (known as the Waki Commission after the Chairman of the Commission Justice Philip Waki) investigated the reasons for the violence and released its findings in a report known as the Waki report. The Independent Review Commission (commonly known as the Kriegler Commission after the Chairman Judge Johann Kriegler) investigated the Kenyan electoral system.

The Kenya National Commission for Human Rights (KNCHR) released their report, *On the Brink of a Precipice*, based on investigations and victim's testimonies. See Waki Commission (2008) Ibid. Kriegler Commission (2008) Ibid. Kenyan National Commission on Human Rights (2008) *On the Brink of the Precipice: A Human Rights Account of Kenya's Post-2007 Election Violence*:

www.knchr.org/Portals/0/Reports/KNCHR_REPORT_ON_THE_BRINK_OF_THE_PRECIPE.pdf

31 Ibid, Para 636, p131

32 Kenyan National Commission on Human Rights (2008) Ibid. Para 71, p29. See also Kriegler Commission (2008) Ibid. p23.

“We say no more innocent Kikuyu blood will be shed. We will slaughter them right here in the capital city. For justice, compile a list of Luos and Kalus(ph) [ethnic communities] you know at work or in your estates, or elsewhere in Nairobi, plus where and how their children go to school. We will give you numbers to text this information.”³³

Government actions to combat the spread of hate speech during the post election violence included considering shutting down the SMS network (which they ultimately did not do) but on December 30 2007, the government ordered a blackout of mainstream media (television, radio and newspapers). The Kriegler report argues that this action fueled the violence:

“Indeed the information blackout, engrafted onto the ECK’s [Electoral Commission of Kenya] lamentable failure to keep the people of Kenya informed, could well be seen as a direct link in the chain of causation that led to the tragedy.”³⁴

By the 2007/2008 elections, a vibrant blogging community had emerged in Kenya and some bloggers and technologists were extremely active in filling the void left by the government imposed media blackout. Bloggers provided information on polling results, dispelled rumours and warned of violent ‘hotspots’; out of this Ushahidi was born. However, bloggers often struggled to contain the comments that incited violence or hatred, which other users left on their blogs in the “Comments” section. Orly Okolloh, who writes as ‘Kenyan Pundit’ and is a founder member of Ushahidi, reported spending a lot of time moderating comments and repeatedly pleaded with users not to use the comment section on her blog for “tribal based vitriol”.³⁵

The popular online bulletin board, Mashada, administrated by David Kobia, another founder of Ushahidi, became overwhelmed with “divisive and hostile” messages and took the decision to shut down the site temporarily in January 2008 due to the difficulties of moderating so much content.³⁶ He redirected Mashada users to a new site called “I Have No Tribe” which explicitly promoted constructive speech and attracted comments from Kenyans around the world on the theme of “I Have No Tribe...I am a Kenyan.” Mashada was then re-opened on February 14th 2008. According to Ethan Zuckerman, a leading American Internet activist, Mashada had “elegantly demonstrated that one possible response to destructive speech online is to encourage constructive speech.”³⁷

In the years since the 2007 elections, several politicians had been arrested for disseminating hate speech under the National Cohesion and Integration Act. This included two MPs and an assistant minister arrested during the constitutional reform in 2008 for engaging in hate speech³⁸; a minister arrested in 2012 for inciting violence that killed 100 people in the Tana River area³⁹ and a deputy minister arrested in 2012 for alleged incitement and hate speech against ethnic Maasai.⁴⁰

In July 2009, information was passed to the International Criminal Court (ICC) prosecutor, who opened an investigation into crimes against humanity committed during the post election violence of 2007/2008. Among the charges confirmed were those against Uhuru Kenyatta (then the Deputy Prime Minister), William Ruto (Senior member of

33 Ofeibea Quist-Arcton, *Text Messages Used to Incite Violence in Kenya*, NPR: www.npr.org/templates/story/story.php?storyId=19188853

34 Kriegler Commission (2008) *Ibid*, p64

35 See www.kenyanpundit.com/2007/12/29/election-2007-dec-29-1030-am-update/ and www.kenyanpundit.com/2007/12/29/till-tomorrow/

36 See Goldstein and Rotich, *Digitally Networked Technology in Kenya’s 2007-2008 Post Election Crisis*, Berkman Centre for Internet and Society at Harvard University, 2008, p5: cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Goldstein&Rotich_Digitally_Networked_Technology_Kenyas_Crisis.pdf

37 Ethan Zuckerman, *Kenya: Citizen Media in a Time of Crisis*, 20th June 2008: www.ethanzuckerman.com/blog/2008/06/20/kenya-citizen-media-in-a-time-of-crisis/

38 Humphrey Malalo and Duncan Miriri, Reuters, *Kenyan Politicians Arrested Over Hate Speech Charges*, June 15th 2010: www.reuters.com/article/2010/06/15/us-kenya-politics-idUSTRE65E20K20100615

39 BBC News online, 12th September 2012, *Kenya’s Tana River clashes: MP Charged With Incitement*: www.bbc.co.uk/news/world-africa-19575973

40 BBC News Online, 25th September, *Kenya MP Ferdinand Waititu Accused of ‘Hate Speech’*: www.bbc.co.uk/news/world-africa-19713466

ODM) and Joseph Arap Sang (head of operations for radio station Kass FM in Nairobi). In 2012, Uhuru Kenyatta and William Ruto formed the Jubilee coalition and went on to win the 2013 Presidential election. Uhuru Kenyatta currently serves as the fourth President of Kenya with William Ruto as Deputy President.

(i) The challenge of documenting examples of hate speech

It is difficult to give accurate statistics on the content and number of SMSs inciting violence that were circulated during the 2007/2008 elections in Kenya because there is very little existing data from the period in the public domain. One reason is that after the post-election violence, people were cautious about giving information regarding inciting messages they may have received or saved, for fear of implicating themselves in incitement to violence. Even some of those working in human rights organisations deleted such SMSs from their mobile phones because the content was considered so toxic. It is not known how widely these messages were disseminated. While it has been recorded in enquiries into the post election violence that SMS did play a part, there are only a handful of existing examples of the actual content of the messages. Therefore, much of what is known is based on anecdotal or second-hand evidence, such as some bloggers recording the content of messages.

Studying hate speech presents a particular dilemma, in that republishing examples perpetuates the sentiments in the message, and it is now illegal for newspapers in Kenya to reprint alleged hate speech. This can limit the scope of any study of hate speech in that it can frustrate efforts to understand the context and establish the nature of that speech, which as will be discussed later is extremely important. The UN Special Rapporteur on Freedom of Opinion and Expression, Frank La Rue, acknowledges the problem, saying in a recent report:

“A clear distinction between expression that constitutes incitement, hate speech and merely offensive speech should be at the core of any such data-collection and analysis exercise ... In most countries, comprehensive data on such areas are glaringly absent. Consequently, policies and legislation are frequently based on perception.”⁴¹

In addition, some telecommunication companies, as is the case with Safaricom, maintain they do not have the legal authority or technical capacity to even store SMS messages for retrieval and analysis. Companies may also be precluded from collecting evidence of hate speech due to the risk of violating user privacy. It is therefore crucial that civil society organisations take on a role of documenting evidence in order to aid further investigation and analysis and that companies respond to reports that patterns of speech are emerging that are typical of those before mass violence or genocide.

Lack of data from the 2007 elections regarding the spread of hate speech via electronic media led civil society organisations to take the lead in monitoring the online space during the 2013 elections. For example, the Umati project (‘Umati’ means ‘crowd’ in Swahili) set up by iHub Research and Ushahidi, aimed to document instances of hate speech and also ‘dangerous’ speech, as outlined in the methodology by Susan Benesch of the Dangerous Speech Project, in the Kenyan online space.⁴² The Umati project recognised a need to identify the type of speech that has the “potential to trigger violence, so that the violence can be avoided or diminished”.⁴³ The aim was to operate as an early warning mechanism and test prevention strategies rather than advocate for prosecution.

41 UN General Assembly, *Promotion and Protection of the Right to Freedom of Opinion and Expression: Note by the Secretary-General*, 7th September 2012, A/67/357, P19 para 68: daccess-dds-ny.un.org/doc/UNDOC/GEN/N12/501/25/PDF/N1250125.pdf?OpenElement

42 The Dangerous Speech Project’s goal is to find new ways to prevent collective violence, without curbing freedom of expression. The project developed a set of guidelines to try to predict the likelihood of certain speech catalyzing violence in certain situations. “The guidelines are based on the insight that the dangerousness of a particular speech act, in the context in which it is made or disseminated, depends on five variables: the speaker, the audience, the speech itself, the historical and social context, and the means of dissemination. For example, some speakers are more influential than others, and some audiences are especially vulnerable”: voicesthatpoison.org

43 *Umati: Monitoring Online Dangerous Speech*. February and March 2013 report, p2: www.research.ihub.co.ke/uploads/2013/april/1365508815_819_823.pdf

However, in the run up to and during the 2013 Presidential elections, there was a palpable will on the part of the government, business, civil society and the *wananchi* (citizens) to avoid repetition of events at the 2007 polls and restore Kenya's reputation as a peaceful and progressive member of the African community. There was a concerted effort by the government and mainstream media to advocate for peace, with advertisements placed regularly on TV, radio, and billboards. Presidential candidates appeared on stage at Uhuru Park in the centre of Nairobi at a peace rally the week before the elections, pledging their commitment to a peaceful election. They agreed to concede defeat gracefully, and promised that if there were any dispute over the results, they would settle it in court and not on the streets.

3. Dealing with the aftermath: Enacting new laws in Kenya

In the wake of the post election violence and subsequent enquiries, the Kenyan government initiated several important legislative, institutional, and policy reforms⁴⁴, including passing a new Kenyan Constitution, as there was growing realisation that a culture of impunity prevailed during the elections and the country was not equipped to deal with unrest on such a scale. Among the measures, the Electoral Commission of Kenya (ECK) was found to have failed its mandate and was disbanded. It was replaced by the Independent Electoral Boundaries Commission (IEBC) and the government enacted changes to the elections process in the Elections Act and Political Parties Act. The term 'hate speech' then made its way into the 2010 Kenyan Constitution, which presents a revised definition of the right to freedom of expression and its limitations:

33. (1) *Every person has the right to freedom of expression, which includes—*

- (a) freedom to seek, receive or impart information or ideas;*
- (b) freedom of artistic creativity; and*
- (c) academic freedom and freedom of scientific research.*

33. (2) *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).*

33. (3) *In the exercise of the right to freedom of expression, every person shall respect the rights and reputation of others.*

In 2008, the government established the National Cohesion and Integration Act (NCIA), which outlawed discrimination on ethnic grounds in an attempt to reconcile deep-rooted tribal differences in Kenya. The NCIA created the National Cohesion and Integration Committee (NCIC) which was mandated to "facilitate and promote equality of opportunity, good relations, harmony and peaceful coexistence between persons of different ethnic and racial backgrounds in Kenya and to advise the government thereof."⁴⁵

44 Milly Lwanga, Article 19, Freedom of Expression and Harmful Speech: The Kenyan Situation, 27th September 2012: www.article19.org/join-the-debate.php/82/view/#_ftnref1

45 www.cohesion.or.ke

It was the NCIC which would eventually be the channel by which hate speech was reported. The NCIA has a section on 'hate speech', defining it thus:

Hate speech.

(1) A person who-

[a] uses threatening, abusive or insulting words or behaviour, or displays any written material;

[b] publishes or distributes written material;

[c] presents or directs the performance the public performance of a play;

[d] distributes, shows or plays, a recording of visual images; or

[e] provides, produces or directs a programme; which is threatening, abusive or insulting or involves the use of threatening, abusive or insulting words or behaviour commits an offence, if such person intends thereby to stir up ethnic hatred, or having regard to all the circumstances, ethnic hatred is likely to be stirred up. Ethnic hatred means hatred against a group of persons defined by reference to colour, race, nationality (including citizenship), or ethnic or national origins.⁴⁶

Section 62 sets out the criminality of certain forms of speech.

62. *(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 (3) commits an offence and shall be liable on conviction to a fine not exceeding one million shillings.

The Kenya Information and Communication Act appears to specifically refer to SMS and could be extended to online communications, but does not appear to refer to hate speech as such,

A person who by means of a licensed telecommunication system—

(a) sends a message or other matter that is grossly offensive or of an indecent, obscene or menacing character;
or

(b) sends a message that he knows to be false for the purpose of causing annoyance, inconvenience or needless anxiety to another person commits an offence and shall be liable on conviction to a fine not exceeding fifty thousand shillings, or to imprisonment for a term not exceeding three months, or to both.⁴⁷

The term 'hate speech' does not have a workable definition in international human rights law. Further analysis features in greater detail later in this study. At this point, however, it should be noted that the International Covenant

⁴⁶ Section 13, National Cohesion and Integration Act 2008: www.cohesion.or.ke

⁴⁷ *Kenya Information and Communication Act 1998 (as amended in 2008)*, paragraph 29:
www.cck.go.ke/regulations/downloads/Kenya-Information-Communications-Act-Final.pdf

on Civil and Political Rights (ICCPR) outlines legitimate restrictions on freedom of expression, in order to protect the rights of others. Article 19, paragraph 3, of the ICCPR lays the groundwork:

The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:

- (a) for respect of the rights or reputation of others*
- (b) for the protection of national order (ordre public), or of public health or morals.*

Article 20 of the ICCPR examines these restrictions in more depth:

- 1. Any propaganda for war shall be prohibited by law*
- 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.*

In other words, according to Article 20 (2), hatred, by itself, would not be subject to restriction. It is only when advocacy of national, racial or religious hatred constitutes incitement to discrimination, hostility or violence that it is considered illegal under international law. Hate speech is defined only in ethnic terms in Kenyan law and it falls short of Article 20 (2) of the ICCPR, as it does not reflect that both advocacy of hatred and incitement to discrimination, hostility and violence are the criteria for illegal speech.⁴⁸

Although the Kenyan government criminalised discriminatory speech as well as speech inciting violence, which is in accordance with international law, each law in the Kenya context presents restrictions on speech slightly differently. There has been criticism about the additional use of undefined terms in these laws such as “abusive words”, “incite feelings of contempt” “vilification of others”, “stir up ethnic hatred” and “causing annoyance or inconvenience” that could be misinterpreted and lead to undue restrictions on freedom of expression. Stephanie Muchai, Legal Officer at ARTICLE 19 in Nairobi, stated at the Kenya IGF:

“Because we have a problematic constitutional definition, then we have a problematic legal definition coming out of that... when you are trying to regulate at this lower level, it [speech] may very well be dangerous, it may very well be other things, but then you are putting limitations on speech that is not justifiable, you are limiting debate because freedom of expression does not just protect nice or agreeable statements, it protects things that are shocking, offensive and disturbing because that is what freedom of expression is about”.⁴⁹

Stephanie Muchai also said that this inconsistency was already having a chilling effect on freedom of expression: “We spoke to the Bloggers Association of Kenya and they said there was such a ‘chilling effect’ online because people don’t know what to say and what not to say as they don’t know what hate speech is, there is no categorisation.”⁵⁰

The emphasis on criminalising many forms of speech not only creates a “chilling effect” on freedom of expression, but also results in the public expectation that many people will be prosecuted. This puts pressure on both the government and companies to act, which may result in legitimate views being restricted and citizens being unduly arrested. It also distracts from efforts to tackle the root cause of such speech.

48 Article 19, June 2010. *Commentary on the Regulation of Hate Speech in Kenya*, para 55, p17:

www.article19.org/data/files/pdfs/analysis/kenya-commentary-on-the-regulation-of-hate-speech-.pdf

49 Comments made during panel on hate speech and social media, chaired by IHRB, Nairobi 26th July 2013: bit.ly/kigf13webcast

50 Ibid

4. Safaricom's actions to mitigate hate speech on its SMS platform

By 2012, Kenya was making significant investments to boost its ICT infrastructure. Benefits to the broader economy and to rural areas and to the poor were now visible. Elections were approaching, and there was widespread concern that hate speech, which had contributed to the violence in the last elections, may be disseminated more effectively, and efficiently, through greater use of mobile technology. In the run up to the 2013 elections, the challenge for companies like Safaricom was to operate in an environment where concerns of another outbreak of violence were common. This, on top of new, untested laws; fears over the potential of SMS to simultaneously send messages that incite violence; a government that hadn't clarified how it intended to enforce the law; and an expectation from the public - at least civil society—that some action was needed to curb violence.

In the period from 2007-2013, mobile subscriptions in Kenya had increased from 12 million to 29.8 million. Safaricom projected that 2 billion SMSs would be sent in the run up to the election, in comparison with 1 billion sent during the same period in the 2007 election.⁵¹ As the market leader in terms of SMS traffic, the company perceived an urgent need to address the risk of hate speech spreading via SMS sent through its network.

(i) Starting Point

Safaricom began internal discussions in early 2012 with representatives from the Regulatory & Public Policy, Legal, Communications/Public Relations, Technical, Risk Management and Commercial teams (in particular the Commercial Business Unit, which dealt with content service providers). They met to brainstorm ideas in a combination of scheduled and ad-hoc meetings. They decided to take a pre-emptive step – shaping a system that would give them some control over the content of bulk SMS messages sent by political parties, politicians and aspirants.

Bulk SMS are different to sending mass SMS or 'one to many', which involves the user sending SMS messages to their contacts via their own mobile phone. The focus on bulk SMS was critical for Safaricom, as premium rate SMS services are a large revenue earner, with services like games and ringtones. The company also sells the right to send 'bulk SMS' (often consisting of millions of SMS) to other companies. Businesses, including banks, political parties and the media use a bulk SMS facility regularly to send SMSs to a large number of people who, in theory, have subscribed to a particular service.⁵² Bulk messages are free for the user to receive.

Safaricom does not sell bulk SMS bundles directly to the person or entity wanting to send messages and Safaricom does not transmit the messages itself either. Content service providers (CSP) develop tools and software to enable bulk SMS to be sent, using the network belonging to a particular mobile network operator. A client, such as a politician, would approach a CSP to send a certain number of messages on a certain day and submit the content to be sent in advance. The CSP will have a license agreement with a mobile network operator, such as Safaricom, to send messages to customers on that particular mobile network. CSPs often have such a license agreement with a number of mobile network operators which enables bulk SMS to be sent to a large number of people on different mobile networks.

⁵¹ Edwin Mutai, Business Daily Africa, *Two Billion SMS to Be Sent in the Run-up to March 4 Election*, February 7th 2013: www.businessdailyafrica.com/Corporate-News/Two-billion-SMSs-to-be-sent-in-run-up-to-March-4-election-/539550/1688064/-/12vhp2mz/-/index.html

⁵² For example, this could be news updates from media houses, health information from the government, or messages from banks or political parties.

(ii) Why political bulk SMS?

As mentioned previously, the post election violence enquiries concluded that politicians who engaged in hate speech at rallies and in campaigns drew large crowds. Devolution is a major change in Kenya's political landscape. The new Constitution had changed the election process and created 47 counties in an effort to devolve government. This meant there were thousands of candidates campaigning for newly created positions in each county. Each voter would vote for six positions: President, County Representative, Governor, Senator, Member of Parliament and Women's Representative. The all-important Presidential race had eight candidates.

Safaricom felt that political parties, politicians, and aspirants would use bulk SMS to communicate with voters and feared these bulk SMSs of a political nature could contain hate speech. Although bulk SMS was a big revenue earner, in an interview with IHRB, a Safaricom representative said the company did not want to sell the service "at any cost". Safaricom discussed internally the limitations of what it could do. The main risk with bulk SMS is that it can be forwarded many times as a peer-to-peer message, at which point it becomes a private communication, and Safaricom cannot intervene due to the implications for privacy and in accordance with the 2009 Kenyan Information and Communications Act.⁵³ In the Act, mobile operators are prohibited by law and license from disclosing content of SMS messages:

"A licensed telecommunications operator who otherwise in the course of his business-

- (a) intercepts a message sent through a licensed telecommunication system; or
- (b) discloses to any person the contents of a message intercepted under paragraph (a); or
- (c) discloses to any person the contents of any statement or account specifying the telecommunication services provided by means of that statement or account, commits an offence and shall be liable on conviction to a fine not exceeding three hundred shillings or, to imprisonment for a term not exceeding three years, or to both."

Safaricom wanted to ensure that the bulk SMS sent through the Safaricom network – by politicians, aspirants and political parties only – would meet a certain threshold, and the content was of a certain standard, in accordance with Kenyan law, without infringing on user privacy. By focusing their efforts on bulk SMS of a political nature only, Safaricom could work within its legal boundaries of leverage.

The Electoral Code of Conduct, part of the 2011 Elections Act⁵⁴ also specifically prohibited hate speech in campaigns, which here specifically mentions speech that may lead to violence:

"In addition to the foregoing obligations incidental to the very act of subscribing to the Code, the Code further requires of those bound by it to do the following:

- (a) Publicly and repeatedly condemn violence and intimidation and avoid the use of hate speech, language or any kind of action which may lead to violence or intimidation whether to demonstrate party strength, gain any kind of advantage or for any other reason;
- (b) Refrain from any action involving violence or intimidation"

53 Kenya Information and Communication Act 1998 (as amended in 2008), paragraph 31:
www.cck.go.ke/regulations/downloads/Kenya-Information-Communications-Act-Final.pdf

54 The Elections Act (2011) including the Electoral Code of Conduct, Second Schedule, p94:
www.cickenya.org/index.php/legislation/acts/item/65-the-elections-act-2011#.Umqe8hZU2ZY

(iii) Bulk SMS originating from outside Kenya

During the 2007 elections, Safaricom identified that bulk SMS bundles were being bought from outside Kenya and used to spread hate speech. Some bulk SMS also used modified alphanumeric codes⁵⁵ that falsely claimed to be a particular person or entity. Safaricom had identified that political parties and aspirants were buying bulk SMS bundles from overseas operators, which could be sent to mobile phones over the Internet, which were being received in Kenya without an identity number or Caller ID, which would determine the origin of the message and allow it to be traced. This method would also bypass any attempts to vet content in Kenya.

Safaricom put in place a 'one-to-many' filtering tool, which can block an SMS if over 100 texts containing the same message were being sent from the same number simultaneously via Safaricom's network. The intention was to prevent the spread of hate speech. But it also prevented organisations, such as unofficial election observers, to communicate and co-ordinate their efforts. Organisations such as Ushahidi, which rely on SMS for crowdsourcing, are also concerned of the effect this policy may have on collecting data in the future.

In March 2012, Safaricom announced that it would be blocking SMS with a foreign alphanumeric code on its network.⁵⁶ This would prevent alphanumeric codes being used to bypass attempts to vet bulk SMS content for hate speech.⁵⁷ A member of Safaricom's senior management suggested that,

"By blocking non-licensed bulk providers, we will also be looking out for our customers in the sense that if measures are not taken, this problem will could [sic] be a big issue especially now that we are moving towards the elections."⁵⁸

Safaricom would only accept bulk SMS messages from CSPs inside Kenya with which it had a licensing agreement. While the measure was primarily taken to reassure users, CSPs criticised Safaricom for announcing this measure with immediate effect and giving no notice. Although this may raise concerns regarding the impact on competition, this discussion is outside the scope of this study.

(iv) CEO approval

The first step in initiating any process was to pitch the idea of action to address the risk of hate speech spreading via SMS sent through its network to the CEO of Safaricom, Bob Collymore. The project would require a huge investment of time (not so much financial investment), and Collymore's commitment was important, as well as establishing who would be responsible for implementing any plans or processes. Once the CEO had given the go-ahead, the senior Public Policy team led the process with input from the other departments, 12 members of staff in all were assigned to drafting and implementing the system.

55 Bulk SMS are sent using a 4 or 5 digit 'short code' (such as 1234) and blocks of numbers are issued to MNOs by the telecoms regulator, the Communications Commission of Kenya (CCK). Clients can pay CSPs extra for their allocated short code to be displayed as an 'alphanumeric' code instead. For example, if Kenya Power wanted to send bulk SMS to their customers to notify them of charges, instead of the customer receiving a message from the sender 1234, they receive a message from the sender Kenya Power. It is possible, and often cheaper, to buy bulk SMS bundles from international companies to send into Kenya. It is this aspect of the business of bulk SMS in particular that has caused controversy.

56 Gythan Munga, HumanIPO, *Safaricom Blocking International Bulk SMS*, February 24th 2013 www.humanipo.com/news/127/Safaricom-Blocking-International-Bulk-SMS

57 Paul Wafula, Standard Digital, January 4th 2013. *Kenya Working With Interpol to Fight Hate SMS*. www.standardmedia.co.ke/?articleID=2000074198&story_title=Kenya-Kenya-working-with-Interpol-to-fight-hate-SMS

58 Gythan Munga, HumanIPO, *Interview: Safaricom Clarifies Position on Bulk SMS Blocking*, March 7th 2013. www.humanipo.com/news/140/Interview-Safaricom-Clarifies-Position-on-Bulk-SMS-Blockage

(v) Developing internal guidelines to mitigate hate speech sent via bulk SMS

In February 2012, Safaricom approached the Communications Commission of Kenya (CCK), National Cohesion and Integration Commission (NCIC), the Independent Electoral and Boundaries Commission and the Registrar of Political Parties to develop national guidelines on bulk political messages. By June 2012, the government had not enacted any guidelines for mobile network operators, but Safaricom had developed its own internal guidelines. The aim was to avoid any language sent via their bulk SMS platform that was “partisan” or sought to divide communities, according to the Safaricom Regulatory & Public Policy team. It would also show that the company was conscious of the harm hate-filled messages can do, and by taking such steps, it felt it may avoid liability if hate speech was found on their platform. It had been reported that mobile network operators might be held liable if hate speech messages were found to have been transmitted via their network.⁵⁹

(vi) Reviewing Existing Laws

The Safaricom team reviewed applicable Kenyan hate speech laws, regulations, and guidelines as well as the terms of agreements with Content Service Providers (CSPs). This started with NCIA provisions on hate speech as well as cases where politicians had been arrested for disseminating hate speech under the Act.

One of the main sources that helped model Safaricom’s internal guidelines were those issued by national newspapers on political advertisements. After the mainstream media were held partly responsible for some of the post election violence, The NCIC wrote and issued guidelines to Kenyan media organisations on monitoring hate speech by the NCIC in 2010, and some publications had issued their own guidelines on political advertising in their publications. For example, the Nation Media Group (NMG), which owns Kenya’s most popular newspaper, the Daily Nation, released its Guidelines for political advertising in early 2012⁶⁰, which stated content would be vetted and the identity of the person or organisation sponsoring the advertisement must be clear. They added that advertisements would be rejected if they, amongst other issues, contain:

“Obscene or profane language or pictorial representation that, when taken in context, tends to or is likely to expose an individual or a group or class of individuals to hatred or contempt on the basis of race, national or ethnic origin, colour, religion, sex, sexual orientation, age or mental or physical disability.”

NMG’s guidelines also stated that no political advertisements were to be published 24 hours prior to or on the election day itself.

Safaricom said it looked to other “international benchmarks” such as the UK Guidelines on Premium Rate Messaging and the International Covenant on Civil and Political Rights (ICCPR), the Convention on the Elimination of All Forms of Racial Discrimination (ICERD) and the European Convention for Protection of Human Rights and Fundamental Freedoms. The UN Guiding Principles on Business and Human Rights was not a standard the company had considered. To be sure, the Guiding Principles, adopted only in 2011, underscore the importance of due diligence processes, including identifying risks and impacts, but do not necessarily provide guidance on the steps companies should take to mitigate specific risks.⁶¹

59 Paul Wafula, Standard Digital, January 4th 2013. *Kenya Working With Interpol to Fight Hate SMS*: www.standardmedia.co.ke/?articleID=2000074198&story_title=Kenya-Kenya-working-with-Interpol-to-fight-hate-SMS

60 Daily Nation, January 11th 2013, *NMG Guidelines for Political Advertising*: www.nation.co.ke/News/NMG-guidelines-for-political-advertising-/1056/1663614/-/cqcqeczl/-/index.html

61 The European Commission recently released the *ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights*, written by IHRB and Shift, to help ICT companies “translate” the UN Guiding Principles into their own systems and cultures: www.ihrb.org/pdf/eu-sector-guidance/EC-Guides/ICT/EC-Guide_ICT.pdf

(vii) Guidelines for Political Mobile Advertising on Safaricom's Premium Rate Messaging Network

The resulting publication, *Guidelines for Political Mobile Advertising on Safaricom's Premium Rate Messaging Network*⁶², drew heavily on the efforts developed by other media organisations and tackled a range of issues that contributed to the spread of hate speech during the 2007 elections. Instead of a CSP buying bulk SMS bundles and sending any message a client wanted, CSPs now had to make an application to Safaricom at least 48 hours before they wanted to send a message that was political in nature only from June 1, 2012, to March 2, 2013, the main election period, although the guidelines would continue to operate after the election results were announced.

Prior to the guidelines, there was no distinction between political bulk SMS and other types of bulk SMS such as news or commercial offers. Those CSPs who chose to offer bulk SMS services to political parties and had a license agreement with Safaricom were required to follow the guidelines. Messages reporting political news were not considered to fall within the guideline as the media have their own governing guidelines and the 48-hour vetting process was considered unsuitable for news. The guidelines tackled a number of issues thought to have contributed to the 2007 post election violence including:

Identifying the sender: During the post election violence of 2007/2008, mobile phone users reported receiving SMS messages from unknown numbers instructing them to attack their neighbours. It is thought that some politicians were behind the origin of these messages which may have originated from a batch of bulk SMS. Safaricom insisted that SMS must feature the signature of an identifiable person. It was felt that politicians could hide behind anonymity and ultimately deny sending hate speech messages. To apply to send political bulk SMS, the CSP had to instruct the applicant to fill out a special form containing specific information:

- The verbatim content of the political message;
- A signed authorisation letter from the political party representative or individual sending the message;
- Certified copies of political party registration documents or a copy of the applicants' national ID card.

Language: There are 42 languages in Kenya, and many are spoken by specific ethnic groups, and not more widely. In addition, many young people speak Sheng, a form of slang and a mixture of Swahili and English. During the post election violence of 2007/2008, some politicians spoke in their mother tongue at rallies to deliberately exclude other communities. In some such speeches heavily coded messages were issued. Incitement uttered in local languages also meant the risk of detection was lowered. In order to combat this, Safaricom insisted in its new guidelines that content for political bulk SMSs was sent in the official Kenyan languages of Kiswahili and English only.

Content: Safaricom would 'vet' the content of the proposed message in accordance with the guidelines, which drew on the NCIA and the Electoral Code of Conduct, and reserved the right to refuse transmission of messages. The guidelines pertaining to content stated:

3.3 Political Messages shall not contain offensive, abusive, obscene or profane language.

3.4 Political Messages shall not contain inciting or discriminatory language that may or is intended to expose an individual or group of individuals to hatred, hostility or ridicule on the basis of ethnicity, tribe, race, colour, religion, gender or otherwise.

3.5 Political Messages shall focus on Party manifestos and shall not dwell on unnecessary attacks on individual persons, their families, their tribe or their associations.

62 Download the Guidelines here: bloggers.or.ke/safaricom-issues-tough-guidelines-for-political-messaging/

The messages then had to be personally approved and signed off by the Senior Management from the Internet and Content and Regulatory and Public Policy team before the CSP could transmit the message.

Unsolicited messages: To date, Kenya has no specific data protection law (although a draft bill is currently under review), but the Kenyan Information and Communications Act specifically prohibited the sending of 'unsolicited messages' i.e. receiving messages from a service to which the user had not explicitly signed up.⁶³

During the post election violence of 2007/2008, some mobile users reported receiving up to 50 'unsolicited' SMS per day.⁶⁴ The Safaricom guidelines therefore state explicitly that sending messages of a political nature must be on an opt-in basis, or through subscription. If users complained they had received a message from a short code to which they had not signed up, Safaricom could ask the CSP to produce a list of subscribers, to look for evidence that they had consented to the service. Otherwise they would reserve the right to immediately terminate the license agreement with the CSP. Politicians would most likely target their party members, but there were suspicions that lists of numbers were being procured from other sources, for example, details given when signing up for a super-market loyalty card or even details being sold by M-PESA agents⁶⁵, and there are no laws in place to prevent such data sharing.

5. National Guidelines to Prevent Transmission of Undesirable Bulk Political Content via Electronic Communications Networks

By July 2012, the CCK had formed the *Technical Committee to Develop the Guidelines for Undesirable Bulk Content*, consisting of the NCIC, the Independent Electoral and Boundaries Commission (IEBC), and mobile operators. As the CCK had already begun the process of issuing guidelines for the media, they agreed a separate process for SMS was necessary. The CCK technical committee developed draft *Guidelines for the Prevention of Transmission of Undesirable Bulk Political Content via electronic Communications Networks*⁶⁶ by August 2012. The CCK draft Guidelines were thereafter published and subjected to stakeholder and public consultations. Throughout the drafting of their own guidelines, which were finished by June 2012, Safaricom did not consult with local NGOs or freedom of expression experts to consider the implications of vetting messages. Later on in the process, Safaricom was criticised by civil society for this and were seen as interfering with the right to freedom of expression.

The stakeholder consultation for the national guidelines involved the publication of the Guidelines for public and stakeholder comments (including CSPs), review of these comments and a stakeholder workshop to discuss the final Guidelines in September 2012. A public notice regarding the consultation was published in national newspapers. However some civil society groups have said it wasn't made clear they could provide input.

63 Kenya Information and Communications Act 1998 (as amended in 2008), para 17, p241:
whiteafrican.com/wp-content/uploads/2012/08/Kenya-Information-Communications-Act-Final.pdf

64 Tim Querengesser, Cellphones Spread Kenyans Messages of Hate. The Globe and Mail, Canada. 29th February 2008:
www.theglobeandmail.com/technology/cellphones-spread-kenyans-messages-of-hate/article1052421/

65 DW, Kenya, Privacy Concerns in Kenya as Users Turn to MPESA to Catch Cheating Partners. 12th July 2013:
www.dw.de/privacy-concerns-in-kenya-as-users-turn-to-m-pesa-to-catch-cheating-partners/a-16947446

66 Communications Commission of Kenya, Guidelines for the Prevention of Transmission of Undesirable Bulk Political Content via Electronic Communications Networks.
www.cck.go.ke/regulations/downloads/Guidelines_for_the_prevention_of_transmission_of_undesirable_bulk_political_content_via_sms.pdf

The Guidelines were similar to Safaricom guidelines but now applied to all mobile network operators and licensed CSPs. They also incorporated concerns about complaints handling and confirmed that instances of hate speech, or messages mobile network operators were unsure about, would be referred to the NCIC for a final decision on whether they could be sent or not. The final Guidelines were released on 24 October, 2012.

The CCK guidelines sent a strong message to politicians that they were expected to adhere to certain standards. One of NCIC's fears was that it would be overloaded with messages referred by MNOs for clarifications that they contained hate speech, but this did not happen. During the time IHRB staff visited Safaricom headquarters, the company received, on average, 3-4 requests daily for political messaging, and none could be characterised as hate speech under international human rights law. Safaricom referred one message to the NCIC throughout the period IHRB spent at its offices. This message was referred because it spoke of land rights, which is a controversial subject in Kenya. However, the NCIC deemed the message suitable for transmission.

The system was not perfect. One CSP reported that its system was hacked, which allowed a message to be sent without being 'vetted' by Safaricom first. That message alleged that a particular political candidate was an alcoholic. The individual complained to the police, which took the investigation out of Safaricom or CCK's hands. In another instance, an applicant claimed to be an MP when he wasn't, and in one case a signature on a message did not match the name on the national ID card provided.

IHRB observed that the majority of messages requested by political parties and aspirants to be sent out as bulk SMS had innocuous content. The text was straightforward with little room for interpretation. Messages typically notified users about rallies, thanked their supporters, or highlighted particular political policies of candidates. A few specifically advertised their SMS short code on their website or in the local or national media. But in reality, the process of finding political SMS services to sign up to was quite difficult as they were not widely advertised and IHRB's researcher was not a member of any Kenyan political party and therefore not on any subscriber lists the parties may have already had. Anecdotal evidence also suggests that Safaricom customers were still receiving unsolicited messages and there was concern that personal details, including national identity numbers, were being sold to political parties by third parties to boost their membership.

6. The Elections

Monday, March 4 2013, was declared a national holiday and Kenya went to the polls. Apart from terrible attacks on one polling station in Mombasa, in which 17 people died, the rest of the country voted largely in peace. An election monitoring project set up by Ushahidi called Uchaguzi (meaning 'election' in Swahili) reported that the most popular topic being texted in by the public was how 'peaceful' the voting was across the country.⁶⁷ The voter turnout was reported at 86%⁶⁸, the highest turnout recorded in Kenya since the introduction of the multi-party system in 1992. IHRB staff visited 3 Nairobi polling stations in Riruta, Kawangware and Kibera and found the atmosphere was quiet and calm.

The final figures from Safaricom on the implementation of the *Guidelines for the Prevention of Transmission of Undesirable Bulk Political Content via Electronic Communications Networks* showed that out of around 100 licensed CSPs, eight offered political messaging services. One content service provider interviewed by IHRB said that a large number of CSPs would not be offering a political messaging service as it was felt there was too much liability placed on the CSP. CSPs were very concerned about who would bear responsibility for the messages, although there is a provision in the CCK content service provider license that assigns the liability for content on the CSP and CSPs seek indemnity from their clients.

Safaricom provided IHRB with the final figures regarding requests for political bulk SMSs only, from June 1st, 2012 (when Safaricom implemented their own guidelines) to March 2, 2013 (two days before the elections):

- 27 clients used Safaricom's network to send political bulk SMSs, consisting of 21 individual political candidates, three political parties and three independent organisations.
- 68 requests to send political messages were submitted, which translated to 963,762 SMSs being sent in total.
- Some 62 of the requests to send political messages were approved by Safaricom within 48 hours after the content was vetted for hate speech.
- The Regulatory & Public Policy team rejected at least 18 message requests outright for various reasons, such as, for example, failing to submit a copy of ID, or specifying who was signing off the message. These message requests were sent back to the client for amendment due to non-compliance with the guidelines. Five out of these 18 were never returned.
- Safaricom referred one message to the NCIC, concerning land rights as mentioned above, as this was viewed a controversial subject to be mentioning in a political message. The NCIC approved the message and it was allowed to be sent out.

These numbers reflect that bulk SMS was used much less for political communication than Safaricom originally predicted. It should be acknowledged that these numbers are a fraction of the total number of SMSs transmitted and communicated during the election period. Some civil society organisations criticised the guidelines for tackling what they described as 'low hanging fruit' and ignoring the wider problem presented by peer-to-peer SMSs. But if Safaricom monitored the peer-to-peer SMSs, it would probably run the risk of breaking privacy laws, and certainly affect privacy rights, and by not doing so, it was acting within the law and recognising privacy rights.

In the days following the elections, people mostly stayed at home. The national television stations streamed voting results from the IEBC centre in Bomas 24 hours a day. This was a complete contrast to the 2007/2008 elections when there was a government imposed media blackout, which fuelled suspicions about the results. The official results of the March 4 elections were announced on March 9, with Uhuru Kenyatta receiving 50.7% of the votes,

⁶⁷ Uchaguzi. Analysis Team Report, March 5th, 20:30: sitroom.uchaguzi.co.ke/2013/03/05/analysis-team-report-march-5th-2030-eat/

⁶⁸ The Independent Electoral and Boundaries Commission (IEBC) Summary of 2013 Presidential Results Declared on 9/3/2013: www.kenyaelections.com/wp-content/uploads/SUMMARY-OF-2013-PRESIDENTIAL-RESULTS-DECLARED-ON-9_3_2013.pdf

compared to Odinga's 43%. Odinga contested the outcome, claiming voting irregularities and petitioned the High Court to overturn the judgment and order fresh elections. The Court ruled against Odinga and Kenyatta was inaugurated as President of Kenya on April 9 2013.

(i) Confusion over SMS blocking

After the elections, it was reported that the Permanent Secretary in the Ministry of Information and Communications at the time, Bitange Ndemo, announced that telecommunications providers in Kenya had blocked 300,000 SMS per day during the elections and that telecommunications companies had installed software that would flag words like 'kill'.⁶⁹ The statement was unclear as to whether it was referring to peer-to-peer SMSs or bulk SMSs coming from abroad. If Dr. Ndemo was referring to the blocking of peer-to-peer messages, it would raise questions about possible privacy violations. It is difficult to clarify exactly what has happened in this case, which raises issues of transparency and accountability with regards to filtering content. Safaricom had announced some months earlier the policy of blocking bulk SMS messages originating from abroad and maintains it does not block peer-to-peer messages based on any criteria.

Interviews with industry experts and civil society groups suggest that keyword filtering of bulk SMS does take place in some countries, and there are documented examples from Egypt and Pakistan. This can take place to prevent spam, as the keyword list in Pakistan purports to do by blocking a list of "harmful" words from SMS sent "in bulk"⁷⁰, although there is no definition of "in bulk". There are obvious concerns with how this practice could be misused in some States to stifle political opposition and freedom of expression, as the following example from Egypt demonstrates.

In 2010, a regulation in Egypt led to a court case over a regulatory decree to monitor all bulk SMSs. In Egypt's case, the trigger was a message circulating from abroad that former president Hosni Mubarak had died. Documents seen by IHRB detail meetings between the Egyptian Ministry of Communications and three mobile telecommunications companies to discuss not only blocking messages from abroad, but also restricting Internet access and mobile communications to particular areas. The Egyptian National Telecommunication Regulation Authority (NTRA) issued a decree that imposed restrictions on bulk SMS in that CSPs should seek approval of *all* message content from government authorities, including security services. The decree also allowed the NTRA and security services to enter the premises of CSPs and monitor their work and also impose high fees on CSPs wanting to provide any kind of bulk messaging. Newspapers and some human rights organisations relied on bulk messaging to disseminate information and the concern was that this was an attempt to restrict the flow of information and the efforts of civil society. An Egyptian civil society organisation, the Association for Freedom of Thought and Expression (AFTE), challenged the decree in court and won; the court ordered a cessation of the decree on the grounds that it violated freedom of expression.⁷¹

However, one critical finding from the Kenyan elections is that while the Guidelines for the Prevention of Transmission of Undesirable Bulk Political Content via Electronic Communications Networks sent a clear message to politicians, political parties and aspirants that their speech was being scrutinized, hate speech was not eradicated from the campaign trail, and ICTs were still being used to spread it. As the next section discusses, it seems most of the contentious traffic moved from SMS to social media. The next section examines the impact of social media.

69 Daily Nation, Kenya. Phone Firms Block 300,000 Hate Texts Daily, says Ndemo. March 21st 2013: www.nation.co.ke/news/Phone+firms+block+300+000+hate+texts+daily+says+Ndemo+/-/1056/1726172/-/f4oj2iz/-/index.html

70 propakistani.pk/wp-content/uploads/2011/11/72911632-PTA-Letter-on-Content-Filtering.pdf
propakistani.pk/2011/11/16/breaking-mobile-companies-to-filter-sms-content/
propakistani.pk/2009/08/19/protection-from-spam-unsolicited-fraudulent-and-obnoxious-communication-regulations-2009/

71 AFTE sues the National Telecommunications Regulatory Authority for Imposing Restrictions on Collective SMSs: www.en.afteegypt.org/index.php?newsid=39

7. The impact of social media on efforts to curb hate speech during the Kenyan election

Internet users in Kenya have increased from 3 million in 2007 to 16.4 million in 2013.⁷² The increase in connectivity and the rise in the number of social media users have presented a new challenge to Kenya in combatting hate speech. With tightened controls in the telecommunications sector and the print media, hate speech seemed to find a new outlet during the 2013 elections on social media sites such as Facebook and Twitter, both of which are US-based companies. There were warning signs from the 2007/2008 elections that social media may present a challenge in 2013, as some bloggers felt they had to moderate content.

Despite the Kenya National Commission on Human Rights 2012 pre-election briefing warning that instances of “hate, stereotypes and harmful propaganda against fellow Kenyans on the basis of their ethnicity” were making their way onto social media, in particular Facebook and Twitter⁷³, there is a sense that the Kenyan government, stunned during the 2007 elections with the spread of hate speech on traditional media and SMS, were taken by surprise again in 2013 by the conversations on social media. In the run up to the elections, the government publicly and repeatedly reported it was clamping down on hate speech and claimed law enforcement officials were actively monitoring social media. However, there is little evidence of this and it may have been a tactical move to make people believe that the government was monitoring content so as to dissuade people engaging in hate speech online. With all the attention to hate speech online during the 2013 elections, it has been reported that 6 people are being investigated, including two prominent bloggers who deny the charges.⁷⁴

One police officer was quoted as saying:

“We are trying our best to combat hate speech online, but it’s hard for us because many officers fighting cyber crime have insufficient or no knowledge at all about the dynamics of technology.”⁷⁵

A recent report by The Institute for War and Peace Reporting⁷⁶ quotes Stanley Cheruiyot, Kenya’s assistant Police Commissioner:

“We prepared to provide physical monitoring of hate speech [by providing recorders to monitor political rallies]...But the offence mutated... it was displaced to social media. The police are not trained so much on monitoring online hate speech.”

The Umati project, set up by iHub Research and Ushahidi to monitor the Kenyan online space during the elections, found online calls to discriminate against certain tribes, with invocations to kill, beat and forcibly evict those ethnic groups in the weeks leading up to and following the elections.⁷⁷ The majority of this content was found on Facebook, which in Kenya has 2 million users and is the most popular social media platform, in both public and private group pages.⁷⁸ In interviews conducted by IHRB with government representatives and civil society organisations in Kenya, there seemed to be confusion over how to notify Facebook about potentially illegal content, the expectations of the actions Facebook would take, and the lack of response on Facebook’s part to user concerns. The issue received

72 Communications Commission of Kenya, Sector Statistics Report, January-March 2013: www.cck.go.ke/resc/downloads/Sector_Statistics_Report_for_3rd_Quarter_2012-2013.pdf

73 KNCHR, 15th February 2013. *We Haven’t Learnt Yet: A Periodic Update on Hotspots and Campaign Monitoring*, para 11: http://www.knchr.org/Portals/0/PressStatements/Preliminary_report_on_KNCHR_Election_Monitoring.pdf

74 Judie Kaberia, Capital FM Online, July 11th 2013 *Kenya: Too Little Action on Hate Speech?* www.capitalfm.co.ke/news/2013/07/kenya-too-little-action-on-hate-speech/

75 Bernard Koech, Kenya Online, February 26th 2013, *Tackling Online Hate Speech in Kenya*: reportingkenya.net/reports/tackling-online-hate-speech-in-kenya/

76 Institute for War and Peace Reporting, *Kenyan Authorities in the Dock Over Hate Speech*, 3 May 2013, ACR Issue 347: www.refworld.org/docid/5189153c4.html

77 Umati: Monitoring Online Dangerous Speech. February and March 2013 report, p13: www.research.ihub.co.ke/uploads/2013/april/1365508815_819_823.pdf. Umati Final Report June 2013, p20: blog.ushahidi.com/2013/06/28/umati-final-report-released/

78 Umati Final Report June 2013, p24: blog.ushahidi.com/2013/06/28/umati-final-report-released/

coverage in the Kenyan press and rumours circulated: on the one hand the government was blaming Facebook for inaction⁷⁹, on the other “contracts had been signed” between the government and Facebook in which software would track pages for hate speech.⁸⁰

The Umati project had repeatedly made attempts to reach out to Facebook through formal and informal networks, but received no response. Instances of hate speech in Kenya are to be reported to the NCIC, but it is unclear whether the NCIC had reached out to Facebook on receiving reports. This confusion is the reason why some activists in Kenya criticise Facebook and other websites which feature user-generated content, for a lack of clarity in their terms and conditions, the difficulty of reporting content, and failure of the company to communicate reasons for decisions, which adds up to a lack of trust and a negative image for the company. If companies do not have clear policies, and are not transparent about their processes or fail to communicate with users, then they can often be blamed or be used as a scapegoat for failings, which may not be theirs alone.

Due to the attention focused on Facebook during the 2013 elections, the remainder of this chapter will focus largely on Facebook activity during the elections and the actions the company took once IHRB and Umati alerted them to the potential problem.

(i) Facebook

Facebook’s Community Standards, which outline what material is not allowed on the company’s social media platform, defines ‘hate speech’ thus:

*Facebook does not permit hate speech, but distinguishes between serious and humorous speech. While we encourage you to challenge ideas, institutions, events, and practices, we do not permit individuals or groups to attack others based on their race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or medical condition.*⁸¹

While these standards describe in general terms how Facebook distinguishes between types of speech, it is unclear what content the company specifically will or will not allow on the site. As will be discussed in more detail, there is no definition in international law of hate speech, and in the statement above there is the use of an undefined term like ‘attack’, which is open to interpretation.

Facebook recently pledged to review its policies on hate speech, after criticism that it failed to remove content depicting or glorifying violence against women⁸², particularly after a few high profile instances in the UK. The announcement that Facebook has joined the Global Network Initiative is a positive step forward in this regard.⁸³ Marine Levine, VP of Global Public Policy at Facebook, stated in a Facebook blog:

“In recent days, it has become clear that our systems to identify and remove hate speech have failed to work as effectively as we would like, particularly around issues of gender-based hate. In some cases, content is not being removed as quickly as we want. In other cases, content that should be removed has not been or has been evaluated using outdated criteria. We have been working over the past several months to improve our systems

79 Sabahi, Kenyan Monitoring Committee Urges Facebook to Expose Hate Speech, February 21st 2013: sabahionline.com/en_GB/articles/hoa/articles/newsbriefs/2013/02/21/newsbrief-05

80 Kenya News Updates, 29th April 2013. *Has NCIC Failed to Curb Rampant Hate Speech on Social Media?* www.kenyanewsupdates.com/opinion/item/509-has-ncic-failed-to-curb-rampant-hate-speech-on-social-media.html

81 Facebook Community Standards - there is a separate section for Violence and Threats: www.facebook.com/communitystandards

82 See *Women, Action and Media*, Open Letter to Facebook, May 21st 2013: www.womenactionmedia.org/facebookaction/open-letter-to-facebook/.

Dave Lee, BBC News online, *Facebook Bows to Campaign Groups over ‘Hate Speech’*, May 29th 2013: www.bbc.co.uk/news/technology-22701082

83 Global Network Initiative, *Facebook Joins the Global Network Initiative*, May 22nd 2013: www.globalnetworkinitiative.org/news/facebook-joins-global-network-initiative

to respond to reports of violations, but the guidelines used by these systems have failed to capture all the content that violates our standards. We need to do better – and we will.”⁸⁴

At a panel on social media and hate speech at the Kenya Internet Governance Forum (IGF) in July 2013, IHRB invited Richard Allan, Facebook’s Director of Policy for Europe, Middle East and Africa, to take part and explain Facebook’s policies in dealing with reported instances of hate speech and in particular in the context of the 2013 elections in Kenya. Below is a summary of that session.

Facebook does not actively moderate content and relies on reports from users (or ‘members’) who flag content they believe should be removed. Users often flag content they find offensive, but which may not be illegal. Due to the complexity in distinguishing between content that is offensive and harmful and content which is not, and the criticism Facebook receives for removing content, Facebook appears to be moving towards a less ‘cut and dry’ approach on blocking content. Allan spoke of Facebook’s introduction of a tiered response to allegations of hate speech:

“A couple of years ago we had a situation where we just shut down a user’s account if they breached our rules. Now we are moving into an educational space. In terms of the sanctions we will take, the first is to remove the item of content itself and that’s a routine sanction we will take. The next would be to warn the individual at the same time as removal, what we call an “educational checkpoint”. Someone will now get a warning message saying, we’ve removed this content, you need to acknowledge you shouldn’t post content like this in order to regain access to your account. That’s something we developed reasonably recently.”⁸⁵

Once the Facebook review team has made a decision, the content will either be removed or not and a message sent to the user who flagged the content to say that the content does not contravene company standards, if it does not. This can be frustrating to users when they receive a generic message with no further explanation, especially in a time of heightened sensitivity.

One example of note in the Kenyan context concerns the Facebook page *Not Another Kikuyu President, Please*. Although it may not constitute hate speech under the international legal definition, the page title and content do appear to ‘attack’ a group based on its ethnicity. When IHRB tested the ‘report abuse’ function in regard to this page (because it appeared to contravene Facebook’s Community Standards), the response it returned stated that the material did not violate Facebook’s Statement of Rights and Responsibilities (terms a user agrees to when using Facebook), with no further explanation.

When the controversial film *Innocence of Muslims* appeared to instigate violent protests in several parts of the world in 2012, YouTube blocked access to the film in certain territories and displayed a generic message to users. Rebecca Mackinnon, Internet freedom activist and Senior Research Fellow at the New America Foundation, and Susan Benesch, Faculty Associate at the Berkman Centre for Internet and Society at Harvard University and genocide prevention fellow at the United States Holocaust Memorial Museum, wrote that in highly sensitive situations like this, the company should consider a more customised message explaining to the user why certain content had been blocked:

“Explanations would set a good example of transparency and would contribute to debate about the contours of freedom of expression, especially in countries that are working out their own rules for regulating speech as they make transitions toward democracy, often in volatile contexts.”⁸⁶

84 Marine Levine, Facebook blog, *Controversial, Harmful and Hateful Speech on Facebook*, May 28th 2013:

www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054

85 All Richard Allan quotes taken from Comments made during panel on hate speech and social media, chaired by IHRB, Nairobi 26th July 2013: bit.ly/kigf13webcast

86 Rebecca Mackinnon and Susan Benesch, *Foreign Policy*, October 5, 2012. *The Innocence of YouTube*:

www.foreignpolicy.com/articles/2012/10/05/the_innocence_of_youtube?page=0,1

The problem for Kenyans, however, was that Facebook was not removing content individuals felt was hate speech. To address this, Facebook is developing an initiative that provides other options aside from the decision to block or not to block and encourages interaction between users to resolve issues. Allan explained:

“We developed something called social reporting, where we put a stage in there if a user wants to report something we encourage them to report it to the person who posted it... I think this same kind of mechanism could be very useful for social feedback for content that is borderline hate speech that needs to be addressed by the person who posted the content.”

Whether this will work in practice remains to be seen, especially in Kenya. However encouraging “social reporting” takes advantage of the unique aspects of social networking, which allows interaction and a ‘right to reply’ on a scale that SMS does not. For example, one post in the page *Not Another Kikuyu President Please*, posted on March, the day before the election results were officially announced, read:

“Uhuru is President? Luos will have their dicks chopped off and Kalenjins will not be safe in their land no matter what. Pretend but it is the sad truth.”

This post attracted 320 comments largely denouncing the statement, which in turn dilutes the threatening sentiment therein. If this statement were distributed as an anonymous leaflet or SMS, the message would carry more weight, as there would be no right to reply, nor would such replies be seen. Here lies the power inherent in social networking to counter negative speech with positive speech, something both business and government should be promoting. At the time of writing⁸⁷, the administrator of that page has not posted any further material since March 30th, but the page still remains.

Part of the problem seemed to be that some instances of perceived hate speech were not in wall posts, which can be reported to Facebook, but in the comments on those posts; a situation for which there is no direct reporting mechanism. According to Facebook, the page Administrator is responsible for moderating the page, including comments, in the same way a blogger would moderate comments on their own blog. This means a user must contact the page Administrator to request a comment be removed. But what if the Administrator refuses to do so, by not engaging with the “social reporting”, and/or has set up the page to encourage inflammatory comments and is therefore unlikely to respond to requests for content removal? Allan responds:

“Ultimately, if somebody continues to post in spite of these warnings, their page will be shut down and if it’s a public page which is a common way for people to put forward quite robust views then the public page will be shut down once the administrators have been warned a number of times, and if the administrators themselves continue to breach the rules they themselves will face the loss of their accounts.”

But what about specific situations where, given the context, Facebook may be used as a platform to spread hatred that could lead to mass violence or even genocide? Allan describes Facebook’s actions during the Kenyan elections:

“Once we saw the Umati report and we had feedback the elections were a potential flashpoint, we put the teams on high state of alert...Then we stood ready to make contact with organisations on the ground if we saw incidents occurring. Our press and PR team monitor constantly what is happening in the news so we can see if Facebook is involved in a particular incident...Had we received reports to say there were incidents that were Facebook related, then we would have escalated it further, but we weren’t getting reports. If people in Kenya feel that was inaccurate and we could have done more I’d be interested in hearing that feedback, but I was personally monitoring closely and the team was monitoring closely what was happening in the news and we weren’t getting advice that, fortunately on this occasion as opposed to the previous election, that things had gone in a dangerous direction.”

87 Last checked 27th October 2013.

In order to decide whether to escalate a situation, Facebook relies on a number of sources: reports from members, mainstream media and what Allan describes as “people on the ground who email us and say what is going on.”

It is unclear the extent to which Facebook members in Kenya were reporting content and the extent that mainstream media was reporting instances of violence for fear of being accused of spreading hatred. Therefore it is extremely important that companies have clear policies and mechanisms regarding reporting content and manage users’ expectations as to what will happen next and most importantly, respond to users raising concerns.

Globally, Facebook receives two million requests for content takedown per week⁸⁸ and knowing how to respond appropriately is a difficult job.⁸⁹ One way forward could lie in inviting local groups, such as Umati, which have an extensive collection of data, to join the “less formal channels of communication Facebook has previously established with other groups” that Marine Levine, VP of Global Public Policy at Facebook, refers to in the aforementioned Facebook blog.⁹⁰ Groups like Umati are willing to contribute technical and country-level knowledge to Facebook to assist them in what is a very challenging task. Allan acknowledges this: “Where there are particular incidents and where there are people on the ground who have valuable insights into how people are using our platform then all of that feeds into trying to improve these systems.”

As with any kind of hate speech, there are deep-rooted issues that will not be solved by removing a Facebook post. There is also the risk that in situations of heightened tension, such as elections, speech that is controversial and offensive may be labeled as hate speech when it may not be so. There is the added pressure for companies of popular opinion and the view that offensive material that is not illegal should be removed. While Allan admitted there is still “a long way to go”, and how these measures work in practice is yet to be seen, this tiered response signifies an acknowledgement that there are grey areas when dealing with restrictions on freedom of expression which demand different responses.

In order to form appropriate responses, companies can go back to international human rights law, which sets out the threshold for legitimately restricting freedom of expression. The next chapter analyses freedom of expression and legitimate restrictions as outlined in international human rights law, and highlights cases of conflicting obligations of international and national law.

88 Jeffrey Rosen, The New Republic, The Delete Squad: Google, Twitter, Facebook and the New Global Battle Over the Future of Free Speech, April 29th 2013. Available at, www.newrepublic.com/article/113045/free-speech-internet-silicon-valley-making-rules

89 Matthew Ingram, GigaOm, Do We Really Want Facebook to Decide What Qualifies as Hate Speech and What Doesn't? , May 29th 2013: gigaom.com/2013/05/29/do-we-really-want-facebook-to-decide-what-qualifies-as-hate-speech-and-what-doesnt/

90 Marine Levine, Facebook blog, Controversial, Harmful and Hateful Speech on Facebook, May 28th 2013: www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054

8. Freedom of expression and its restrictions in international law

Almost every country's constitution recognises the right to freedom of expression and opinion in some form. According to the UN Human Rights Council, these rights are considered "indispensable conditions for the full development of a person"⁹¹ and essential for a free and democratic society. Freedom of expression is also seen as an enabling right, essential to promote and protect other human rights. ICTs can enable the right to freedom of expression; as more people become technologically connected, there are more avenues and platforms to seek, receive, and impart information. Access to the Internet is now well regarded as an enabler to freedom of expression.⁹²

The *Universal Declaration of Human Rights*, which first guaranteed freedom of expression in 1948:

*Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.*⁹³

Article 19 of *The International Covenant on Civil and Political Rights (ICCPR) (1966)*, also guarantees freedom of expression⁹⁴:

1. *Everyone has the right to hold opinions without interference.*
2. *Everyone shall have the right to freedom of expression; this right shall include the freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.*

Regional human rights conventions also enshrine the right to freedom of expression, including Article 9 of the *African Charter on Human and People's Rights*⁹⁵:

1. *Every individual shall have the right to receive information*
2. *Every individual shall have the right to express and disseminate his opinions within the law.*

The right to freedom of expression and opinion extends to ideas deemed unpopular, shocking, offensive or disturbing. The UN Special Rapporteur on the Promotion and Protection of Freedom of Opinion and Expression, Frank La Rue, outlines this in a 2012 report:

"The right to freedom of expression implies that it should be possible to scrutinize, openly debate and criticize, even harshly and unreasonably, ideas, opinions, belief systems and institutions, including religious ones, as long as this does not advocate hatred that incites hostility, discrimination or violence against an individual or a group of individuals."⁹⁶

91 Human Rights Committee, "General Comment No. 34: Article 19: Freedoms of opinion and expression," UN Doc CCPR/C/GC/34, 12 September 2011: www.refworld.org/docid/4ed34b562.html

92 See Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue, 16th May 2011. UN General Assembly A/HRC/17/27.

93 UDHR, 1948, Article 19.

94 Kenya ratified the ICCPR on 1st May 1972 and it came into force on 23rd May 1976

95 African Charter on Human and People's Rights, 1981 (also known as the 'Banjul' charter: www.achpr.org/instruments/achpr/)

96 UN General Assembly, Promotion and Protection of the Right to Freedom of Opinion and Expression. Note by the Secretary General. 10th August 2011, A/66/90, Para 30: www.ohchr.org/Documents/Issues/Opinion/A.66.290.pdf

As La Rue indicates, the right to freedom of expression is not an absolute right, such as the right to life, and it comes with caveats in order to protect the rights of others. Article 19, paragraph 3, of the ICCPR lays the groundwork:

The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:

(c) for respect of the rights or reputation of others

(d) for the protection of national order (ordre public), or of public health or morals.

Article 20 of the ICCPR examines these restrictions in more depth:

1. *Any propaganda for war shall be prohibited by law*
2. *Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.*

In other words, according to Article 20 (2), hatred, by itself, would not be subject to restriction. It is only when advocacy of national, racial or religious hatred constitutes incitement to discrimination, hostility or violence that it is considered illegal under international law. Incitement is recognised as a crime in further international human rights standards. *The UN Convention on the Prevention and Punishment of the Crime of Genocide (1948)* criminalises a “direct and public incitement to commit genocide.”⁹⁷

The UN International Convention on Elimination of All Forms of Racial Discrimination (1966) makes it obligatory that State parties,

*Shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or groups of persons or another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof.*⁹⁸

This convention builds on ICCPR Article 20 (2) by declaring dissemination of ideas based on racial superiority and assisting or financing racist activities as a punishable offence by law. One unfortunate omission, however, is gender, which is not specifically considered in these instruments, nor is advocacy of hatred that incites violence towards women in the Convention on the Elimination of all Forms of Discrimination against Women (1976). This is particularly relevant to the current global debate on the level of threats of violence online directed towards women.⁹⁹

However, when it comes to declaring what speech should be restricted, the conventions help clarify that there should be a very high threshold, and that speech should only be restricted in the most severe situations, such as when it incites violence or when such a threat is imminent. While each country has its own versions of discrimination and hostility towards certain members or groups in society, there are concerns that hate speech is often vaguely

⁹⁷ UN Convention on the Prevention and Punishment of the Crime of Genocide (1948) Article III(c). Available at, <http://www.hrweb.org/legal/genocide.html>

The International Crimes Tribunal for Rwanda, *The Prosecutor vs. Nahimana, Barayagwiza and Ngeze*, also known as the “Media” case in which two executives from Radio Television Libre des Milles Collines (RTL) and one newsletter publisher were convicted of “direct and public incitement to genocide” with regard to the violent messages they transmitted/published: www.refworld.org/cgi-bin/texis/vtx/rwmain?page=publisher&publisher=ICTR&type=&coi=&docid=48b5271d2&skip=0

⁹⁸ Article 4(a)

⁹⁹ See Internet Democracy Project (2013) *Keeping Women Safe: Gender, Harassment and Indian Law* available at, <http://beta.internetdemocracy.in/media/keeping-women-safe-gender-online-harassment-and-indian-law-2/>. Women, Action and the Media (2013) *Open Letter to Facebook*, May 21st 2013. Available at, <http://www.womenactionmedia.org/facebookaction/open-letter-to-facebook/>, Sam Jones, *The Guardian*, 29th July 2013. Labour MP Stella Creasy receives Twitter rape threats. Available at, www.theguardian.com/technology/2013/jul/29/labour-mp-stella-creasy-twitter-rape-threats

defined in national laws, and may be inconsistent with Article 20 (2) of the ICCPR.¹⁰⁰

Vaguely defined terms extend to other forms of speech; broad terms in some States' national laws, such as combatting "incitement to religious unrest", "promoting division between religious believers and non-believers", "defamation of religion", "inciting to violation", "instigating hatred and disrespect against the ruling regime", "inciting subversion of state power" and "offences that damage public tranquility", are examples that do not meet the criterion of legal clarity, according to Frank La Rue.¹⁰¹ Nonetheless, we are seeing people being arrested, imprisoned and beaten for expressing opinions which are entirely reasonable and in line with international law, but which are interpreted as illegal under some national laws.

This confusion and concern that national law is being used to suppress dissent, in the name of combatting hate speech, led La Rue to emphasise the need to distinguish between three different types of speech.¹⁰² These different categories call for different responses and depend very much on the context in which they were spoken, but could be adopted by companies to decide appropriate levels of action regarding requests to takedown content:

1. "Those that are an offence under international law and can be criminally prosecuted."

La Rue specifies this as, (a) child pornography¹⁰³; (b) direct and public incitement to commit genocide; (c) advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence; and (d) incitement to terrorism. The categorisation of this type of content indicates the high threshold set for restricting freedom of expression and that it may be necessary for companies to involve law enforcement to find the perpetrators.

2. "Expression that is not criminally punishable but may justify restriction and a civil suit..."

When outlining the conditions for restriction, in which La Rue acknowledges that blocking content is the most common method of restricting prohibited expression, La Rue says:

"States should provide full details regarding the necessity and justification for blocking a particular website, and determination of what content should be blocked should be undertaken by a competent judicial authority or a body which is independent of any political, commercial, or other unwarranted influences to ensure that blocking is not used as a means of censorship."¹⁰⁴

La Rue also references a previous report to the UN Human Rights Committee which affirms that any "restrictions on the operation of websites, blogs or any other Internet-based, electronic or other such information dissemination system, including systems to support such communication, such as Internet service providers or search engines, are only permissible to the extent that they are compatible with paragraph 3 [of Article 19]. Permissible restrictions generally should be content-specific; generic bans on the operation of certain

100 Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence (2012), para 15. Available at, http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf. See also, Concept Paper on OHCHR's Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred (2011): www.ohchr.org/Documents/Issues/Expression/ICCPR/IncitementWorkshopsConceptPaper.pdf

101 These examples are outlined in Para 51 of A/67/357 and are taken from examples outlined in documentation available from the documentation available from <http://www.ohchr.org/EN/Issues/FreedomOpinion/Articles19-20/Pages/ExpertsPapers.aspx>

102 UN General Assembly, Promotion and Protection of the Right to Freedom of Opinion and Expression. Note by the Secretary General. 10th August 2011, A/66/90, para 79. Available at, <http://www.ohchr.org/Documents/Issues/Opinion/A.66.290.pdf>

103 For this case study, IHRB is focusing on incitement, however La Rue also cites dissemination of child pornography as an example of "Exceptional types of expression that States are required to prohibit under international law" according to the Optional Protocol (2000) to the UN Convention on the Rights of the Child on the sale of children, child prostitution and child pornography as defined in Article 2(c).

104 A/66/90, para. 38.

sites and systems are not compatible with paragraph 3. It is also inconsistent with paragraph 3 to prohibit a site or an information dissemination system from publishing material solely on the basis that it may be critical of the government or the political social system espoused by the government.”¹⁰⁵

In terms of what may constitute a civil suit, La Rue stresses that “all other types of expression that are not mentioned above should not be criminalised, including defamation laws aimed at protecting the reputation of individuals, as criminalization can be counter-effective and the threat of harsh sanctions exert a significant chilling effect on the right to freedom of expression”.¹⁰⁶

3. “Expression that does not give rise to criminal or civil sanctions but does raise concerns in terms of tolerance, civility and respect for others.”¹⁰⁷

Here La Rue places importance on focusing efforts to tackle the “root causes of such expression, including intolerance, racism and bigotry by implementing strategies of prevention.” He continues, “To do so, and to bring about real changes in mindsets, perceptions and discourse, a broad set of policy measures are necessary, for example in the areas of intercultural dialogue or education for diversity, equality and justice and in strengthening freedom of expression and promoting a “culture of peace”. Indeed, the Special Rapporteur has previously stated that the strategic response to expressions deemed as offensive or intolerant is more speech: “more speech that educates about cultural differences; more speech that promotes diversity and understanding; more speech to empower and give voice to minorities and indigenous peoples, for example through the support of community media and their representation in mainstream media.”¹⁰⁸

How does this translate to actions companies should take? Experts have argued that content should only be removed as prescribed by law. Restricting content may not be the only option and as we have seen with the example of Facebook, it is possible to implement a ‘tiered response’ to content based on these three categories. Responses to content that falls in this last category can provide an opportunity for companies to be creative in their approach to responding appropriately. Facebook has demonstrated “social reporting” as an option for users to engage with each other, which aligns with La Rue’s call above for more speech that promotes diversity and understanding. In situations when a user is being systematically targeted with abuse or bullied, companies could support the victim by providing effective reporting mechanisms and respond to users who are being systematically targeted in order to provide support and minimize the impact on the user’s well being. Or reports of patterns of offensive or inflammatory speech could serve as an early warning mechanism to an outbreak of violence. But to evaluate the appropriate action, it is important to take into account the context of speech.

¹⁰⁵ A/HRC/17/27, para 70.

¹⁰⁶ A/66/90, para. 40.

¹⁰⁷ This also reflects Chapter II of the African Charter on Human and People’s Rights, Article 28 under Duties states, Every individual shall have the duty to respect and consider his fellow beings without discrimination and to maintain relations aimed at promoting, safeguarding and reinforcing mutual respect and tolerance.

¹⁰⁸ A/66/90, para. 40 and 41.

9. The Importance of Context

Dr. Mutuma Ruteere, The UN Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Racial Intolerance recently commented at the Kenya IGF:

“In terms of determining what kinds of speech in my view should be prohibited...you have to take into account the particular context. In some places there is a reason or justification to prohibit holding certain views or opinions. Depending on the history and context of different countries, that necessitates the kind of speech that should be prohibited.”¹⁰⁹

Each country has its own pressure points or limits to free speech and therefore may legitimately claim its own interpretation of what could be described as hate speech. What constitutes hate speech may be region or culture-specific, rooted in a country’s history and social make-up. For example, Kenya’s definition of hate speech concentrates purely on ethnic issues, with no mention of religion which can often take priority in other countries’ restrictions of freedom of expression in the form of strict blasphemy laws. What is defined as hate speech in one country may not be considered hate speech in another part of the world. Often national laws are enacted to deal with speech considered harmful. Europe has laws restricting glorification of the Nazis, and democracies in the developing world, such as India, also have laws in place to prevent dissemination of communal and religious hatred.

Companies that host user-generated content often practice ‘geographic blocking’, where content is blocked from view in the country that deems it illegal, but it remains visible in other countries. However, it is rarely that simple and there are many grey areas. In countries where laws are not precise, companies operating communication platforms, such as mobile operators offering SMS services and Internet access, and web-based companies offering social media and user-generated content platforms, have to form their own judgment about an issue that remains inadequately defined under international law and often more so in domestic law. This leads to one of two actions: blocking/removing content or not.

A common criticism leveled at companies hosting user-generated content is that they either do not remove enough content or they block too much content, either at the request of a government or non-state actors. It is not an easy task to draw the line on where freedom of expression ends and legitimate restrictions begin. Without clearly defined laws and policies, companies are in danger of restricting legitimate expression, for example views critical of a particular government or opinions that offend certain sections of society, if the government decides to call such opinions or views “hate speech”.

As we have seen, international human rights law and La Rue’s framework place restrictions on advocacy of hatred that constitutes discrimination, incitement to hostility and potentially violent acts, as outlined in Article 20 of the ICCPR. But what about other instances of speech that do not fall under the legal definition, yet are considered inflammatory especially when disseminated in times of heightened social tension?

Certain forms of speech have long been used to fan ethnic tensions in Kenya and other countries. Such language is used by men and women on the street all the way up to government and can create conditions that lead to violence and even genocide.¹¹⁰ Scholars have found that describing people as certain kinds of animals or insects (as during the Rwandan genocide, when Tutsis were constantly described as ‘cockroaches’ on the radio) de-humanises the potential victim. Kenyan human rights activist and former political prisoner, Koigi wa Wamwere, recounts in his

¹⁰⁹ Comments made during panel on hate speech and social media, chaired by IHRB, Nairobi 26th July 2013: bit.ly/kigf13webcast

¹¹⁰ Simon Adams, The Huffington Post, Killer SMS: Incitement and the Kenyan Elections, 26th February 2013: www.huffingtonpost.com/simon-adams/killer-sms-incitement-and_b_2759884.html

book *Negative Ethnicity*, that “any name provoking fear, stirring the desire to destroy and justifying the death of the other, is used: hyena, rat, snake, lice and cockroach.”¹¹¹

The Kriegler Commission stated in its report that several Kenyan FM radio stations incited ethnic animosity during the 2007/2008 elections, particularly during call-in programs:

“Words and phrases such as ‘settlers’, ‘let’s claim our land’, ‘people of the milk to cut grass’, ‘mongoose has come and stolen our chicken’, ‘madoadoa [spots]’ and ‘get rid of the weeds’ aired by Kass FM and songs such as ‘talking very badly about beasts from the west’....Luo stations also played a song, ‘the leadership of the baboons’, which vilified the Mount Kenya people.”¹¹²

While such examples may not fall under the restrictions laid out in the ICCPR, academics have noted this type of speech rises dramatically before an outbreak of mass violence and there have been efforts to test the direct correlation between such speech and subsequent acts, whatever the means are to communicate.¹¹³

The development of the Internet and the use of social media provides a platform that can amplify this kind of speech and attempts are underway to predict the likelihood of certain speech catalysing violence in certain situations, as can be seen through the work of Susan Benesch of the Dangerous Speech Project¹¹⁴ and Umati. In addition, the UN Rabat Plan of Action¹¹⁵ sets out a threshold test developed to assess the threat of expression to incite discrimination, hostility and violence; the freedom of expression organisation ARTICLE 19 has worked to define speech prohibited under Article 20 of the ICCPR.¹¹⁶

The importance of context signifies yet again the grey areas when it comes to freedom of expression. As outlined earlier regarding SMS keyword filtering, some companies or governments may create a list of words, a ‘blacklist’, and then use software to ‘filter’ the words, which blocks content containing these blacklisted words from being transmitted over the Internet. However, this method can block legitimate content and there are more and more examples of this method being increasingly unreliable, including an academic study dating back to 2002.¹¹⁷ Over-reliance on words in identifying hate speech, by documenting a list of words, does not give the full picture of the nature and extent of hate speech. Language and its usage change over time and are context-dependent, and classifying critical speech as hate speech complicates matters. From here, companies can explore further appropriate actions that extend beyond blocking or not blocking content.

111 Koigi Wa Wamwere (2003) *Negative Ethnicity: From Bias To Genocide*, Seven Stories Press, New York. P46

112 Kriegler Report on 2007 Kenyan Elections, p65 and 100: kenyastockholm.files.wordpress.com/2008/09/the_kriegler_report.pdf

113 See David Yanagizawa (2012) *Propaganda and Conflict: Theory and Evidence from the Rwandan Genocide*: www.hks.harvard.edu/fs/dyanagi/Research/RwandaDYD.pdf
The statistical study shows that killings were 65-77% higher in villages that received the Radio Television Libre de Milles Collines (RTL) signal. Two RTL executives were convicted of incitement to genocide in 2007. See also Koigi Wa Wamwere (2003) *Negative Ethnicity: From Bias To Genocide*, Seven Stories Press, New York. Pp103-105, which describes anonymous leaflets circulated in Nazi Germany propagating hatred against Jews. More recently, leaflets were circulated provoking ethnic hatred during the break up of Yugoslavia, during the violent end to apartheid in South Africa instigating violence between the Xhosa and Zulus and have also been circulated in Kenya warning certain communities to leave their homes or be killed.

114 See the contribution to the OHCHR initiative on incitement to national, racial or religious hatred by Susan Benesch, consultant to the United Nations Special Adviser on the Prevention of Genocide, 2011: www.ohchr.org/Documents/Issues/Expression/ICCPR/Other2011/SBenesch.doc

115 See OHCHR (2012) Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence, available at http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf and UN OHCHR, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/67/357, 7th September 2012, p8: daccess-dds-ny.un.org/doc/UNDOC/GEN/N12/501/25/PDF/N1250125.pdf?OpenElement

116 ARTICLE 19, Towards an interpretation of Article 20 of the ICCPR: Thresholds for the prohibition of incitement to hatred (Work in Progress). A study prepared for the regional expert meeting on article 20, Organized by the Office of the High Commissioner for Human Rights, Vienna, February 8-9, 2010: www.ohchr.org/Documents/Issues/Expression/ICCPR/Vienna/CRP7Callamard.pdf

117 Caroline R. Richardson, MD; Paul J. Resnick, PhD; Derek L. Hansen, BS; Holly A. Derry, MPH; Victoria J. Rideout, MA Does Pornography-Blocking Software Block Access to Health Information on the Internet? The Journal of American Medical Association, December 11th 2002. Available at, <http://jama.jamanetwork.com/article.aspx?articleid=195607> For further examples see, Joe Miller, BBC Technology Online, British Library Wi-Fi Service Blocks ‘Violent’ Hamlet, 13th August 2013: www.bbc.co.uk/news/technology-23680689;

10. Applying a Systematic Approach to Dealing with Hate Speech through the UN Guiding Principles on Business and Human Rights

The *UN Guiding Principles on Business and Human Rights*¹¹⁸ were developed by the former Special Representative of the UN Secretary General for Business and Human Rights, Harvard Professor John Ruggie, over the six years of his mandate from 2005-2011. Based on extensive research and consultations with representatives from government, business and civil society (including trade unions, NGOs and legal and academic experts) across all continents, they have gained broad acceptance and support in recent years.

The *Guiding Principles* were unanimously endorsed by the UN Human Rights Council in 2011 and are now the authoritative global reference point on business and human rights. They are based on the UN's "Protect, Respect and Remedy" Framework, which recognises the complementary but distinct roles of states and business in protecting and respecting human rights. The three pillars of the Framework are:

- **The state duty to protect** against human rights abuses by third parties, including businesses, through effective policies, legislation, regulations and adjudication;
- **The corporate responsibility to respect** human rights, meaning that companies should avoid infringing on the rights of others and address negative impacts with which they are involved; and
- **The need for greater access to effective remedy for victims** of business-related human rights abuses, through both judicial and non-judicial means.

There are six core elements of the corporate responsibility to respect human rights. The core elements are:

- **A human rights policy commitment:** the company's overarching, public commitment to respect human rights, and the processes for driving that commitment into the company's culture.
- **Human rights due diligence:** the set of on-going processes through which the company "knows and shows" that it is respecting human rights in practice. This involves:
 - **Assessing** actual and potential human rights impacts
 - **Integrating** the findings and acting to prevent or mitigate the impacts
 - **Tracking** how effectively impacts are addressed
 - **Communicating** how impacts are addressed
- **Remediation:** the processes through which the company actively engages in the remediation of impacts it has caused or contributed to.

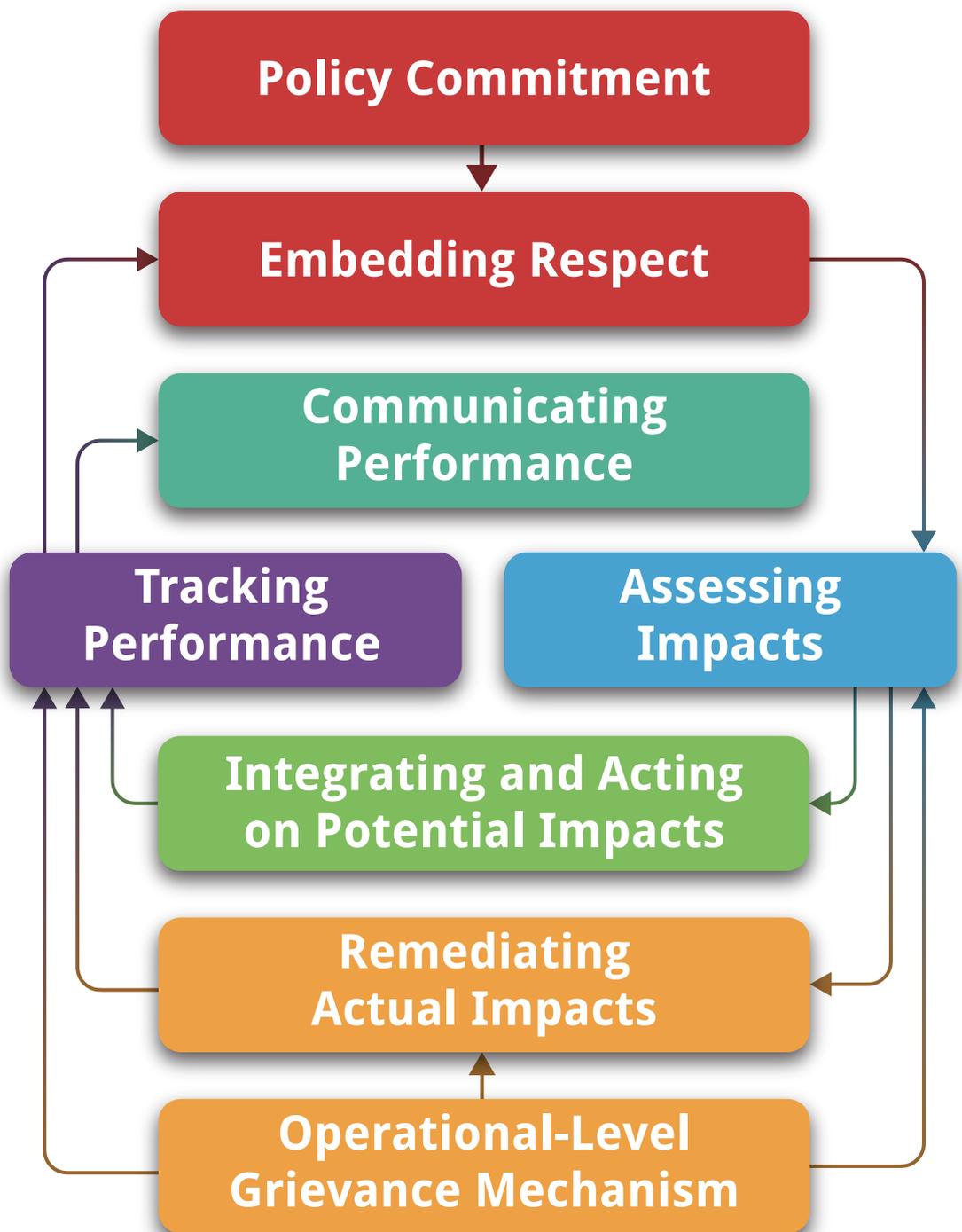
This chapter draws on the steps and recommendations outlined in the European Commission ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights, written by IHRB and Shift, and published in June 2013.¹¹⁹

Figure 1 (opposite) illustrates the relationship between the six core elements of the corporate responsibility to respect human rights.¹²⁰ This chapter outlines how ICT companies can apply the framework to the issue of tackling hate speech, applying them to the examples set by Safaricom and Facebook.

118 www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

119 Download the ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights: www.ihrb.org/pdf/eu-sector-guidance/EC-Guides/ICT/EC-Guide_ICT-13_Annexes.pdf

120 *Ibid*, p15



(i) Developing A Policy Commitment and Embedding Respect for Human Rights

A policy commitment demonstrates a general commitment to respect all internationally recognised human rights. If the commitment comes from a CEO and senior management, this sets the “tone at the top” and helps draw attention to the policy and embed it into the business culture.

This statement needs to be reflected in other company policies, procedures and practices in order to embed respect for human rights throughout the business. In the case of Facebook in the context of Kenya, questions concerning how a request to remove content perceived to contain hate speech would be made were evident. Therefore it is extremely important that companies have clear policies and mechanisms in place regarding reporting content and manage users’ expectations as to what will happen next and most importantly, respond to users raising concerns.

(ii) Embedding

It is important for the company to ensure that staff members, who have to make decisions about keeping or deleting specific messages, are properly trained in human rights law and international human rights standards. They are called upon to act as evaluators, if not judges, and their decisions may have far-reaching consequences on respect for human rights. It is important therefore that individuals making the decisions are aware of their responsibility and are therefore properly trained. Safaricom’s regulatory and legal teams involved in drafting their internal guidelines had completed human rights legal training, but other team members had not.

Engaging stakeholders is an important part of implementing the UN Guiding Principles at each step. When Safaricom were developing its internal guidelines, consulting with stakeholders before implementing blocking bulk SMS from abroad may have alerted it to possible implications for freedom of expression. Civil society organisations have raised this issue, as there is a concern it could spread to peer-to-peer communication such as emails.¹²¹ Blocking or putting in place a one-to-many filtering tool may also have consequences for organisations relying on this method to communicate and gather data. Organisations such as Ushahidi, which rely on SMS for crowdsourcing, are also concerned of the effect this policy may have on collecting data.

(iii) Identifying risks and potential impacts

Safaricom clearly identified that its SMS platform was at risk of being used as a channel for hate speech and acted to mitigate the risk, at a time when there was no legal requirement for the company to do so, and the government had not yet prepared its own plan. In this, Safaricom showed its foresight and its sense of responsibility.

In this case, Safaricom had the advantage of being a locally based company with expert knowledge of Kenyan culture and people. It can be harder for multinational mobile operators or companies offering web-based services to understand the social context in which they operate, but this is extremely important in order to fulfill the corporate responsibility to respect human rights, including the right to freedom of expression, by assessing potential risks to human rights that the company may cause or contribute to through its operations or that may be directly linked to the company. Further stakeholder consultation can assist international companies identify risk and establish the context in which they operate. Companies must understand the context in which they work, and numerous civil society reports can help.¹²²

¹²¹ WCIT meetings. Article 5b of ITRs: www.accessnow.org/blog/2012/12/14/wcit-watch-analysis-of-the-new-itrs-part-ii-

¹²² See annex on Key Resources and Key concepts from EC ICT Sector Guide on Implementing the Guiding Principles on Business and Human Rights: www.ihrb.org/pdf/eu-sector-guidance/EC-Guides/ICT/EC-Guide_ICT-13_Annexes.pdf

(iv) Prioritising impacts for action

In this case, preventing outbreaks of violence due to messages being spread via ICTs was clearly the priority. Based on Safaricom's knowledge and experience of the 2007 elections, it identified the factors that would enable hate speech to be spread via their SMS platform. These were:

- Anonymous sender;
- Language;
- Content of message;
- Unsolicited messages.

Safaricom could not interfere with private peer- to-peer messages and identified the source of the majority of hate speech coming from politicians, and concentrated on scrutinising bulk messages sent by politicians, political parties and aspirants to curb the spread of hate speech.

Companies can find guidance on the kind of content that may pose a risk by referring back to international law. The threshold for hate speech is high when tested against international law – advocating national, racial or religious hatred that constitutes incitement to discrimination, hostility and violence. As mentioned earlier in the study, Special Rapporteur La Rue has emphasised the need to distinguish between three different types of hate speech in his 2011 report¹²³:

- Those that are an offence under international law and can be criminally prosecuted;
- Expression that is not criminally punishable but may justify restriction and a civil suit, and;
- Expression that does not give rise to criminal or civil sanctions but does raise concerns in terms of tolerance, civility and respect for others.

As outlined earlier, these categories not only assist with prioritising impacts, but can also assist companies develop the most appropriate forms of action. The UK Crown Prosecution Service recently published Guidelines on Prosecuting Cases Involving Communications Sent Via Social Media¹²⁴, which may also be useful for companies in prioritising content that should be reported to law enforcement.

(v) Integrating and Acting

The grey area between international and national law means that companies are often left to judge for themselves the best course of action. This is difficult when faced with public or governmental pressure to act and it is often tempting for companies to cast the net wide and block offensive speech to avoid criticism or liability. Companies can use Frank La Rue's framework to formulate policies and build a tiered response in deciding on the best course of action. The debate around actions companies should take has largely focused on blocking content or not. The reality is that other options can be explored. This does not mean that any speech that does not qualify as a criminally prosecutable offence under international law should be ignored.

Companies that provide services for user-generated content could conceivably have a role in contributing to early warning mechanisms for outbreaks of mass violence by escalating to the senior management (and in turn to local or international agencies) reports of content that fall into these categories or patterns of speech in territories that have historically suffered as a result. It is therefore important that those responsible for receiving reports and mod-

¹²³ UN General Assembly, Promotion and Protection of the Right to Freedom of Opinion and Expression. Note by the Secretary General. 10th August 2011, A/66/90, para 79: www.ohchr.org/Documents/Issues/Opinion/A.66.290.pdf

¹²⁴ The Crown Prosecution Service (UK), Guidelines on Prosecuting Cases Involving Communications Sent via Social Media: www.cps.gov.uk/legal/a_to_c/communications_sent_via_social_media/

erating content understand the context in which they operate in order to not only remove content that falls under the legitimate restrictions in the ICCPR (of hatred that directly incites discrimination, hostility and violence) but also to recognise patterns of speech and content which could serve as an early warning mechanism of an outbreak of violence.

In Safaricom's case, a wide range of departments were consulted in the best course of action and in drafting the company's own guidelines, which in turn informed guidelines set by the Kenyan telecommunications regulator. In addition to the guidelines, Safaricom acted on another problem during the 2007 election: a severe shortage of pre-pay airtime credit, which is used by 99% of mobile phone subscribers in Kenya.¹²⁵ At the time, most subscribers topped up their airtime by buying 'scratch cards' from local agents and entering a code to access the amount of credit they had purchased. Deliveries of these scratch cards to local agents were restricted due to violence at the time, as suppliers were unable to deliver stock to local agents resulting in a shortage of credit, and some retailers began charging more than the face value for credit.¹²⁶ This added to the climate of fear, as customers were unable to check on their loved ones or alert each other of 'hotspot' areas. Safaricom had prepared adequately for any possible shortage of pre-pay credit by delivering extra scratch cards to local agents and enabling airtime to be purchased through M-PESA. Through M-PESA, a subscriber could also purchase airtime and allocate it to another subscriber.

(vi) Tracking performance

Safaricom has no plans to publish a final report on implementation of its Guidelines unless specifically required by CCK. A report was circulated by the Commercial Team internally which included the statistics mentioned in the previous chapter. This was important in order for employees to know that they contributed to a successful initiative to keep peace during the elections as a way to boost employee morale. In order to evaluate the success of the guidelines and identify areas where they could be improved or strengthened, some public reflection on the process would be useful for the sake of transparency, and it could be used as an example of best practice for other companies facing similar challenges.

(vii) Communicating

Clearer, straightforward communication and increased transparency of the company's actions, such as steps to block SMS from abroad, could have minimised some of the apparent public confusion that existed concerning Safaricom practices. Such efforts to engage proactive communication can help prevent companies being viewed with suspicion and reduce the likelihood the company is blamed when situations are no longer in their control. Companies should ensure the efforts they put in place to tackle certain forms of speech are transparent, as blocking content without due process could lead to "mission creep", where the criteria for blocking content is extended to legitimate forms of expression.

Likewise, during the Kenyan elections, the confusion around Facebook's policies, reporting mechanisms and lack of response to civil society concerns led to an impression that Facebook was not taking the situation seriously. In fact, there was much activity 'behind the scenes' where teams were monitoring the situation and put on high alert during the elections. Communicating this to Kenyans may have eased fears in this situation of heightened sensitivity.

125 Denis Mbuvi, CIO East Africa, Safaricom Donates 50 million SMS to Kenya Election Peace Campaign, February 7th 2013: www.cio.co.ke/news/main-stories/safaricom-donates-50-million-sms-to-kenya-election-peace-campaign

126 Kenyan Pundit blog post, Jan 1st 2008: www.kenyanpundit.com/2008/01/01/post-media-blackout-update-jan-1-1000-am/
Miwa Tsuda, Kenya after the 2007 'Post Election Violence': Constitutional Reform and the National Accord and Reconciliation Act. Institute of Developing Economies (2013), p3: www.ide.go.jp/English/Publish/Download/Dp/pdf/381.pdf

(viii) Working with Business Relationships

The Guiding Principles define “leverage” as the ability of a company “to effect change in the wrongful practices on an entity that causes harm”, or, in short, its ability to influence the behaviour of others. According to Safaricom, the Kenyan government considered shutting off the entire SMS network early on in the post election violence of 2007/2008 in an attempt to stem the flow of hate speech and curb the violence. States often resort to suspending networks in times of unrest, the most well known example being the government ordered national network disconnection during the Egyptian revolution in January 2011, which in this case was intended to disrupt legitimate protests. Safaricom and other mobile operators pushed back on the Kenyan government and instead convinced the government to use the network in a positive way to disseminate messages of peace and appeal for calm.¹²⁷ Shutting of the SMS network in a country so dependent on mobile communication such as Kenya could cause further panic and add to the climate of fear and would, for example, make it harder for those without banking facilities to undertake simple transactions. Although the government may have intended to stem the flow of hateful SMS, mobile operators were right in resisting the pressure and pushing back on this request and ensuring communications stayed open.

IHRB has long been concerned about the trend of governments asking/demanding that ICT companies block access to services in times of ‘national unrest’ or ‘emergency’ and companies’ complying without any clear guidelines as to where authority and responsibility reside. This example shows the trend beginning to be reversed, with companies promoting the positive aspects of ICT networks in appealing for calm, instead of letting their technology be used for harm. Network disconnection was reportedly not raised as an option at any point during the 2013 elections.

Telecommunications companies often argue that they are restricted from acting to prevent human rights abuses because of their position as a heavily regulated industry and their presence on the ground, which means they have local staff that may be put in danger. This study shows that such presence can be a strength given the inside knowledge of the challenges facing a country and the ability to identify where the ‘hotspots’ might be or how their products or service can be used to violate human rights and acting to mitigate that risk.

(ix) Issues of Context and Remediation

In situations of this nature, where national law conflicts with or is at variance from international standards, the UN Guiding Principles on Business and Human Rights state that business enterprises should: “Seek to honour the principles of internationally recognized human rights when faced with conflicting requirements.”¹²⁸ This means the high threshold for legitimate restrictions on freedom of expression should be reflected in a company’s terms and conditions. As mentioned earlier, this does not mean that companies can turn a blind eye to reports of other forms of speech, such as abuse, which should be considered in context and alternative responses explored.

127 The first peace SMS sent, as reported here by prominent blogger Juliana Rotich, blogging as ‘Afromusing’, in Eldoret, said:

“The ministry of Internal security urges you to please desist from sending or forwarding any SMS that may cause public unrest. This may lead to your prosecution.”
afromusing.com/2008/01/01/quick-update-from-eldoret-rift-valley/

CEO of Safaricom at the time, Michael Joseph, said in an interview that Safaricom then “toned it down a bit” and sent the following messages:

“In the interest of peace, we appeal to Kenyans to embrace each other in the spirit of patriotism and exercise restraint to restore calm to our nation.” They then added the tagline: “Prevent trouble, choose peace” and sent the messages in both national languages of Swahili and English.

Ofeibia Quist-Arcton, NPR, February 20th 2008. *Text Messages Used to Incite Violence in Kenya*: www.npr.org/templates/story/story.php?storyId=19188853

128 UN Guiding Principles on Business and Human Rights, 23 (b): www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

If content does not fall in the criteria of illegal speech under international human rights law, it does not mean it can be ignored. Thus, if a company is asked to block or remove content that does not fall within legitimate restrictions outlined in international law, a company could have in place other mechanisms to counter inflammatory or offensive speech. This could involve warning users, supporting victims and promoting positive counter speech.

(x) Continuing Stakeholder Engagement

Kenya has an active civil society dedicated to the preservation of freedom of expression; bloggers and groups such as Ushahidi and Mashada are active in this space to educate Kenyans on the impacts of online speech and encourage social media users to combat offensive and destructive speech with constructive speech. During the elections, civil society-led public initiatives countered hate speech, such as the Nipe Ukweli project, which encouraged people to tackle false rumours to prevent the escalation of tension and violence.¹²⁹ Expedient, short-term solutions such as blocking content will not help heal social and ethnic divisions and as Frank La Rue has stated, more speech is needed to promote diversity and understanding. ICT companies would benefit by engaging in this mission and supporting groups to do this work.

However, in the guise of decency, morality, or security, any group can prevent certain views from being heard. Companies like Safaricom and Facebook should listen to those voices, but not cede authority.

129 IHub Research, Nairobi, Nipe Ukweli: www.research.ihub.co.ke/uploads/2013/february/1361639967_819_689.pdf

Conclusion and Next Steps

The 2013 Presidential elections in Kenya passed peacefully, in part due to a shift in attitude and efforts on behalf of the government, business and the people to make it so.

This study has analysed the challenges in forming appropriate responses to hate speech, for companies providing a variety of services, from SMS to social media, operating domestically and internationally. Given the focus on one specific instance – Kenya – and the experiences of two companies in different contexts as well as lessons from elsewhere, its findings do not claim to represent similar situations everywhere. However, it is hoped that the insights and lessons of tactics deployed in this case will be of value to companies and governments, with civil society support, which aim to reduce the likelihood of spreading messages that incite violence or hatred in other contexts. For example, Safaricom’s decision to put in place mechanisms to vet political content may not be a viable option in countries where political opposition is known to be persecuted. Doing so in some circumstances could compound the problem. As outlined earlier, context is extremely important when dealing with restrictions on freedom of expression.

The rise of mobile Internet means more people will be connected and social media users will increase rapidly over the coming years. As the ICT sector converges, mobile network operators, which previously had little to do with providing content, may begin to expand services and provide extra features like apps and content and therefore may be called upon by governments or users to block content. Therefore, it is important to consider these challenges at the start of operations so that companies are prepared for situations that may arise quickly.

This study has highlighted the difficulty of tackling ‘hate speech’ in ICTs. Hate speech has taken prominence on the world stage and is certainly not restricted to Kenya, but each jurisdiction takes a different approach by enacting different laws to combat it, which may not be aligned with international standards. Companies operating in multiple jurisdictions often have to evaluate many different interpretations of the restrictions placed on freedom of expression. As this study has shown, there is no internationally agreed legal definition of hate speech, but in international law there is a high threshold for illegal speech.

When companies receive requests from governments and non-state actors to block or remove content, they must test the legality of speech against international law, consider the context of the country in question, and explore other solutions apart from blocking or removing content. Tackling the root cause of such speech and encouraging robust debate is ultimately the only way that harmful speech will begin to subside online.

Some questions remain around what actions the Kenyan government took in terms of blocking content during the most recent elections. In January 2013, research conducted by Citizen Lab¹³⁰ into Bluecoat’s Packetshaper appliance, which filters Internet traffic by content category, was discovered in 18 countries, including Kenya, although it is uncertain whether the device has been implemented.¹³¹ In March 2012, it was reported that the CCK had sent letters to telecoms service providers informing of plans to install Internet traffic monitoring equipment known as the Network Early Warning System (NEWS)¹³² to help detect “cyber threats” as part of an agreement with the ITU. Operators were reportedly opposed to the proposal due to privacy implications and again it is unclear whether this has been implemented.

130 Citizen Lab (2013) Planet Blue Coat: Mapping Global Censorship and Surveillance Tools, p25: citizenlab.org/wp-content/uploads/2013/01/Planet-Blue-Coat.pdf

131 Freedom House, Freedom on the Net 2013. Kenya: www.freedomhouse.org/report/freedom-net/2013/kenya

132 *Experts Criticise Kenya Email Spying*, ITWeb, 23 March, 2012: www.itweb.co.za/index.php?option=com_content&view=article&id=52888

This highlights the importance and urgency of addressing the challenge of hate speech. The consequences of not placing importance on this issue could result in the introduction of restrictive laws and measures, such as increased monitoring and surveillance of online communications in an attempt to combat hate speech or even offensive speech, which in turn may infringe on privacy and have a chilling effect on freedom of expression.¹³³

Members of Kenyan civil society are concerned that following the elections which passed relatively peacefully, the NCIC has moved away from combatting the root causes of hate speech and onto the next pressing issue, which is largely viewed as devolution. Civil society groups in Kenya feel that it is important to keep the conversation concerning hate speech and freedom of expression going, as there are still challenges to face and tensions to resolve. The worry is that if not, the same problems will present themselves at the next elections in around five years' time.

For Kenya, further dialogue is needed, including companies, civil society and governments so that each party can fully understand the context and challenges of combatting hate speech online and ask questions about the appropriate roles of all actors in society. Clear communication from companies is crucial so that others understand their positions and policies. Then, as a consequence of these actions, during the next Kenyan election, political debate can thrive without undue restrictions placed on freedom of expression and privacy.

133 CCK Sparks Row With Fresh Bid To Spy On Internet Users, Business Daily Africa, March 20th 2012:

www.businessdailyafrica.com/Corporate-News/CCK-sparks-row-with-fresh-bid-to-spy-on-Internet-users-/539550/1370218/-/item/0/-/edcfmsz/-/index.html

Notes

A series of horizontal dotted lines for writing notes.



Institute for
Human Rights and Business

Shaping Policy • Advancing Practice • Strengthening Accountability

ISBN 1-908405-06-7



9 781908 405067

Institute for Human Rights and Business
34b York Way
London N1 9AB
UK

Phone: (+44) 203-411-4333
Email: info@ihrb.org