UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A RIGHTS-RESPECTING APPROACH TO PREVENTING ONLINE HARMS, PROTECTING ONLINE EXPRESSION AND ENSURING EFFECTIVE PLATFORM GOVERNANCE IN NIGERIA AND SOUTH AFRICA

by

OLUWATOMIWA TIMOTHY ILORI

STUDENT NUMBER: U18362878

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Laws (LLD)
in the Faculty of Law, University of Pretoria

Supervisor: Professor Magnus Killander, Centre for Human Rights, University of Pretoria

Co-supervisor: Professor Jay Aronson, Department of History, Dietrich College of Humanities and Social Sciences, Carnegie Mellon University

MAY 2022

© University of Pretoria

DECLARATION

I declare that this thesis, 'A rights-respecting approach to preventing online harms, protecting online expression and ensuring effective platform governance in Nigeria and South Africa,' which I submit for the degree of Doctor of Laws (LLD) in the Faculty of Law, University of Pretoria, is my work and has not been previously submitted by me for a degree in the University of Pretoria or any other higher institution.

Name: Oluwatomiwa Timothy Ilori

Student no: u18362878

Date:

© University of Pretoria

DEDICATION

This one is for me.

© University of Pretoria

ACKNOWLEDGMENTS

This thesis is not done without thanking the people who have shared themselves with me through their kindness, patience and cheers.

To my mother, Yé'jẹbú, who taught me the virtue of fortitude for life, thank you for showing me all the dimensions of love.

To my supervisors, Professor Magnus Killander and Professor Jay Aronson, thank you for the rod, thank you for the staff. I am grateful for your firmness, patience and insights.

I would also like to thank all my examiners for their constructive feedback on how to improve the thesis.

Nelly, *nyagot*, *mpenzi wangu*, *Rafiki*, thank you.

Sam, the friend, thank you.

To the Oyedeles, Akintundes and the Alabis, thank you for all the sacrifices.

Professor Frans Viljoen, thank you for your kindness.

To my friends at the Centre for Human Rights, Dr Ayo Sogunro, Micheal Nyarko, Dr Trésor Makunya, Foluso Adegalu, Johannes Buabeng-Baidoo, Thiruna Naidoo, thank you for the gift of laughter.

To my colleagues at the Expression, Information and Digital Rights Unit, Hlengiwe Dube and Marystella Simiyu, thank you for making us work.

Dr Shyllon, thank you for making time and for your tough questions.

Adebiyi Olusolape, thank you. Yes, you don't run marathons like sprints.

Dr Adeola Adedapo, Ridwan Oloyede, Mayowa Ajileye, Dr Dami Ajayi, Kendy Mbatha, Sally Omotto, Aunt Mary, Lulu, Bonnie, thank you.

To Dr Dunmade, thank you for seeing the light at the end of each tunnel.

To everyone I cannot mention, it is the lack of space, not of gratitude. Thank you for being part of this journey with me.

© University of Pretoria

SUMMARY

This thesis examines how the prevention of online harms and protection of online expression can be carried out in Nigeria and South Africa through a rights-respecting approach. It uses postcolonial legal theory to argue that the concepts of expression in African indigenous societies are close enough to the normative principles that underpin the right to freedom of expression online. However, despite these conceptions, the right is unprotected in African countries and one of the reasons for this is the continuing impacts of colonial-era legal provisions on the right to freedom of expression. These provisions, found in criminal and penal codes of most former British colonies, provide for various offences like 'publication of false information,' 'sedition,' 'abuse or insulting language to religion or person,' 'criminal defamation' and others.

This thesis argues that these provisions have negatively influenced cybercrime and electronic communications laws in these countries. These colonial-era codes and laws make regulating online harms and protection of online expression on social media platforms particularly difficult. It argues that this difficulty may be attributed to a new form of digital colonialism, the continued use of colonial laws to violate expression on social media platforms in these African countries. It also argues that for these harms to be prevented and for online expression to be protected on social media platforms, it is necessary for lawmakers to replace these holdovers of colonial-era legal regimes with a rights-respecting approach to platform governance.

Focusing on Nigeria and South Africa, this thesis proposes that such an approach must be anchored to international human rights law, and its governance must be built on a dynamic regulatory matrix that allows for constant communication, openness and multistakeholderism. For actors in Nigeria and South Africa to prevent these harms and protect online expression, they must rethink their idea of governance starting with legal reforms. This includes repeal and amendment of relevant laws that violate the right to freedom of expression online and enactment of laws that prevent online harms and protect online expression. These laws must be creatively designed, normatively sound and generative in their processes.

For these laws to be creatively designed, proximate actors must include as many stakeholders as possible. This means that the rules that govern online expression must be driven by multistakeholderism, that is meaningful engagements among specific actors like governments, national human rights institutions, social media platforms, the United Nations and African Union human rights systems, international non-governmental organisations, local civil society and academia. This thesis also argues that for these laws to be normatively sound, they must be anchored to international human rights standards. This anchoring means that these laws must be rights-respecting. It also notes that the processes involved in coming up with new

© University of Pretoria

laws must be incremental – they should first be developed into soft laws before they are enacted as hard laws.

Using doctrinal analysis, this thesis shows that African countries have both legal obligations to protect online expression and the responsibility to prevent online harms like information disorder and targeted online violence. It shows that in order for such protection and prevention to be effective, the aforementioned specific stakeholders must collaborate in order to carry out the necessary legal reforms. It also notes that governments like those of Nigeria and South Africa have a fundamental obligation to govern social media platforms in a rights-respecting way. However, in order for this to work, these actors must commit to a generative process. This means that they must apply international human rights standards and multistakeholderism to first shape social media charters as soft laws and later online harms acts or laws as hard laws in order to prevent online harms and protect online expression in Nigeria and South Africa.

This thesis makes four major original contributions to conversations on the prevention of online harms, protection of online expression and effective platform governance. One, it connects the concept of expression in African indigenous societies to the development of the right to freedom of expression online. It also traces the history of problematic laws on online expression, which exacerbate online harms in some African countries, to colonial legal provisions. Two, it examines the impacts of these harms on the right to freedom of expression online in African contexts. Three, it proffers a model that could be used to ensure a rights-respecting approach to combatting online harms. Four, it identifies the roles of specific actors in ensuring such an approach.

KEYWORDS: content moderation, cybercrime, digital colonialism, human rights, information disorder, legal reform, multistakeholder, Nigeria, online expression, online harm, online violence, platform governance, rights-respecting, social media, South Africa.

© University of Pretoria

LIST OF ABBREVIATIONS

AU	African Union
BSR	Business for Social Responsibility
CDA	Communications Decency Act of 1996
CGI	Computer-Generated Imagery
CRA	Child Rights Act of 2003
CSAM	Child Sexual Abuse Material
DVA	Domestic Violence Act of 1998
ECOWAS	Economic Community for West African States
ECTA	Electronic Communications and Transactions Act of 2002
FPB	Films and Publications Board
GIFCT	Global Internet Forum to Counter Terrorism
GNI	Global Network Initiative
HRIAs	Human Rights Impact Assessments
ICASA	Independent Communications Authority of South Africa
ICCPR	International Covenant on Civil and Political Rights
ICERD	International Convention on the Elimination of Racial Discrimination
ICESCR	International Covenant on Economic, Social and Cultural Rights
IGF	Internet Governance Forum
ISP	Internet Service Providers
ISPA	Internet Service Providers' Association
ITU	International Telecommunications Union
NCC	Nigeria Communications Act of 2003
NCPHS	National Commission for the Prohibition of Hate Speeches
NGOs	Non-Governmental Organisations
NHRC	National Human Rights Commission
NHRIs	National Human Rights Institutions
NSRP	Nigeria Stability and Reconciliation Programme
OAU	Organisation of African Unity
PEPUDA	Promotion of Equality and Prevention of Unfair Discrimination, 2000

© University of Pretoria

PIFM	Protection from Internet Falsehoods and Manipulation
SAHRC	South African Human Rights Commission
SMChs	Social Media Charters
SMCs	Social Media Councils
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNGPs	United Nations Guiding Principles on Business and Human Rights
UNSP	United Nations Special Rapporteur for Freedom of Opinion and Expression
WIPO	World Intellectual Property Organisation
WSIS	World Summit on the Information Society

© University of Pretoria

CONTENTS

© University of Pretoria

© University of Pretoria

© University of Pretoria

© University of Pretoria

© University of Pretoria

© University of Pretoria

© University of Pretoria

CHAPTER ONE: INTRODUCTION

1.1 Introduction

This thesis examines a rights-respecting approach to preventing online harms, protecting online expression and ensuring effective platform governance in Nigeria and South Africa. It identifies the roles of specific actors in ensuring such a rights-respecting approach. In examining the challenges facing the governance of online speech in African countries, this thesis adopts postcolonial legal theory to interrogate the sources of various legal provisions on online expression. This thesis argues that the legacies of colonial legal provisions negatively impact contemporary cyber laws, which seek to limit the right to freedom of expression online in many African countries.

The linear relationship between these cyber laws and colonial legal provisions has made preventing online harms difficult.[1] In addition, it has become even more challenging to prevent these harms without adopting a rights-based approach.[2] The significant contribution of this thesis is to identify how to prevent online harms while the right to freedom of expression online is also protected on social media platforms. It further highlights the roles of state and non-state actors in Nigeria and South Africa in ensuring this approach.

In examining its central focus, this thesis makes four major arguments. The first argument is twofold. The first part highlights the connection between African indigenous systems and online expression. The second part shows that the right to freedom of expression online is not adequately legally protected in many African national contexts due to legal provisions dating back to the colonial-era. It further argues that this lack of protection exacerbates online harms on social media platforms. The second argument is that these harms pose threats to online expression in African national contexts. These harms, which include information disorder and targeted online violence, are being weaponised by various actors to violate online expression in African countries. The third argument is that the most feasible means of preventing online harms and protecting online expression on social media platforms in African

[1] United Nations General Assembly 'Disinformation and freedom of opinion and expression: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/HRC/47/25' 13 April 2021 http://undocs.org/en/A/HRC/47/25 (accessed 26 August 2021) paras 52 & 53.

[2] See MK Land 'Regulating private harms online: Content regulation under human rights law' in RF Jørgensen (ed) *Human rights in the age of platforms* (2019) 287-310; E Douek 'The limits of international law in content moderation' (2021) 6 *UC Irvine Journal of International, Transnational, and Comparative Law* 72; EM Aswad 'The future of freedom of expression online' (2018) 17 *Technology Review* 45; D Kaye *Speech police: The global struggle to govern the Internet* (2018) 88; J Andrew 'Introduction' in J Andrew & F Bernard (eds) *Human rights responsibilities in the digital age: States, companies and individuals* (2021) 1-23.

© University of Pretoria

contexts must be through carrying out legal reforms such as the repeal or amendment of applicable laws and the enactment of better laws. However, for these reforms to be effective, international human rights standards of protecting online expression must apply. The fourth argument, using Nigeria and South Africa as case studies, is that specific actors have clearly identifiable roles that are necessary in ensuring an effective rights-respecting approach. These specific actors include internal and external as well as state and non-state actors. Some of them include governments, National Human Rights Institutions (NHRIs), social media platforms, the United Nations (UN) and the African Union (AU) human rights systems, international non-governmental organisations (NGOs), local civil society and academia.

1.2 Background

Platform governance is problematic.[3] No person, institution or government has been able to respond adequately to the complex challenges facing governing online expression.[4] Today, the regulation of online speech has moved from the strict regulation of governments and now lies in the hands of 'new governors.'[5] These new governors, who are majorly social media platforms, have a global reach that significantly influences online expression everywhere.[6] This influence lies in how social media platforms now shape permissible expression and communication online.[7] However, governing this influence has been particularly difficult.

This particular difficulty is due to many reasons. One of them is the complexity of interests and actors involved in preventing online harms while also protecting online expression. These harms include information disorder and targeted online violence that are mainly perpetrated on social media platforms, which manifest differently in different societies based on various factors.[8] For example, governments have passed

[3] T Gillespie *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media* (2018) 9.

[4] See A Rochefort 'Regulating social media platforms: A comparative policy analysis' (2020) 25 *International and Comparative Perspectives on Communication Law* 225-260; A Callamard 'The human rights obligations of non-state actors' in RF Jørgensen (ed) *Human rights in the age of platforms* (2019) 191-218; See MK Land 'Against privatised censorship: Proposals for responsible delegation' (2019) 60 *Virginia Journal of International Law* 2020.

[5] See K Klonick 'The new governors: The people, rules and processes governing online speech' (2018) 131 *Harvard Law Review* 1603.

[6] E Donahoe & FE Hampson 'Governance innovation for a connected world protecting free expression, diversity and civic engagement in the global digital ecosystem' (2018) *Centre for International Governance Innovation: Special Report* 11 https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/stanford_special_report_web.pdf (accessed 15 August 2020).

[7] See N Persily & JA Tucker 'Introduction' in N Persily & JA Tucker (eds) *Social media and democracy: The state of the field and prospects for reform* (2020) 1-9.

[8] United Nations General Assembly 'Contemporary challenges on freedom of expression: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/71/373' 6 September 2016 http://undocs.org/en/A/71/373 (accessed 26 August 2021) paras 9-49.

© University of Pretoria

laws aimed at combating these harms but these laws violate the right to freedom of expression instead.[9] Some of these violations include arbitrary arrests of government critics, journalists and human rights defenders.[10] In some instances, these laws have also been cited as the reason for banning social media platforms and online websites in some African countries.[11] Social media platforms also defer to these laws as standards when dealing with moderating expressions on their platforms, which makes protection of online expression more difficult.[12] Treaty-monitoring mechanisms like the Special Rapporteurs on the Right to Opinion and Expression and Civil Society have constantly pushed back against these two powerful interests to safeguard online expression and human rights in general.[13] Given this background, preventing these harms and protecting online expression in any context will not be easy.

The right to freedom of expression online, especially on social media platforms has assumed new contexts and forms, thereby opening up spaces, ideas and people to newer meanings and debates.[14] However, online harms pose unique challenges to the

[9] United Nations General Assembly (n 1 above).

[10] A Sugow *et al* 'Appraising the impact of Kenya's cyber-harassment law on the freedom of expression' (2021) 1 *Journal of Intellectual Property and Information Law* 91-114. The paper argued that vague words in Kenya's Computer Misuse Act might be used to violate the rights to freedom of expression online as it is being done in Nigeria and Uganda; R Kakungulu-Mayambala & S Rukundo 'Digital activism and free expression in Uganda' (2019) 19 *African Human Rights Law Journal* 167-192; Media Defence 'Mapping digital rights and online freedom of expression litigation in East, West and Southern Africa' 1 October 2021 https://www.mediadefence.org/resource-hub/wp-content/uploads/sites/3/2021/08/Media-Defence-Mapping-digital-rights.pdf (accessed 30 October 2021).

[11] G De Gregorio & N Stremlau 'Internet shutdowns and the limits of the law' (2020) 14 *International Journal of Communications* 4224-4243; E Marchant & N Stremlau 'The changing landscape of internet shutdowns in Africa' (2020) 14 *International Journal of Communications* 4216-4220.

[12] Twitter 'About country withheld content' https://help.twitter.com/en/rules-and-policies/tweet-withheld-by-country (accessed 12 February 2020); Facebook 'Government request to remove content' https://transparencyreport.google.com/government-removals/overview?hl=en (accessed 13 February 2020).

[13] United Nations General Assembly 'Online content regulation and freedom of opinion and expression: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/HRC/38/35' 6 April 2018 http://undocs.org/en/A/HRC/38/35 (accessed 15 October 2021); United Nations General Assembly 'Gender justice and freedom of opinion and expression: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/76/258' 30 July 2021 http://undocs.org/en/A/76/258 (accessed 26 August 2021); Provisions of the Declaration on Principles of Freedom of Expression and Access to Information Online, 2019; African Declaration on Internet Rights and Freedoms (2014) https://africaninternetrights.org/sites/default/files/African-Declaration-English-FINAL.pdf (accessed 15 October 2020).

[14] In addressing the challenges free speech faces in the age of new technologies, Sunstein argues for five forms of regulation: indirect regulation through disclosure requirements; self-regulation; publicly subsidised websites; content labelling; and content flagging. See CR Sunstein 'The future of free speech' in LC Bollinger & GR Stone (eds) *Eternally vigilant: Free speech in the modern era* (2002) 285; DA Strauss 'Freedom of expression and the common-law Constitution' in LC Bollinger & GR Stone

© University of Pretoria

right to freedom of expression online due to problematic laws and policies steeped in colonial-era provisions.[15] These provisions, which focus on the criminalisation of false information, criminalisation of abusive and insulting language, sedition, and criminal defamation, have various influences on cyber laws and policies that seek to regulate online harms. These provisions underpin the relationship between colonialism, online harms and the need to protect the right to freedom of expression online in African countries. This thesis examines these issues and proffers a human rights-based approach to platform governance for state and non-state actors in Nigeria and South Africa.

1.3 Problem statement

Debates on effective means of governing social media platforms are relatively recent.[16] These debates have become topical due to the impact of online harms on the global democratic decline across the world and vulnerable persons who are adversely affected by these harms.[17] The prevention of these harms, especially information disorder and targeted online violence, has been a challenge for critical stakeholders, including governments and social media platforms.[18] Furthermore, examination of these harms, their impact on the right to freedom of expression online and the need for an effective governance model is not only limited; they are yet to receive any serious academic attention in the African context.

(eds) *Eternally vigilant: Free speech in the modern era* (2002) 33; E Bell 'The unintentional press: How technology companies fail us as publishers' in LC Bollinger & GR Stone (eds) *The free speech century* (2010) 235; M Bickert 'Defining the boundaries of free speech on social media' in LC Bollinger & GR Stone (eds) *The free speech century* (2010) 254; RF Jørgensen 'Human rights and private actors in the online domain' in MK Land & J Aronson (eds) *New technologies for human rights law and practice* (2018) 243.

[15] H Hannum 'Reinvigorating human rights for the twenty-first century' (2016) 16 *Human Rights Law Review* 439; MF Rice 'Information and communication technologies and the global digital divide' (2003) 1 *Comparative Technology Transfer and Society* 74; Land's approach to new technologies through international law focuses on the movement of the UN and its agencies towards the realisation of the implications of new technologies on human rights. See MK Land 'Towards an international law of the Internet' (2013) 54 *Harvard International Law Journal* 393-458; O Spijkers 'The United Nations, the evolution of global values and international law' (2011) 47 *School of Human Rights Research Series* 13-57.

[16] See E Bietti 'A genealogy of digital platform regulation' 3 June 2021 https://bit.ly/3j5YX0W (accessed 15 October 2021).

[17] D O'Connor & M Schruers 'Against platform regulation' September 2016 http://blogs.oii.ox.ac.uk/ipp-conference/sites/ipp/files/documents/OConnor-Schruers%2520-%2520Against%2520Platform%2520Regulation.pdf (accessed 15 October 2021).

[18] Digital Act 'Online harms white paper: Seven expert perspectives' 8 April 2019 https://www.politico.eu/wp-content/uploads/2019/04/Seven-expert-perspectives-on-the-UK-online-harms-White-Paper-.pdf?utm_source=POLITICO.EU&utm_campaign=723cb52285-EMAIL_CAMPAIGN_2019_04_10_05_07&utm_medium=email&utm_term=0_10959edeb5-723cb52285-189780761 (accessed 15 October 2021).

© University of Pretoria

One of the essential principles of protecting democracies is the respect for fundamental human rights, one of which is the right to freedom of expression online in the digital age.[19] Among other reasons, the increasing use of social media platforms in Africa has also necessitated the protection of this right online. Some of these uses include commercial activities, participation in public debates, demanding accountability from duty-bearers, amplifying social justice causes and political mobilisation.[20] External state actors, like treaty-monitoring bodies, have provided guidance for African states to protect the right online.[21] However, despite this guidance, African countries still struggle with protecting the right.[22] This struggle is due to anti-free speech provisions in laws dating back to the colonial era and cyber laws on the one part, and the complexity of governing social media platforms to protect online speech on the other part. This thesis considers the challenges posed by these laws and the complexity of governing these platforms to protect online expression in Nigeria and South Africa. Primarily, this thesis tackles the challenges posed by online harms in the African context and how various actors can prevent these harms and promote online expression through a rights-respecting approach.

1.4 Research questions

The primary question this research seeks to respond to is: How can a rights-respecting approach be applied in Nigeria and South Africa to ensure the prevention of online harms and protection of online expression on social media platforms? In responding to the question above, the thesis seeks to proffer answers to the following sub-questions:

a. To what extent have the theoretical perspectives and recent normative developments on protecting the right to freedom to expression online been implemented in African countries?
b. What are the impacts of online harms on the right to freedom of expression online in African countries?
c. How can rights-respecting platform governance be used to prevent online harms and protect freedom of expression online on social media platforms in African countries?

---

[19] D Grimm 'Freedom of speech in a globalised world' in I Hare & J Weinstein *Extreme speech and democracy* (2009) 11.

[20] A Olojo & K Allen 'Social media and the state: Challenging the rules of engagement' 24 June 2021 *Institute for Security Studies* <https://issafrica.org/iss-today/social-media-and-the-state-challenging-the-rules-of-engagement> (accessed 16 November 2021); A Okunola & K Mlaba 'From #EndSARS to #AmINext: How young Africans used social media to drive change in 2020' 23 December 2020 *Global Citizen* https://www.globalcitizen.org/en/content/endsars-aminext-young-african-social-movement-2020/ (accessed 16 November 2021); M Dwyer & T Molony 'How social media is changing politics in Africa' 23 February 2021 *Democracy in Africa* https://democracyinafrica.org/socialmedia/ (accessed 16 November 2021).

[21] United Nations General Assembly (n 13 above).

[22] O'Connor & Schruers (n 17 above).

© University of Pretoria

d. What are the roles of state and non-state actors, internal and external, in ensuring a practical rights-based approach to social media platform governance in Nigeria and South Africa?

1.5 Definition of key terms

This thesis uses some key terms and they are defined as follows.

Rights-respecting approach: This term is used to mean the application of international human rights standards to the regulation of online expression. In the context of this thesis, a rights-respecting approach means the use of these standards to prevent online harms and protect online expression as used by various scholars.[23] It is also used to denote a rights-based approach or principles such as participation, accountability and transparency. It is used interchangeably with international human rights standards.

Digital colonialism: This term is used to show the impacts of colonial legal provisions on online expression in African countries. Digital colonialism is used in this thesis to illustrate one of the effects of colonialism on the protection of online expression. Just as traditional forms of expression were violated under these colonial legal provisions before the digital era, its online form is also under threat in African countries. It is used to highlight the linear relationship between colonialism, online harms and online expression in African countries.

Online harms: The term online harms[24] is used in this thesis to explain two major forms of harm that may occur as a result of electronic communications. These two major forms are information disorder and targeted online violence.

Information disorder: This term is used in the context of this thesis as the manipulation of electronic communication in such a manner that the true facts of such communication are misconstrued or misunderstood. Examples of information disorder include misinformation (non-intentional misrepresentation of facts), disinformation

[23] B Sanders 'Freedom of expression in the age of online platforms: The promise and pitfalls of a human rights-based approach to content moderation' (2020) 43 *Fordham International Law Journal* 955-1004; D Sive & A Price 'Regulating expressions on social media' (2019) 136 *South African Law Journal* 51-83; TD Oliva 'Content moderation technologies: Applying human rights standards to protect freedom of expression' (2020) *Human Rights Law Review* 607-640; K Gill 'Regulating platforms' invisible hand: Content moderation policies and processes' (2020) 21 *Wake Forest Journal of Business and Intellectual Property Law* 173-212; RJ. Hamilton 'Governing the global public square' (2021) 62 *Harvard International Law Journal* 117-174.

[24] L Belli & N Zingales *Glossary of platform law and policy terms* https://cyberbrics.info/wp-content/uploads/2020/11/Glossary-on-Platform-Law-and-Policy-CONSOLIDATED17472-1.pdf 106 (accessed 10 November 2020).

© University of Pretoria

(intentional misrepresentation of facts) and malinformation (suppression or projection of parts of a communication).[25]

Targeted online violence: This refers to the use of electronic communications to cause emotional, psychological and verbal abuse to another. Targeted online violence is used in this thesis to show how these forms of communications are used to target vulnerable persons or groups. Its examples include cyberstalking, cyberbullying, cyberagression, online gender-based violence, online violence against children and online hate speech.[26]

Online expression: Online expression is used as a term in this thesis to denote the right to freedom of expression online as protected under international human rights law.[27]

Platform governance: Platform governance, as used in this thesis, is the regulation of social media platforms through soft and hard laws.[28] It is used interchangeably with social media platform governance.

Generativity: This term is used in this thesis to show how to apply a rights-respecting governance. Generativity means that in ensuring that international human rights law is effectively applied to the prevention of online harms and the protection of online expression, actors' approach should be incremental i.e. they should adopt soft laws they can learn from before adopting hard laws that can further secure stakeholders' commitments.[29] It is also used interchangeably with generative model and approach to platform governance.

1.6 Objectives

This thesis aims to examine the prevention of online harms and protection of online expression through a rights-respecting approach. It seeks to achieve this aim in two ways. The first way is to analyse the protection of the right to freedom of expression online, identify the challenges posed by the impacts of online harms to the right and consider the approach that would best govern them in African countries. It seeks to provide analyses on how to tackle these dynamic, complex and obstinate questions facing the protection of online expression today by focusing on two country contexts, Nigeria and South Africa.

[25] See section 3.3.1 below.
[26] See section 3.3.2 below.
[27] See section 2.4 below.
[28] See section 5.3. below.
[29] See sections 4.6, 5.3 and 6.4 below.

© University of Pretoria

The second way is to contribute to the recent debates on social media platform governance from an African perspective. While the debates on how to govern social media platforms are recent, they have been focused mainly on Western systems. In addition to this limited focus, there is a need to contribute to the debate on how best to govern these platforms to prevent online harms and promote online expression in African contexts. Therefore, this thesis makes contributions on how to prevent online harms and protect online expression on social media platforms in Africa, also by focusing on the contexts in Nigeria and South Africa.

1.7 Methodology

This thesis adopts a doctrinal legal methodology by providing analyses on the protection of the right to freedom of expression online in African countries. The rationale for this methodology is that it assists in interrogating the role of law as a tool for rights-respecting solutions to platform governance in African countries. Where possible, this thesis also considers the various perspectives on freedom of expression through secondary sources. The thesis adopts the systematic method of analysing domestic and international laws on how they impact the prevention of online harms and the protection of online expression in these countries. It examines the right to freedom of expression as a stand-alone right and the challenges of protecting it today in Nigeria and South Africa.

1.8 Literature review

Major conversations on how to prevent online harms are reasonably new.[30] When these conversations are contextual, they are Global North-facing, for example focused more on the United States, European Union and other Western systems.[31] There are hardly any engagements on how these harms impact the right to freedom of expression online and the various contextual challenges faced in non-Western contexts. In addition to this challenge, there are limited conversations on the prevention of these harms, the protection of online expression and the roles proximate actors should play in ensuring these, especially from an African perspective.

For example, the right to freedom of expression has been identified as one of the cornerstones of most organised societies.[32] Before colonial conquests in African countries, indigenous systems understood the importance of expression in building communities.[33] These systems range from the Ga-Dangme, Akans, Somalis, the Igbos

[30] O'Connor & Schruers (n 17 above); B Zankova & V Dimitrov 'Social media regulation: Models and proposals' (2021) 10 *Journalism and Mass Communication* 75-58; See R Gorwa 'What is platform governance?' (2019) 22 *Information, Communication & Society* 2.
[31] As above.
[32] See G *Ayittey Indigenous African institutions* (2006).
[33] As above.

© University of Pretoria

in indigenous African societies who understood the importance of expression. Records show that these indigenous societies were so organised that colonialists had to co-opt their existing structures to which free expression contributed.[34] Despite this understanding of expression by African indigenous societies, the foundations of the right, especially as formulated under international human rights law, is often claimed by Western liberal theorists.[35]

Donnelly, Franck and others have noted that human rights development cannot be the sole preserve of any culture.[36] Viljoen and Murray have also analysed the various contributions of the African human rights system to the development of international human rights law.[37] However, no works have dealt with the specific conception of expression in African indigenous societies and its relationship with the development of the right to freedom of expression online. This relationship is necessary to draw traditional and normative justifications for the right to freedom of expression online in African countries. This necessity is because the right to freedom of expression online in some African countries is under threat as they are home to various laws that violate the right to freedom of expression. One of the reasons for this threat is the impact of old colonial laws that exacerbate online harms and violate online expression in these countries.

Arewa has noted that even though law sometimes plays catch up with regulating technologies, there is an often-overlooked concern – colonial legacies that are not limited to law.[38] Her work further focuses on such concerns, including how colonial legacies that have breathed into new laws impede technological development in Nigeria and other African countries. She writes:

> Colonial legal relics are evident in laws, legal approaches and legal interpretations that might arise in contexts where customary law and English law conflict. Colonial legal relics are relevant to the past and have significant implications for digital-era participation, particularly because digital economy policies and laws relating to business, entrepreneurship, and motivation in many countries in Africa continue to lag behind those of other countries areas of the world.[39]

[34] M Lechler & L MacNamee 'Indirect colonial rule undermines support for democracy: Evidence from a natural experiment from Namibia' (2018) 51 *Comparative Political Studies* 1861.
[35] See P Fitzpatrick & E Darian-Smith 'Laws of the postcolonial: An insistent introduction' in P Fitzpatrick & E Darian-Smith (eds) *Laws of the postcolonial* (1999) 249.
[36] J Donnelly *Universal human rights: Theory and practice* (2003) 62; TM Franck 'Is personal freedom a Western value' (1997) 91 *American Journal of International Law* 595 & 625.
[37] See F Viljoen 'Africa's contribution to the development of international human rights and humanitarian law' (2001) 1 *African Human Rights Law Journal* 19; R Murray 'International human rights: Neglect of perspectives from African institutions' (2006) 55 *The International and Comparative Law Quarterly* 193-204.
[38] OB Arewa *Disrupting Africa: Technology, law and development* (2021) 157.
[39] Arewa (n 38 above) 152.

© University of Pretoria

Since colonialism has many after-effects and is not limited to law, it means it also affects law.[40] However, how it affects law and specifically the right to freedom of expression online from an African perspective has not received enough attention.

This issue is more pronounced given new developments on protecting the right to freedom of expression online at the international level.[41] Online harms now harm vulnerable groups, influence elections and drive 'infodemics.'[42] The impacts of harm are no longer limited to physical violence as they can now be virtual in addition to resulting in offline violence.[43] Online harms like misinformation, disinformation, propaganda, cyberstalking, cyberbullying, cyberharassment, online gender-based violence, online violence against children and online hate speech now impact online expression in African countries. For example, in Ethiopia, online hate speech has precipitated violence offline, while in Uganda, women in politics have been victims of online gender-based violence.[44] In Kenya, organised information disorder has become more pronounced.[45] In some cases, governments have claimed to shut access to the Internet or social media platforms precisely due to the impacts of these harms on their elections or national security.[46] While these are debatable claims, especially on the

[40] Arewa (n 38 above) 164; J Clarke 'Law and race: The position of indigenous people' in S Bottomley & S Parker (eds) *Law in context* (1997) 231, 275 & 246; See M Hawkins *Social darwinism in European and American thoughts 1860-1945: Nature as model and nature as threat* (1997) 185.

[41] United Nations General Assembly (n 8 above).

[42] E Douek 'Governing online speech: From "posts-as-Trumps" to proportionality and probability' (2021) 121 *Columbia Law Review* 803; M Kivuva 'Online violence in times of COVID-19' 29 May 2020 *KICTANET* https://www.kictanet.or.ke/online-violence-in-times-of-covid-19/ (accessed 15 October 2021); G Achieng 'How harassment keeps women politicians offline in Uganda' 1 September 2021 *restofworld* https://restofworld.org/2021/women-politics-social-media-uganda/ (accessed 19 October 2021); N Iyer *et al* 'Alternate realities, alternate internets: African feminist research for a feminist Internet' (2020) *Pollicy* 10 https://www.apc.org/sites/default/files/Report_FINAL.pdf (accessed 15 October 2020); See E Pauwels 'The anatomy of information disorders in Africa' July 2020 *Konrad-Adenauer-Stiftung* https://www.kas.de/documents/273004/10032527/Report+-+The+Anatomy+of+Information+Disorders+in+Africa.pdf/787cfd74-db72-670e-29c0-415cd4c13936?version=1.0&t=1599674493990 (accessed 15 October 2021).

[43] T Ilori 'Facebook's censorship of the #EndSARS protests shows the price of its content moderation errors' 27 October 2020 *Slate* https://slate.com/technology/2020/10/facebook-instagram-endsars-protests-nigeria.html (accessed 13 March 2021).

[44] YE Ayalew 'Uprooting hate speech: The challenging task of content moderation in Ethiopia' *Centre for International Media Assistance* (CIMA) 27 April 2021 https://www.cima.ned.org/blog/uprooting-hate-speech-the-challenging-task-of-content-moderation-in-ethiopia/ (accessed 15 October 2021); Achieng n 42 above).

[45] O Madung & B Obilo 'Inside the shadowy world of disinformation for hire in Kenya' 2 September 2021 *Mozilla* https://foundation.mozilla.org/en/blog/fellow-research-inside-the-shadowy-world-of-disinformation-for-hire-in-kenya/ (accessed 15 October 2021).

[46] J Campbell 'Nigerian President Buhari clashes with Twitter Chief Executive Dorsey' 8 July 2021 *Council on Foreign Relations* https://www.cfr.org/blog/nigerian-president-buhari-clashes-twitter-chief-executive-dorsey (accessed 16 July 2021); Al Jazeera 'Chad slows down internet to curb hate speech on social media' 4 August 2020 https://www.aljazeera.com/news/2020/8/4/chad-slows-down-internet-

© University of Pretoria

proportionality of governments' response through these shutdowns, the fact remains that online harms negatively impact human rights across the world.

Therefore, stemming these harms and, at the same time, protecting online expression has become a considerable challenge even for international human rights law. This challenge is because what might constitute online hate speech in Ghana might not be the same in Nigeria, just as disinformation is often tricky to spot due to various factors.[47] The dynamic nature of these harms and the roles of actors in regulating them have formed the biggest challenge for protecting online speech worldwide. So far, special procedures of both the UN and the AU human rights systems, like the Special Rapporteurs on the Right to Freedom of Opinion and Expression have identified these harms and proposed ways through which governments can prevent them while also protecting online harms.[48]

The cross-cutting recommendations made to states by these systems are for states to only criminalise expression in ways that comply with international human rights law, in order to effectively combat these harms and also promote online expression. The recommendations also focus on states adopting more alternative measures like digital literacy, proactive access to public information, civil and administrative sanctions and a multistakeholder approach. The focus has equally been on social media platforms to ensure more transparency and accountability on how they prevent these harms. However, these developments on regulatory guidance do not reflect in many African contexts, and there has been limited enquiry as to what contributes to such policy lag.

For example, there are two significant approaches by most governments in regulating these harms. The first approach has been through laws that define online harms as a whole concept, addressing its various aspects and the sharing of regulatory responsibilities between governments and social media platforms in a proposed law. Examples of these governments include Canada and the UK.[49] The second approach is the use of laws to address an aspect or aspects of online harms but not online harms as a whole concept. Examples of these governments include France, Germany and

---

to-curb-hate-speech-on-social-media (accessed 15 October 2021); H Athumani 'Ugandan government restores social media sites, except Facebook' 10 February 2021 *Voice of America* https://www.voanews.com/a/africa_ugandan-government-restores-social-media-sites-except-facebook/6201864.html (accessed 16 April 2021).

[47] United Nations General Assembly (n 13 above).

[48] As above.

[49] UK online harms white paper https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973939/Online_Harms_White_Paper_V2.pdf, UK (accessed 21 August 2020); Government of Canada 'Discussion guide' https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/discussion-guide.html (accessed 21 August 2020).

© University of Pretoria

Ethiopia.[50] So far, there is no example of any stakeholder seeking to define online harms as a whole concept and the approach that best regulates them in any law or policy in the African context.

Given the complexity of social media platform governance, many proposals seek to ensure that online harms are prevented, and online expression is protected. There have been government-led initiatives like the enactment of laws. In some African countries like Uganda,[51] Tanzania,[52] Kenya,[53] these laws are mainly cybercrime and electronic communication laws that provide for online disinformation and hate speech offences. In other countries like Nigeria, it has been through specific laws on information disorder and targeted online violence.[54] For platform-focused solutions, the Facebook Oversight Board was put in place as a response to calls for holding Facebook responsible for their regulatory decisions.[55] There have also been sectoral efforts like the Global Internet Forum to Counter Terrorism (GIFCT) by social media platforms and the Santa Clara Principles by the academia and civil society actors.[56] In addition to these, initiatives like the ARTICLE 19's Social Media Councils (SMCs), the Business of Social Responsibility's report on content governance, the Global Initiative Network's human rights analysis for content moderation all seek to define a more global-facing set of norms that is anchored to international human rights law.[57]

All of these initiatives are ground-breaking in their own ways in that they grapple with a novel, complex and dynamic issue like social media platform governance. However, none of these initiatives has shown understanding of the legal and regulatory landscape that can ensure a rights-respecting approach to preventing harms while

[50] France: LAW n° 2020-766 of 24 June 2020 aimed at combating hateful content on the internet; Ethiopia: Computer Crime Proclamation No 958/2016; Germany: Network Enforcement Act of 2017; Hate Speech and Disinformation Prevention and Suppression Proclamation (2019).

[51] Sections 24 & 26 of the Computer Misuse Act of 2011 provides for the offences of cyberharassment and cyberstalking, respectively.

[52] Tanzania's section 16 of the Cybercrime Act of 2015.

[53] Section 22 of Computer Misuse and Cybercrime Act of 2018.

[54] Ethiopia (n 50 above); the Protection from Internet Falsehoods and Manipulation and Other Related Matters (PIFM) Bill of 2019

[55] K Klonick 'The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression' (2020) 129 *Yale Law Journal* 2437.

[56] Global Internet Forum to Counter Terrorism https://gifct.org (accessed 15 March 2021); Santa Clara principles https://www.eff.org/files/2015/07/08/manila_principles_background_paper.pdfaccessed (accessed 15 March 2021).

[57] ARTICLE 19 'Social media councils: Consultation paper' June 2019 https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf (accessed 16 June 2019); Business for Social Responsibility (BSR) 'A human rights-based approach to content governance' March 2021 https://www.bsr.org/reports/A_Human_Rights-Based_Approach_to_Content_Governance.pdf (accessed 2 April 2021); Global Network Initiative (GNI) 'Content regulation and human rights' September 2020 https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf (23 February 2021).

© University of Pretoria

also protecting online expression from an African context. This understanding is necessary in order to examine the reasons for these policy lags especially from the major stakeholders involved in platform governance debates.

For example, in Nigeria, laws that seek to regulate online harms and expression suffer from a colonial hangover. The Criminal Code Act and the Penal Code that provide for various offences in Nigeria before its flag independence in 1960 are still in force. Since then, there has been no meaningful effort to examine the propriety of these laws in Nigeria's current legal environment, which has undoubtedly changed due to new developments. These old laws form the basis of traditional governance that limit online expression.[58] The impact of this foundation is even more protracted today that the conversations about regulating social media platforms have reached a crescendo in many African countries. This suggests a linear relationship between colonial legal provisions and cyber laws that seek to prevent online harms in Nigeria.[59]

On the other hand, while South Africa also experienced colonialism, its legal system shows interesting developments, especially with respect to governing online speech. One of the ways these developments have been possible is how amenable its legal system is to considering international human rights standards in deciding issues that deal with the regulation of online expression on social media platforms.[60] This consideration may be attributable to a non-linear relationship between old laws, despite its colonial history and cyber laws that seek to regulate online harms.

---

[58] Section 59 of the Criminal Code Act, Cap C38 Laws of the Federation of Nigeria 2004 (Criminal Code Act) and section 418 of Penal Code (Northern States) Federal Provisions Act, Cap P3, Laws of the Federation of Nigeria (Penal Code) provide for the offence of false information; Section 24(1)(b) of the Cybercrime (Prohibition, Prevention etc.) Act of 2015 (Cybercrime Act) provides for the offence of false information online; Section 399 of the Criminal Code and section 204 of the Penal Code provide for the offence of insulting language and insult to religion respectively; Section 24(1)(b) of the Cybercrime Act provides for the offence of insulting and annoying language online; Section 373 of the Criminal Code and section 391 of the Penal Code provide for the offence of criminal defamation; Section 24(1)(b) of the Cybercrime Act of 2015 provide for criminalisation of false statements meant to annoy or cause ill will; Sections 50-52 of the Criminal Code and Sections 416-422 of the Penal Code provide for the offence of sedition; Section 3 of the PIFM Bill provides for the offence of causing disaffection against the state online (sedition).

[59] H Essien 'Installing Twitter seditious under Penal Code of Northern Nigeria, AGF Malami tells Court' 21 September 2021 *Peoples Gazette* https://gazettengr.com/installing-twitter-seditious-under-penal-code-of-northern-nigeria-agf-malami-tells-court/ (accessed 14 October 2021).

[60] J Duncan 'Monitoring and defending freedom of expression and privacy on the internet in South Africa' (2012) *Global Information Society Watch* 14 https://giswatch.org/sites/default/files/southafrica_gisw11_up_web.pdf (accessed 1 December 2021); T Bosch & T Roberts 'South Africa digital rights landscape report' in T Roberts (ed) *Digital rights in closing civic space: Lessons from ten African countries* (2021) 143 https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/15964/South_Africa_Report.pdf (accessed 1 December 2021).

© University of Pretoria

A clear link between both countries is that they have been victims of colonial legal legacies. However, South Africa, compared to Nigeria, is moving farther away from its colonial legal legacies and embracing more contextually relevant laws with respect to online harms. Nigeria could learn from South Africa's journey so far on developing contextually relevant laws on online harms prevention. For example, Nigeria could learn from the processes involved with South Africa's recent moves to adopt a social media charter.[61] Both countries have a lot to learn from international human rights standards in designing rights-respecting legislative policies that prevent online harms and protect online expression. Therefore, this thesis presents an opportunity to consider these two systems, highlight their similarities, differences and rethink solutions in preventing online harms and protecting online expression on social media platforms. These are the issues this thesis focuses on. It considers the importance of protecting online expression in African countries on social media platforms in a rights-respecting way. It borrows from the strengths and weaknesses of some of these ground-breaking initiatives to illustrate how to ensure effective prevention of online harms and protection of online expression in Africa using Nigeria and South Africa as examples.

1.9 Limitations

This thesis has two major limitations. One, its scope is limited to Nigeria and South Africa and the comparison of both countries with respect to ensuring a rights-respecting approach to the prevention of online harms and protection of online expression. There are fifty-two other African country contexts that this thesis does not focus on in detail that could show peculiar challenges just as they would identify new prospects in ensuring a rights-respecting approach to preventing these harms and protecting online expression. The second limitation, which is in respect to one of the major findings of this thesis, is that in order to effectively govern platforms, such governance must be generatively applied. The limitation is that such an approach is incremental in application and will therefore be dependent on the will of key stakeholders to work together to prevent these harms and protect online expression.

1.10 Structure

This thesis is divided into six chapters. Chapter one is the introduction and general overview of the study. Chapter two to five focus on the four research sub-questions raised under section 1.4, while chapter six concludes the thesis. The first sub-question, which focused on the theoretical perspectives and recent normative developments on the protection of the right to freedom to expression online in Africa, is examined in chapter two. It examines the contribution of the African indigenous societies to the

[61] See here for South African Human Rights Commission's call for developing a social media charter: SAHRC 'Terms of reference: Develop a draft social media charter for the South African Human Rights Commission https://www.sahrc.org.za/home/21/files/Terms%20of%20reference%20-%20Social%20Media%20Charter%20-%20Final.doc (accessed 15 August 2021).

© University of Pretoria

right to freedom of expression before colonialism and in the digital age. It notes that this right is not protected in the national context, explains a new form of colonialism as one of the reasons and highlights how this colonialism engenders online harms.

The second sub-question on the impacts of online harms on the right to freedom of expression online in Africa is examined in chapter three. It considers the various forms of online harms in African contexts and how they impact the right to freedom of expression online. Chapter four examines the third sub-question on how a rights-respecting approach to platform governance can be used to prevent online harms and protect freedom of expression online in Africa. It highlights the major ways platforms are governed and suggests a rights-based approach as the most feasible approach to ensuring an effective mode of governance. The fifth chapter examines the roles of internal and external state and non-state actors in ensuring an effective rights-based approach to platform governance in Nigeria and South Africa. It considers the legal and regulatory landscape of online harms in both countries and highlights the roles of specific actors in both contexts in ensuring a rights-respecting approach to platform governance. The sixth chapter summarises the findings in each chapter. It notes that even though colonial provisions negatively impact online expression in both countries, if actors collaborate creatively, normatively and generatively, they can effectively prevent online harms and promote online expression in Nigeria and South Africa.

© University of Pretoria

CHAPTER TWO: THEORETICAL PERSPECTIVES AND RECENT DEVELOPMENTS ON THE PROTECTION OF THE RIGHT TO FREEDOM OF EXPRESSION ONLINE IN AFRICA

2.1 Introduction

This chapter examines the theoretical perspectives and recent normative developments on the protection of the right to freedom of expression online in Africa. This examination is done by assessing the perspectives of indigenous African societies and Western societies on the importance of freedom of expression. It also highlights their meeting points and how these meeting points have impacted on standard-setting for protecting the right online especially in Africa. It then focuses on recent developments on the right in the digital age and how it has raised concerns for its protection in African countries. In essence, this chapter draws out the relationship between the major theoretical perspectives and their contributions to the right to freedom of expression in Africa, the recent developments on protecting the right online and how these developments are yet to be implemented in African national contexts.

In answering the first sub-question of this thesis, which examines the extent of implementing the theoretical perspectives and recent normative developments on the protection of the right to freedom to expression online in Africa, further questions to be answered under this chapter are:

a. How have major theoretical perspectives contributed to the protection of the right to freedom of expression in Africa?
b. What are the recent normative developments on the protection of the right to freedom of expression online in Africa?
c. What are the gaps between these recent updates and their implementation in African national contexts?

This chapter is divided into six sections. The first section introduces the chapter while the second section analyses the theoretical framework that supports the analyses carried out in this chapter. The third section focuses on how the right to freedom of expression is perceived in indigenous African societies and Western societies. The fourth section considers the recent updates on the protection of the right in Africa and the fifth section examines one of the reasons why these developments are not being reflected in national contexts in Africa. The sixth section concludes that extant provisions of colonial laws have found their ways into cyber laws that seek to govern online expression, therefore exacerbating online harms and violating online expression in some African countries.

© University of Pretoria

2.2 Placing postcolonial legal theory in context

The main theory applied by this thesis is postcolonial legal theory, which is a subset of postcolonial theory. Postcolonial theory is also a subset of critical legal studies.[1] In understanding postcolonial legal theory and how it applies to this research, it is important to first analyse postcolonial theory itself. Referred to as postcolonial or postcolonial theory or postcolonialism, Prakash explains that it exists as:

> An aftermath, as an after – after being worked over by colonialism. Criticism formed in this process of enunciation of discourses of domination occupies space that is neither inside or outside the history of western domination but in a tangential relation to it.[2]

Spivak terms it as 'reversing, displacing and seizing the apparatus of value-coding.'[3] The term postcolonial became popular after the World Wars, the codification of international human rights norms and the rise of new states granted independence from colonial systems including those in Africa. Now, the term is used to denote the history and effects of colonialism right from its start until the present day. The periodisation of the postcolonial is often used to consider the sum total of the effects of colonialism especially with respect to how it has affected cultures, slavery, displacements, dispossession, representation, hybridity, suppression, multiculturalism and a host of other issues that arose as a result of Western conquests in indigenous societies.[4]

While theorising the postcolonial is well beyond simplistic expressions, one of its major aims is to carefully interrogate the impacts of colonialism on the quality of life of colonised societies.[5] Beyond the material dispossession of property, the postcolonial also looks to focus on the dispossession of indigenous systems that are not easily categorised under material dispossession like culture, social practices and – as it relates to this research – law.[6] In order to engage postcolonial legal theory and the relationship between the coloniser and the colonised properly, it requires critical analysis. Such analysis is to show that the colonial laws, as foisted on indigenous societies do not immediately serve the purpose of cohesion in these societies. Rather, what exists, and as Fitzpatrick and Darian-Smith put it, is that:

---

[1] M Davies 'Race and colonialism: Legal theory as white mythology' in M Davies (ed) *Asking the law question: The dissolution of legal theory* (2002) 257 94; I Ward *An introduction to critical legal theory* (1998) 172.
[2] G Prakash 'Postcolonial criticism and Indian historiography' 31/32 *Social Text* 8.
[3] GC Spivak 'Post-structuralism, marginality, postcoloniality and value' in P Collier and H Geyer-Ryan (eds) *Literary Theory Today* (1990) 228.
[4] Spivak (n 3 above) 295 & 351.
[5] A Roy 'Postcolonial theory and law: a critical introduction' (2008) 29 *Adelaide Law Review* 317.
[6] Roy (n 5 above) 278.

© University of Pretoria

> postcolonialism would not only oppose those who perceived law as a great civilizing mode of colonization or as an instrument of development or of modernisation… Self-evidently, postcolonialism will not deny all valency to such critical views. It would, however deny their claims to completeness and finality.[7]

As a result, postcolonial legal theorists have focused on the impact of the law before, during and after the colonial process. According to Roy, postcolonial legal theorists trace ways colonial laws are imposed on annexed cultures and the ideological effects of this imposition.[8] An important aspect of postcolonial legal theory is its critique of legal positivism.[9] The main argument against legal positivism is that while it creates a theoretical opportunity to discuss legal neutrality, formal inequality and legal objectivity, it is unwilling to see the impacts of other positions which inevitably results in the promotion of large-scale and substantive inequality.[10] To Davies, 'the Western legal project framed in its liberal positivist tradition has not recognised other sources and forms of law.'[11] Davies's position points to two important issues of note. One, not only are liberal positivist theories uncritical of its relevance, source and control, they reject, often through force, the need to compare and interrogate a system designed outside its existence. Second, the 'otherness' of liberal positivism, which seeks to make Western legal traditions complete in themselves while also regarding cultures outside it as non-existent is the bane of the theory which needs to be constantly subjected to legal theorisations and inquiries.

In addition, postcolonial legal theory is an emerging area of critical legal studies which is the way postcolonial legal theorists' view of liberal positivism emerged.[12] These studies, which includes scholarship in postmodernism, deconstruction, postcolonial theory, critical race theory, feminist legal theory critiqued liberal positivism from several angles. According to Roy, one of the major feature of critical legal theory, from which postcolonial legal theory derives strength, includes challenging the static monolithic categories by liberal positivist law by insisting on the 'necessity of recognising partial realities, subjugated knowledges and subaltern positions.'[13] Postcolonial legal theory also provides an opportunity to interrogate the taxonomic legal order and pre-occupation of colonialism with the aim to dismantle the oppression

[7] P Fitzpatrick & E Darian-Smith 'Laws of the postcolonial: an insistent introduction' in P Fitzpatrick & E Darian-Smith (eds) *Laws of the postcolonial* (1999) 4.
[8] Roy (n 5 above) 319.
[9] S Bottomley & S Bronitt *Law in context* (2006) 16, 58; Davies (n 1 above) 90, 104.
[10] See P Fitzpatrick *The mythology of modern law* (1992); Fitzpatrick & Darian-Smith (n 7 above) 61; RM Unger *Law in modern society: Toward a criticism of social theory* (1986) 181.
[11] Davies (n 1 above) 277.
[12] Davies (n 1 above) 167, 195.
[13] Davies (n 1 above) 169, 257.

© University of Pretoria

of 'otherness' through the legacies of colonialism that exists in imperialism and neo-colonialism – to re-engage the effects and after-effects of these structures.[14]

In making a case for why postcolonial legal theory has become more important, Fitzpatrick stated that 'from a postcolonial perspective, it is clear that the portrayal of Western legal systems as superior has not been confined to history but continues in contemporary legal thought in this era of postcolonialism.'[15] He also noted that the colonial legal systems were an instrumental part of the imperial project as colonial laws were established as the natural default for colonised systems. As a result, Roy cautioned that territories that have been victims of colonialism, which also include their legal systems, should not be underestimated as it is gravely naive to assume that colonial legal cultures have not enormously distorted and displaced local cultures since the laws of a culture essentially reflects the essential underlying values that defines how a society is protected and for how long.[16]

In identifying the relationship between postcolonial legal theory and African legal theory, Silungwe's categorisation of African legal theory is useful. In his work, he categorises this relationship into three. First, there is the sentimentalist approach which favours African indigenous societies and their norms as being useful in spite of colonialism.[17] Second, there is the revisionist approach which focuses more on how the colonial project is the source of African legal theory.[18] The third approach, referred to as legal pluralism, is the one that calls for the re-imagination, recreation and reinterpreted forms of existing legal norms in Africa, to apply such legal norms that

[14] J Clarke 'Law and race: the position of indigenous people' in S Bottomley and S Parker (eds) *Law in context* (1997) 231, 275, 246; M Hawkins *Social darwinism in European and American thoughts 1860-1945: Nature as model and nature as threat* (1997) 185.

[15] Fitzpatrick (n 10 above).

[16] Roy (n 5 above) 329.

[17] In Silungwe's work on African legal theory, this first category comprises of TO Elias, Okoth-Ogendo and others. See CM Silungwe 'On African legal theory: a possibility, an impossibility or mere conundrum?' in O. Onazi (ed) *African legal theory and contemporary problems: Critical essays* (2014) 18. For T Elias's work on the subject, see TO Elias *British colonial law: A comparative study of interaction between English and local laws in British dependencies* (1962); TO Elias *Nigerian land law* (1971). For Okoth-Ogendo, see HWO Okoth-Ogendo 'Some issues of theory in the tenure relations in African agriculture' (1989) 59 *Africa: Journal of the International African Institute* 6.

[18] The second category comprises Snyder, Fitzpatrick, Chanock, Ranger, Mamdani and others. For Snyder, see F Snyder 'Colonialism and legal form: The creation of "customary" law in Senegal' in C Summer (ed) *Crime, justice and underdevelopment* (1981) 90-121; F Snyder 'Customary law and the economy' (1984) 28 *Journal of African Law* 34. For Fitzpatrick, see P Fitzpatrick 'Traditionalism and tradition law' (1984) 28 *Journal of African Law* 20; P Fitzpatrick *Modernism and grounds of law* (2001) 214. For Chanock, see M Chanock *Law, custom and social order* (1985). For Ranger see T Ranger 'The invention of tradition in colonial Africa: Anthropological contribution' in E Hobsbawm & T Ranger (eds) *The invention of tradition* (1983) 211. For Mamdani, see M Mamdani *Citizen and subject: contemporary Africa and legacy of late colonialism* (1996) xix, 286.

© University of Pretoria

best serve the end of justice regardless of their source.[19] This chapter, and by extension this thesis, focuses on the third approach in order to carry out a postcolonial legal assessment of the right to freedom of expression online in African countries. In carrying out this assessment, the next section focuses on how examples from indigenous African societies are not far removed from Western perspectives on the right to freedom of expression.

2.3 Mapping theories of human rights on freedom of expression: Africa versus the West or Africa and the West?

Development of the right to freedom of expression and other human rights across the globe is a product of more than one system or culture. The right to freedom of expression like other human rights was influenced by various perspectives as the need to have a more globalised system of order and peace that is planted in democratic principles became more expedient.[20] As many scholars have argued on the sources of human rights, one fact cuts across their positions: many cultures in one way or the other understand the essence of human dignity either through struggle or way of life.[21]

2.3.1 African indigenous societies and the right to freedom of expression

In addressing the contributory nature of universality of human rights, Donnelly acknowledged the role of other cultures in human rights values that:

> Although human rights are indeed an important part of the Western heritage, my focus on universality and overlapping consensus clearly indicates they have also become part of the heritage of every culture, religion, or civilization. I am not claiming that 'all human rights imagination [i]s the estate of the West.[22]

[19] The third category includes the likes of Benda-Beckman, Woodman and Obilade. For Benda-Beckman, see F Benda-Beckman 'Law out of context: A comment on the creation of traditional law discussion' (1984) 28 *Journal of African Law* 28-33. For Woodman and Obilade, see GR Woodman & AO Obilade *African law and legal theory* (1995).

[20] See MB Demobour 'What are human rights: Four schools of thought' (2010) 32 *Human Rights Quarterly*; S Mohney 'The great power origins of human rights' (2014) 35 *Michigan Journal of International Law* 828-832.

[21] B Ibhawoh *Human rights in Africa* (2018) 24; The major contention in this regard has been that the indigenous African human rights system did not have the force of command in the human rights corpus as the Western human rights culture does. While some Western scholars have argued that the predominant idea of human rights that existed in these indigenous systems was human dignity, scholars like Mutua, Ibhawoh and others have disputed these claims. See B Ibhawoh *Imperialism and human rights: Colonial discourses of rights and liberties in African history* (2007); M Mutua *Human rights: A political and cultural critique* (2002); V Sewpaul 'The West and the rest divide: Human rights, culture and social work' (2016) 1 *Journal of Human Rights and Social Work* 31.

[22] J Donnelly *Universal human rights: Theory and practice* (2003) 62.

© University of Pretoria

Following most narratives on the right to freedom of expression and how human rights systems across the world may have contributed, the role of the African indigenous system to the right is not always obvious or known. For example, many scholars have attributed the development of the right to freedom of expression as a Western human rights concept rather than a contributory process. However, this attribution has been argued against by Franck, who states that human rights are 'the current manifestation of a very long chapter in the history of ideas. Progress, as one might expect, has been uneven: remarkable in some places, not much evident in others.'[23] Even though scholars like Donnelly argue that the universality of human rights makes it ideally contributory, the point he makes in his works are not confident of African contributions to the human rights corpus.[24]

In making a case for how the human rights norms across the world came to be, Lauren argues that no culture or civilisation can lay claim as being the source of human rights in the world. He stated that human rights have a long history and that this history 'emerged instead in many ways from many places, societies, religious and secular traditions, cultures, and different means of expression, over thousands of years.'[25] Further, he argued that the drafters of the Universal Declaration of Human Rights (Universal Declaration), the document which formally formalised a globally accepted normative standards on human rights gleaned from the past. This past according to him was an aggregation of cultures that believe human dignity and rights must be protected from authoritarian states.

A close look at the indigenous African human rights systems shows that the organisation of Africa's indigenous state and stateless societies prior to colonialism points to the existence of a body of rules that are used to guide social conduct.[26] In buttressing this point, the British as one of the Western systems that perpetrated colonialism introduced the indirect system of governance because indigenous societies were already organised in many parts of the continent.[27] A logical conclusion from this shows that African indigenous societies had organised systems which may not immediately fall within the democratic or human rights ideals of the colonisers but are ideal enough for the Africans guided by such systems.[28] This conclusion whets the

[23] TM Franck 'Is personal freedom a Western value' (1997) 91 *American Journal of International Law* 595-625.
[24] Donnelly (n 22 above) 62.
[25] PG Lauren 'The foundations of justice and human rights in early legal texts and thought' in D Shelton (ed) *The Oxford handbook of international human rights law* (2015) 163.
[26] Donnelly (n 22 above).
[27] M Lechler & L MacNamee 'Indirect colonial rule undermines support for democracy: Evidence from a natural experiment from Namibia' (2018) 51 *Comparative Political Studies* 1861.
[28] See N Cheeseman & J Fisher 'How colonial rule committed Africa to colonial rule' 2 November 2019 *Quartz Africa* https://qz.com/africa/1741033/how-colonial-rule-committed-africa-to-fragile-authoritarianism-2/ (accessed 17 June 2020).

© University of Pretoria

point that organisation, which is first a precursor to social existence, was very much African as much as it was Western.

According to Williams' study of 26 African nations and 106 languages, he noted that African indigenous societies had a semblance of modern-day constitutions through their customary laws and practices.[29] He explained that from the semblance emerged a 'Bill of Native Rights' among indigenous Africans. An important aspect of this Bill is:

> The right to comment on and criticize government policy, since the legitimacy of the "ruler" is based upon the consent of the people. The right to express an opinion freely and the right to be heard are fundamental to African culture and participatory democracy… Freedom of Expression (required to debate, criticize policies and participate in the decision-making process).

In restating the importance of freedom of opinion and expression in African indigenous societies, Ayittey argues that not only was freedom of expression taken for granted in many of these societies, its actualisation is premised on consensus.[30] Ayittey also referred to Cruickshank that 'anyone – even the most ordinary youth will offer … opinion, or make a suggestion with equal chance of being heard, as if it proceeded from the most experienced sage.'[31] These thoughts have also been re-echoed by the likes of Busia who noted in his seminal work that:

> The members of a traditional council allowed discussions, a free and frank expression of opinions, and if there was disagreement, they spent hours, even days if necessary, to argue and exchange ideas till they reached unanimity. Those who disagreed were not denied a hearing, or locked up in prison, or branded as enemies of the community… the traditional practice indicated that the minority must be heard, and with respect and not hostility. The traditions of free speech and inter-change of views do not support any claim that the denial of free speech or the suppression of opposition is rooted in traditional African political systems.[32]

An important feature of many indigenous societies before colonialism is organisation based on a set of beliefs which are popularly sourced and adhered to. In describing this set of beliefs and their uniformity across the continent, the then Organisation for African Unity (OAU) stated that:

> African culture, art and science, whatever the diversity of their expression, are in no way essentially different from each other. They are but the specific expression of a single universality.[33]

An important aspect of these beliefs is consensus. For example, the Ga-Dangme society, regarded as one of Africa's foremost indigenous state societies before the

[29] C Williams *The destruction of black civilization* (1987) 175.
[30] G Ayittey *Indigenous African institutions* (2006) 276.
[31] As above.
[32] KA Busia *Africa in search of democracy* (1967) 276.
[33] OAU Charter https://au.int/sites/default/files/treaties/7759-file-oau_charter_1963.pdf (accessed 20 March 2020).

© University of Pretoria

colonial conquests was one of the most organised in the region. Existing in today's Greater Accra in Ghana, evidence from history shows that the *akutso*, a group of complex lineages in the Ga-Dangme society, constituted the basis for socio-political order. This group provided major forums for discussing issues affecting their society. The *man-dzrano,* which is also known as the public square, set in motion by the *akutso*, is a place for public assembly, 'initiating young people into the art of public speaking, political discussion, debates and news exchange.'[34] The *akutso*, through the *man-dzrano*, as a result engineer consensus to drive an organised Ga-Dangme society by playing the role of modern-day press. An important aspect of this constitution is not only rigorous intellectual exercises, but also seeking out truths through debates while also catering for individual self-development.

The Akans, who are now divided among present-day Ghana and Ivory Coast, used to have a commoners' association whose leader is called the *Nkwankwaahene.*[35] Unlike the society's elitist social arrangement, the *Nkwankwaahene* is neither hereditary or inherited, rather it is often occupied by an individual who has demonstrated bravery and eloquence. It was through the *Nkwankwaahene* that the commoners expressed their grievances to the Council of Elders. According to Amoah:

> …the office of the *Nkwankwaahene* provided an effective channel for expressions of popular criticisms against the ruler and his government. It enabled the elders to take action against the ruler without being charged with disloyalty or jealousy.[36]

The laws of the Somalis, who largely operated stateless societies before colonialism, originated from the intersection of people and they believe that these laws are 'a product of reason and the conscience of the community.'[37] Law in indigenous Somali societies recognises that people have inherent freedoms as a result of their existence. An important aspect of these freedoms are freedom of expression, trade, contract and movement.[38] To Cerulli, the Somali legal terminologies which have been used to regulate their societies before colonial conquests have been found to be practically devoid of loan words from foreign language therefore making Somali's body of jurisprudence, including those on freedom of expression truly indigenous.[39]

Among the Bantu in Southern Africa, the importance of freedom of expression especially in legal adjudication is primarily important.[40] In the Bantu system, both the

---

[34] Ayittey (n 30 above) 25.
[35] KA Busia *The position of the chief in the modern political system of Ashanti* (1951) 276.
[36] GY Amoah *Groundwork of government for West Africa* (1988) 176.
[37] M van Notten *The law of the Somalis* (2006) 35.
[38] FD Heath 'Tribal society and democracy' (2001) 5 *The Laissez Faire City Times* 22.
[39] Ayittey (n 30 above) 79 citing E Cerulli 'Somalia: Scritti vari editi et inediti' (3 volumes) *Instituto Polografico dello Stato* (1957–1964).
[40] P Bohannan & L Bohannon *Tiv economy* (1968) 199.

© University of Pretoria

plaintiff and defendant are able to freely express themselves before the community judges who are usually three in number.[41] Winnie Mandela noted the brilliance of the constitution of African indigenous societies when she said that 'the council (of elders) was so completely democratic that all members of the ethnic group could participate in its deliberations.'[42]

Of the Igbos, among African indigenous societies, who now exist in today's south-eastern Nigeria, Harris notes that:

> The village assembly characterized Igbo democracy. It was there that the elders presented issues to the people, everyone had a right to speak *(freedom of expression)*, and decisions had to be unanimous. The village assembly therefore was a body in which the young and old, the rich and poor could be heard. Every citizen's participation was possible and important. Decision-making could often be time-consuming, but the slow procedure guaranteed greater individual participation.[43]

According to van Notten on assemblies in indigenous African systems, 'the reason why the Assembly operates by consensus is easy to understand: It prevents the Assembly from taking decisions that would infringe on anyone's freedom and property rights.'[44] While stressing that assemblies as a form of consensus-building is not necessarily African, Ayittey argues for centring assemblies in traditional African societies as an important form of consensus-building that, 'consensus was the cardinal feature of the indigenous African political system.'[45]

In order for this consensus-building to be possible, two fundamental requirements are important. The first one is participation in the decision-making process. In order to be able to build a sense of belonging in many African indigenous societies through consensus, they must be afforded the opportunity to meaningfully contribute to the process that leads to the adoption of social rules and law making. This feature of consensus lends support to the near-formalisation of basic democratic principles and how they relate to expression in African indigenous societies. By allowing adults to sit in most of these assemblies, they are equally allowed to voice their opinions and make contributions to ongoing debates.

Second, consensus is not achievable without the respect for freedom of opinion and expression in these meetings. Logically, reaching consensus requires that everyone airs their opinion and express themselves fully and it is the aggregate of these expressions that are enforced as a consensus position by the community. Controversial positions by members of the assembly are often vigorously debated until

[41] Ayittey (n 30 above) 85.
[42] W Mandela *Part of my soul went with him* (1984) 53.
[43] JE Harris *Africans and their history* (1987) 121.
[44] van Notten (n 37 above) 82.
[45] Ayittey (n 30 above) 136.

© University of Pretoria

reason prevails. It is this reason that is often considered as the backbone of consensual positions by the various traditional societies.

Drawing the link between assemblies, consensus-building and the right to freedom of expression in African indigenous societies, Ayittey notes that the first and second units of political leadership were the village chief and Council of Elders while the third unit was the:

> "Village Assembly"—public assembly of all citizens. At the village meetings, individuals exercised their freedom of expression without fear of harassment. It was up to individuals to make sensible suggestions or fools of themselves. But their right to freedom of expression was respected and upheld. At such meetings, however, every effort was made to reach a consensus.[46]

He shares further that:

> It is important to note the tradition of *freedom of expression* at such gatherings. Everyone—even including non-tribesmen—expressed their views freely. Their freedom of expression was assured. Sensible proposals or ideas were often applauded and inappropriate ones vocally opposed. Dissent was open and free, with due respect to the chief, of course. Dissidents were not harassed, arrested, or jailed. If a dissident made an intelligent argument, he was praised for having offered an idea that could help the community. If he made a silly remark, he set himself up for ridicule. Or if he offered a proposal with little merit, it was rejected by the assembly.[47]

An aggregate summary of these cultures in African indigenous societies has shown that the basic tenets of organisational cultures in both Western and African societies are not remarkably different. In drawing this connection, Karp argued that:

> A careful comparison of African and Western cultures shows that they share common spheres of concern with the limits on the controls people can hold over their social and natural environment and with how they reassert control or influence in their worlds. In both Western and African cultures this set of questions and problems includes technology, morality and belief…
> The great conclusion of E.E. Evans Pritchard's pioneering study of the Azande systems of thought was that differences between the Azande and Westerners were not differences in logic or thinking capacity. The Azande and other Africans reason as much as people everywhere do.[48]

Given these examples above, both state and stateless societies in indigenous Africa appreciated and understood the importance of the right to freedom of opinion and expression before colonialism.

[46] Ayittey (n 30 above) 293.
[47] Ayittey (n 30 above) 139.
[48] I Karp 'African systems of thought' in I Karp & CS Bird (eds) *Explorations in African systems of thought* (1986) 202.

© University of Pretoria

2.3.2 Western human rights perspectives and the right to freedom of expression

The major Western human rights perspectives on the right to freedom of expression are: truth theory; democracy theory; self-development and self-realisation theory; autonomy theory and the theory of human dignity. These theories are often referred to as the 'classical theories' and regarded to be steeped in the older versions of thought from the likes of John Locke, John Stuart Mill and several other Western scholars.[49] Considering the scholarship on human rights, especially the right to freedom of expression by most Western scholars before the twentieth century, emphasis is placed on the centrality of the individual as the human rights agent and not necessarily the community.[50] This emphasis is from the concept of liberalism – the political philosophy based on the consent of the governed and rule of law.

One of such notable scholars in this area is John Stuart Mill. The question of whether Mill represents or is influenced by the Western idea of human rights or vice versa is largely rested given the scholar's immediate environment and how this has been shown over the years in his writing to have greatly influenced his work.[51] Therefore, the influence and impact of the scholar's constant contact with his immediate political environment seems to have played a pivotal role in how he views human rights in general and free speech specifically. Connected to this influence is also how these ideas impact several others who have largely accepted his postulations.[52] The various theoretical positions on these ideas are next considered in turn.

A The truth theory and the right to freedom of expression

The crux of Mill's ideas on liberty and free speech is that the individual must not be controlled by the community for expressing his unpopular views and these views should be regarded with the same respect as popularly held ones.[53] However, when

[49] Regarded as the 'father of liberalism', John Locke held strong opinions on the need to focus more on the individual in a society as an autonomous unit whose expression must at all times be allowed unfettered. See J Locke *Two treatises on government* (1689); See also RP Krayanak 'John Locke from absolutism to toleration' (1980) 74 *The American Political Science Review* 53-69 where Krayanak argues that Locke's ideas on free speech being an absolute right have been misread to exclude liberal toleration. John Stuart Mill on the other hand agrees with Locke's ideas on how firm the right must be treated but unequivocally departed from the individualistic focus of Locke by including the 'harm principle' which is allowed to limit the right of an individual. See JS Mill *On liberty* (1859) 6-86.

[50] As above.

[51] According to HLA Hart, 'Mill's principles are still very much alive in the criticisms of law.' See HLA Hart *Law, liberty and morality* (1963) 15; For CL Ten, 'it will be a long time before the message of *On Liberty* becomes redundant.' See CL Ten *Mill on liberty* (1980) 204.

[52] Some of these scholars include but are not limited to I Berlin *Essays on liberty* (1969) 15-29; J Rawls *A theory of justice* (1971) 41, 162, 210, 232 & 234; CL Ten 'Was Mill a liberal?' (2002) 1 *Politics, Philosophy & Economics* 355-370; HLA Hart *Law, liberty and morality* (1963).

[53] Mill (n 49 above) 45.

© University of Pretoria

the actions of such individuals constitute danger to others, Mill surreptitiously places the limitation of such individuals in the community.[54] According to him, this limitation is the only time when the right of the individual must be suppressed because of the harm it causes others. The need for this kind of relationship between the individual and the community is because it assists in the search for the truth. Establishing the truth is the most important justification of freedom of expression and this justification is not in any way diminished by the possibility that some forms of expression are false.[55] Here, Mill's thoughts on protecting speech including that which may not immediately appear to be true may be best understood in the contexts of the positives of negativity – that a thing is immediately problematic does not mean that it cannot be positive or influence positive thinking[56] and as a result, stifling an opinion, whether false or correct, would be evil.[57]

However, despite his fierce defence of individualism in guaranteeing the right to freedom of expression, he puts a lid on such guarantees where it stops being 'self-regarding' – failing to recognise others in such expression. According to him, society must allow the individual to be 'self-regarding' until he is unable to be 'other-regarding.'[58] 'Other-regarding' here connotes expressions that do not harm others.

This twin proposition is best understood as where the line is drawn between a form of speech whose harm is only limited to the individual and the one which extends to his neighbours or society. Such an instance is only where Mill suggests that an individual's right to freedom of expression may be limited. It may be regarded as the source of the popular free speech philosophy, 'the harm principle.'[59] Simply put, the individual's right to freedom of expression is no longer free when it begins to pose harm to others.

One of the criticisms against Mill and particularly 'the harm principle' is that it is too abstract to justify its use by the state.[60] Since modern societies are now formed as state entities, formulation of harm is mostly carried out through political and legal institutions. Given the harm principle, its generality is not specific enough to give direct and unequivocal limitations of powers to the state on how to adequately respond to the tensions between individual rights and collective rights. For example, aside from social punishments that do not carry formal sanctions that may be carried out by the public on an erring individual, the state in applying the harm principle tends to define what harm is and in so doing, creates undesirable climate for dissent through state institutions.

---

[54] Mill (n 49 above) 143-144.
[55] Mill (n 49 above) 64.
[56] As above.
[57] Mill (n 49 above) 33.
[58] Mill (n 49 above) 28, 145.
[59] As above.
[60] See DA Dripps 'The liberal critique of the harm principle' (1998) 17 *Criminal Justice Ethics* 3-18.

© University of Pretoria

While this criticism is justified given several examples that could be gleaned from it, it may be argued that Mill's points are not moot nonetheless. It may be argued that Mill's argument is not an end in itself but a means to an end, the end being a system that typifies the scope of limitations that may be exercised by the state in restricting freedom of expression. If this argument is of any value, it would be seen in how the thoughts of the likes of Mill, setting the liberal scope of the right to freedom of expression, is seen in the codification of the rights in international treaties towards mid-twentieth century.[61] In essence, the work of the likes of Mill is fairly done if all that was achieved was to lay some theoretical foundations to conceptualise the right to freedom of expression in subsequent efforts through the law.

B The democracy theory and the right to freedom of expression

Closely tied to the evolution of the right to freedom of expression is democracy theory. The theory is based on the indispensable need for every qualified adult to contribute to the formation of the state.[62] This theory is also best considered with the right to hold opinions. An illustration of the relationship between the right to hold an opinion, freedom of expression and democracy is seen when citizens need information to hold an opinion to express themselves in order to contribute to the democratic process. A value chain for democracy may then be regarded as the strong interdependence of both rights, the guarantee of other rights and their combined effect in making democracies truly democratic. The core justification for the rights to freedom of opinion and expression in relation to its importance for democratic societies is a seldom contested fact. This justification is because people-based governments must involve contributory expressions in several ways including engaging in public policy debates, participating in organisations to further personal or public interests and exercising the rights to vote based on personal opinions and expressions.[63]

One of the major criticisms against the democracy theory of the right to freedom of expression is the tyranny of the majority.[64] As argued by Redish, democracy is not an end but a means to an end which is to establish a system of government that ensures the best output for all. Dahl and Diamond, in their separate works, theorise democracy to be beyond a process but that includes respect for human rights and the rule of law; collective deliberation; choice and participation; representative and accountable governments.[65] All of these, when viewed together cannot be narrowly construed as just a process but rather, an end to be achieved through democracy. A closer look at

---

[61] J Marshall *Personal freedom through human rights law* (2009) 111.

[62] See MH Redish 'Self-realisation, democracy and freedom of expression: a response to Professor Baker' (1981) 130 *University of Pennsylvania Law Review* 681.

[63] As above.

[64] M Redish, 'The value of free speech' (1982) 130 *University of Pennsylvania Law Review* 605.

[65] RA Dahl 'What political institutions do large-scale democracies require?' (2005) 120 *Political Science Quarterly* 196; L Diamond 'The democratic rollback: the resurgence of the predatory state' (2008) 87 *Foreign Affairs* 37.

© University of Pretoria

these ends also show that they are practically impossible without the guarantees of human rights in general and the right to freedom of expression in particular.

C The self-fulfilment theory and the right to freedom of expression

The major thrust of the self-fulfilment theory of the right to freedom of expression is the value added through the optimisation of personal abilities. Redish, in particular, has argued that human beings are naturally wired to seek self-fulfilment through their abilities which are activated and accentuated by being able to express themselves.[66] An important aspect of this theory is how it focuses solely on the individual. Prior to Redish, Emerson also expressed the self-fulfilment theory in two key principles.[67] The first principle is such that the society depends on the individual to express himself for it to hold together while the second contends that the individual also has a duty of being part of his community in cooperation. However, self-development is impossible without access to information which informs opinions. If the need for self-development may be realised only through the right to freedom of expression, expression can only be fully realised by informed opinions which are made available by access to information.

However, this theory is premised on the assumption that the goals of self-fulfilment of the individual will always agree with that of his society. Even though cooperating with the society is one of such duties of the self-fulfilling individual, the theory does not adequately address instances where such cooperation would be impossible given the likely tyranny of the state to coerce expression. A fair response to such criticism would be that in order to advance the self-development theory, clear and narrow instances of where such development would affect the wellbeing of the community must be agreed on by the society.

D The autonomy theory and the right to freedom of expression and information

The concept of autonomy which is sourced primarily from Mill's libertarian thoughts is central to being human. According to Scanlon, it is how to task the individual as a rational being who is trusted to make the best decision given the circumstance.[68] This theory believes that the human being is capable of opinions, expressions and holding on to the logical balance for both to drive his life.[69] The autonomy of any human being

[66] Redish (n 62 above) 621.
[67] T Emerson *The system of freedom of expression* (1970) 15.
[68] T Scanlon 'A theory of freedom of expression' (1972) 1 *Philosophy and Public Affairs* 204-226.
[69] Picking from the elements of Karl Popper's *Open society and its enemies*, of his five theories of what makes an open society which are limitation of state institutions and protection of freedoms, elimination negative utilitarianism – commonwealth benefits for the few, progressive development, *rational criticism and individualism*, an autonomous individual is at the centre.

© University of Pretoria

to freely express themselves is intrinsic and as a result should not be interfered with under any circumstances.

The autonomy principle resonates firmly with the theories of truth, democracy and self-fulfilment which in turn are not possible without first forming an opinion. In order for any of these theories to be useful to the rights to freedom of opinion and expression, the individual must be regarded as a unit capable of rational thoughts and given different circumstances will make the best decision. For example, it is the ability to hold thoughts through logical deduction or induction based on available information that the search for truth, through debates and communication with other members of the society is possible.

A rebuttable argument against autonomy would be when the individual does not only constitute harm to his community but also to himself. While arts in some parts of the world are expressed through bodily art which may immediately constitute harm from the perspective of an outsider, the possibility of carrying out an irrevocable harm like suicide is still a subject of contention on whether it is a form of expression to be protected. Circling back with Mill and his position on self-regarding and other-regarding acts, there is reference to the duty the society owes the individual other than it being negative. The human rights culture in the West is however divided on whether suicide is a form of expression and therefore leaves the theory on autonomy as it relates to self-regarding acts and the society's duty to such.

E Human dignity and the right to freedom of expression

The idea of human dignity when considered with the rights to freedom of expression is often discussed in three ways.[70] The first one is the intrinsic dignity which explains that the human being is inherently accorded a dignified status just by being a human. Being a person is enough qualification to have such a form of dignity and it cannot be given or taken away. The second perspective is the communitarian dignity which supposes that an individual's dignity is treated with respect to that of others in his community. Here, his dignity is usually what the organised society defines as such to be in order to be able to set the limits for co-existence. Some of the examples of such limitations are defamation laws and in more recent context, protection of children's rights and expression. The third form is substantive dignity which applies a certain historical, cultural or political context to what dignity means.

The intrinsic human need for other human relationships has made it possible to be fully human. If these relationships are made and strengthened based on information and opinion, the right to hold an opinion is therefore central to the theory of human

[70] GE Carmi '"Dignity," the enemy from within: A theoretical and comparative analysis of human dignity as a free speech justification' (2006-2007) 9 *University of Pennsylvania Journal of Constitutional Law* 957, 969-970.

© University of Pretoria

dignity and by extension, the right to freedom of expression. While there are contentions on how the concept of dignity may limit the right to freedom of expression of the individual, it is important to note that as established through the considered theories, the right to freedom of expression is not an absolute right. There are continuous formations as to its protection and limitations. However, one of the greatest problems with the right despite the wealth of theories that support it is how to meaningfully regulate in such a way that it establishes the required level of protection while also managing the possible harms that may arise from prohibited speech. This challenge is not particularly solved given that states are often the institutions charged with drawing up such limitations who often end up over-limiting or under-protecting the right.[71]

2.3.3 Cross-cutting perspectives on the right to freedom of expression in African indigenous and Western societies

According to Donnelly, 'we must be very clear that we are drawing on cultural resources for the purpose of human rights advocacy, not defining human rights by culture.'[72] Further, he argued that rights like that of freedom of conscience, speech and association may be determined by relative systems because they thrive most on assumption of autonomous individuals that may challenge traditional concepts of communities. The concept of human rights is universal but since the place of Western human rights culture is settled in modern human rights discourse, some cultures including those from Africa really do not have that human rights-speak to qualify in the first place let alone be universal.

This assertion by him was carried out in less than two pages of his book, *Universal human rights: In theory and practice* to form an opinion about a system of thoughts as complex as that of the African indigenous values which has not only spanned several centuries but whose undoing has been to suffer the documentation Western philosophers do not regard as objective – oral tradition.[73] However, there is more clarity on his position that while he agrees that cultures make up the universality of human rights, some cultures, including 'African traditional societies did not develop a significant body of human rights ideas or practices prior to the twentieth century.'[74] He further noted that any human rights tradition, whether Western or African, can be used to support human rights just as they could be used to infringe on them.[75] This acknowledgement of African traditional systems shows, albeit reluctantly, that these

[71] States have the primary responsibility to implement the provisions of international law. See M Land 'Against privatized censorship: proposals for responsible delegation' (2019) 60 *Virginia Journal of International Law* 29.
[72] Donnelly (n 22 above) 70.
[73] See D Kuwali 'Decoding afrocentrism: Decolonising legal theory' in O Onazi (ed) *African legal theory and contemporary problems: critical essays* (2014) 80-81.
[74] J Donnelly 'The relative universality of human rights' (2007) 29 *Human Rights Quarterly* 286.
[75] As above.

© University of Pretoria

systems are not new to the various conceptions of human rights including the right to freedom of expression.

Where this chapter contributes to the existing works on these theories and perspectives is by examining whether there are any clear relationships between both Western and African human rights perspectives. Major theories often ascribed to many Western philosophers have been found not to be practiced in the West alone. While the likes of Mill have been considered to be one of the earliest proponents of free speech, historical evidence as documented have shown that these theories were merely fine-tuned and not established by them. In fact, what has been shown by various examples of African indigenous societies highlighted above have been that universal ideas on freedom of expression are as Western as they are African. While there are differences in ways of life of many African indigenous societies, the means through which this way is achieved are not markedly different. These means have been largely punctuated by the individual being a social unit capable of rational expression and in turn an important member of the community necessary for societal growth. These forms of expressions have been seen to be grounded in the search for truth, ensuring democratic development, improving self-fulfilment, encouraging individual autonomy and also respecting human dignity.

Taking the Ga-Dangme traditional society for example, the rigour of intellectual exercises which included debates, public speaking, political discussion and many others were evidence of what Western philosophers regard as the truth theory. The truth theory is essentially so because of the ability of every individual to freely express themselves. Considering Ayittey's observation of freedom of expression in these societies, it can be seen that not only is it closely tied to assemblies and cohesion, it also involves allowing unpopular opinions in these traditional societies. While these unpopular opinions may be false sometimes, such falsity is also allowed and further assessed through debates in public gatherings such that censorship is not the immediate solution for disagreeable comments, but rather, more conversations that are able to further tease out the truth.

Also, in setting limits to 'self-regarding' speech according to Mill, such speech must also be 'other-regarding' in that a speech maker must pay attention to the likely harm that may be caused to his society as a result of his speech. This 'other-regarding', according to Mill, which is the only accepted means of limiting speech, is not alien to the Somali jurisprudence which not only expresses that 'one's freedom ends where another's begins' but which is also wholly indigenous.[76] An example of how the African indigenous societies respond to one of the major criticisms of Mill's 'harm principle' on the ever-changing dynamics of how modern states can limit speech, is how the traditional societies focus more on debates in the public square rather than censorship. Here, the assemblies are the state while the chiefs are often seen as the lead enforcer

[76] Ayittey (n 41 above) 78.

© University of Pretoria

of rules made in these assemblies. So, power resides with this public and it is this public that is given the powers to define limits through debates and consensus in most indigenous societies. Therefore, modern state can borrow from these indigenous societies on the role of publics in drawing up limitations on the right to express.

As pointed out by Ayittey, consensus-building is an integral part of the indigenous African societies. In strengthening this point, the British colonial system adopted the indirect rule because of the already-organised societies in the traditional societies they colonised. Both facts show that to a large extent, the right to freedom of expression in these societies was central to their organisation and democratic processes. For example, as explained by Ayittey, the requirements for consensus-building are participation and guarantees of freedom of expression. Therefore, Western philosophies on the need to protect the right to freedom of expression due to its democratic importance is well founded under traditional African systems. Ayittey's position has also been re-echoed by Kuwali when he stated that 'the golden thread in most treaties on the African continent is the provision of making decisions by consensus.'[77]

Considering both Dahl's and Diamond's theories on the need for freedom of expression in ensuring democracy include the rule of law; collective deliberation; choice and participation; representative and accountable governments, all these features can be found in traditional African societies.[78] The nature of publicly sourced law through public assemblies and deliberative communications were key aspects of organisation in traditional African societies. There is not much to take apart in these traditional societies that is not found in the more popular democracy theory of free speech.

Even though not in direct response to Redish, Ayittey pointed out that rather than the tyranny of majority, many African traditional societies are more focused on consensus that is devoid of coercion. He noted that the majority's views rarely bear on the overall decision the same way consensus does. This comparison to a fair extent addresses the points raised by Redish. However, it should be noted that such consensus-building might be difficult to achieve in today's modern societies especially given the advent of technologies and amplification of communications.

The communitarian features of African indigenous societies have made the need for self-fulfilment through freedom of expression more prominent. The nature of these societies, in their organisation and continued social survival has intricately woven the communities' objectives and the individuals' needs. Considering Redish's part on how freedom of expression foregrounds self-fulfilment, the quality of respect given to each member of the community in indigenous societies in Africa, especially during

[77] Kuwali (n 73 above) 87.
[78] Dahl & Diamond (n 65 above).

© University of Pretoria

deliberations of public issues support Redish's claim. The Somalis believe strongly in the agency of the individual to do the right thing while the Ga-Dangme society steeps this individuality in the nurturing of personal gifts through assemblies and freedom of expression. By creating an intricate web of catering to communal needs and individual development, the indigenous African societies have put forward in practice what Redish envisages for the self-fulfilment theory.

Looking closely at Emerson's ideas on self-fulfilment which focuses on the inter-relationship between the individual and his community, most African indigenous societies ensure this kind of relationship and more. Understanding that the individual is the lifeline of the community, these societies place each individual at the core of decision-making but also draw clear lines for when such placement would have dire impacts on the overall well-being of his community. For example, in the Yoruba traditional society in today's south-western Nigeria, the King who is often regarded as the one with the utmost authority stays so until the community he leads decides otherwise. There have been many instances where the communities in Yoruba societies have jointly actioned their right to political participation and expression and banished their kings for misuse of power.[79]

In dealing precisely with the challenges that arise from when an individual's self-development goals clash with his community's well-being, most indigenous societies engineer consensus-building through freedom of expression. The power to limit such a goal at individual fulfilment does not come from the constituted political authority, rather, it is popularly sourced from the views of the community through debates and unanimity.

Closely tied to the idea of autonomy as a theory of freedom of expression is rationality of the human mind. The human mind is presumed to be rational under many circumstances. Therefore, he makes the best option not only for himself but also for others within his community. An important connection to note between this theory and African indigenous practice on freedom of expression is that the public square, which is the nerve centre of public policy in many indigenous societies allows for all kinds of expression.

The reason for such allowance is focused primarily on two key reasons. The first reason is that every adult in such assembly is accorded the respect of being human and being able to fully express their thoughts and feelings with respect to matters that affect them or the community.[80] Second, such allowance to speak freely is also premised on the presence of the rights of others to query and debate such expression with the test of reason and rationality. Therefore, in African indigenous societies, every

[79] Ayittey (n 30 above) 49.
[80] Ayittey (n 30 above).

© University of Pretoria

adult is allowed to have policy opinions because they are deemed rational and their policy opinions are further tested along the lines of rationality in order to determine their acceptance.[81]

Perhaps one of the theories closely linked with the non-absolute nature of the right to freedom of expression, the human dignity theory focuses on the individual and his community and the interplay between them when it comes to the harm possible. The first part of the theory, which is often referred to as intrinsic, leans on the nature of being human – that every human being by their very nature are expressive and must be allowed to be so. It is this intrinsic nature of human dignity that also draws the line with another aspect of the theory that focuses on his community in relation to his right to express himself. His rights are required to be balanced with that of others whose rights are likely to be adversely affected. The substantive aspect of human dignity refers to the possible limitative factor on freedom of expression due to some social backgrounds like historical, cultural or political context. The theory of human dignity sits closely with the ideas of freedom of expression in indigenous Africa, as societies like the Akans, Somalis, Bantu (in southern Africa) all have the semblance of not only ensuring that freedom of expression because it is intrinsically human, but also seeing that the society is also the institution that draws the line for when such expression poses harm to the community in general.

It is important to note that the second aspect of the human dignity theory is more focused towards the limitation of the right to freedom of expression. While the third part of the theory, substantive dignity, also suggests acceptable means of limitation of freedom of expression in modern societies, a combination of both aspects may be seen as a theoretical basis for limiting free speech. Since no two societies are exactly the same, the community direction on what speech is acceptable will differ and so will the context of application of such freedom of speech.

In drawing connections between Western theories of freedom of expression and the social values in African indigenous societies, it is necessary to point out that such comparison is not done in order to place one above the other. Rather, it is to situate the ideas in both settings in each other and establish a link to a more globally accepted system of values – the international human rights system, which is similar to Silungwe's idea of legal pluralism. Therefore, the aim is not to focus on the right to freedom of expression as an ideology of geography but as an ideology of introspection, which is not alien to any society that has been victim of both internal and external oppression. Here, it should be noted that the thoughts of most Western theorists have been regarded as the bedrock of the formalisation of international human rights.[82] However, since no institution, especially that which is global in nature, exists without

[81] As above.
[82] D Smith & L Torres 'Timeline: a history of free speech' 5 February 2006 *The Guardian* https://www.theguardian.com/media/2006/feb/05/religion.news (accessed 23 May 2020).

© University of Pretoria

underlying diverse principles or ideals that establish them, it became necessary to state the foundations of this global system and also state in clear terms how the African indigenous value systems are also not alien to these foundations especially in indigenous Africa.

In applying postcolonial legal theory, African indigenous societies did have ideas that are similar in practice to Western ideas on expression. This points to the tapestry of how the right to freedom of expression has morphed through indigenous societies but displaced by problematic colonial legal provisions which have now been set straight by various developments under international human rights law. However, despite being set straight by this system through recent developments, it is discovered that not only are many of the legal provisions dating back to the colonial period still extant in most African countries,[83] they are being transplanted into the cyber laws of many African countries and they pose dangerous threats to online freedoms. It is therefore important to examine these recent updates by the international human rights system and how African countries are failing to measure up.

2.4 Recent developments on the right to freedom of expression: From the United Nations to the African Union

The number of independent states in Africa had risen to thirty-seven in 1966 when the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR) were adopted by the General Assembly of the United Nations.[84] This rise formalised the reception of independent African states of the international human rights system by adoption of these human rights treaties and becoming part of the United Nations. The International Bill of Rights, which comprises the Universal Declaration, the ICCPR and the ICESCR all had provisions for different generations of rights including civil and political rights and socio-economic rights. The Universal Declaration and the ICCPR provides primarily for the first generation of rights like civil and political rights while the ICESCR provides for the second generation of rights like socio-economic rights. Both generations of rights have been referred to as interdependent, each carrying the same weight, both in substance and implementation.

[83] O Mokone 'The colonial-era laws that still govern African journalism' 10 March 2019 *Al Jazeera* https://www.aljazeera.com/programmes/listeningpost/2019/03/colonial-era-laws-govern-african-journalism-190310080903941.html (accessed 17 June 2020); J Rozen 'Colonial and apartheid-era laws still govern press freedom in southern Africa' 7 December 2018 *Quartz Africa* https://qz.com/africa/1487311/colonial-apartheid-era-laws-hur-southern-africas-press-freedom/ (accessed 17 June 2020).

[84] F Viljoen 'Africa's contribution to the development of international human rights and humanitarian law' (2001) 1 *African Human Rights Law Journal* 19.

© University of Pretoria

Perhaps the most cross-cutting right after the right to human dignity in the UN treaty system, the right to freedom of opinion, expression and information can be found in the major binding human rights treaties of the UN.[85] The engagement of the UN with the right predates the adoption of the Universal Declaration with a 1946 Resolution by the United Nations General Assembly (UNGA) which affirmed the crucial role of the freedom of information which it described as the 'touchstone' of all other human rights.[86] Article 19 of the Universal Declaration on the right to freedom of expression and information provides:

> 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.

The consultations that birthed the Universal Declaration have been regarded as customary, multicultural and wide, even though many African countries which were still under the colonial systems could not participate in its drafting and adoption. This argument has been used to show that in ensuring a universal instrument like the Universal Declaration, it must not only be adopted unanimously but must be culturally relevant. While the former could not have been achieved given colonial systems in non-Western states, the latter could have been said to have been largely developed seventy-two years after it was adopted, given its wide reception.[87] Having assumed the status of customary international human rights law, the Universal Declaration is the fountain head of the other two binding human rights treaties – the ICCPR and the ICESCR, as it laid the formal foundation for the right to freedom of opinion and expression globally.[88]

2.4.1 The ICCPR and recent developments on the right to freedom of expression in the digital age

The draft Convention on Freedom of Information of 1948 is one of the most important starting points for the Convention on Human Rights, which later became the ICCPR and the ICESCR.[89] Referring to the ICCPR, McGonagle, stated that 'given the tabula rasa nature of the drafting exercise, the drafters sought inspiration from a wide range

---

[85] T McGonagle 'Freedom of expression and information in the UN' in T McGonagle & Y Donders (eds) *The United Nations and freedom of expression and information* (2015) 32.

[86] UNGA Resolution 59(1), 14 December 1946.

[87] J Humphrey 'The international bill of rights: Scope and implementation' (1976) 17 *William & Mary Law Review* 527 529. A subsequent work by Humphrey emphasises that the Declaration is now 'binding on all states, including the states that did not vote for it in 1948.' J Humphrey *No distant millennium: The international law of human rights* (1989) 155.

[88] A De Baets 'The impact of the Universal Declaration of Human Rights on the study of history' (2009) 48 *History and Theory* 20.

[89] United Nations, Economic and Social Council *Final act of the United Nations Conference on Freedom of Information* 21 April 1948 https://digitallibrary.un.org/record/3806839?ln=en (accessed 20 June 2019)

© University of Pretoria

of legal, political, religious and philosophical sources.'[90] While the wordings of the text in the draft Convention and what finally made it through into the ICCPR were different, the original provisions of the draft Convention were thorough and this helped to guide the texts towards their meanings, that is accommodate the possible reality of constant changes to contexts and rights. For example, the text produced at the Conference on the right to freedom of expression and information was as follows:

> 1. Every person shall have the right to freedom of thought and the right to freedom of expression without interference by governmental action; these rights shall include freedom to hold opinions, to seek, receive and impart information and ideas, regardless of frontiers, either orally, by written or printed matter, in the form of art, or by legally operated visual or auditory devices.
> 2. The right to freedom of expression carries with it duties and responsibilities and may, therefore, be subject to penalties, liabilities or restrictions clearly defined by law, but only with regard to:
> (a) Matters which must remain secret in the interests of national safety;
> (b) Expressions which incite persons to alter by violence the system of government;
> (c) Expressions which directly incite persons to commit criminal acts;
> (d) Expressions which are obscene;
> (e) Expressions injurious to the fair conduct of legal proceedings;
> (f) Infringements of literary or artistic rights;
> (g) Expressions about other persons natural or legal which defame their reputations or are otherwise injurious to them without benefiting the public;
> (h) The systematic diffusion of deliberately false or distorted reports which undermine friendly relations between peoples and States; A State may establish on reasonable terms a right of reply or a similar corrective remedy.
> 3. Measures shall be taken to promote the freedom of information through the elimination of political, economic, technical and other obstacles which are likely to hinder the free flow of information.
> 4. Nothing in this article shall be deemed to affect the right of any State to control the entry of persons into its territory or the period of their residence therein.

While the final text that was adopted under article 19 of the ICCPR reads as follows:

> 1. Everyone shall have the right to hold opinions without interference.
> 2. Everyone shall have the right to freedom of expression; this right shall include 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.
> 3. 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 reputations of others;
> (b) For the protection of national security or of public order (ordre public), or of public health or morals. [91]

According to McGonagle, the difference between these texts is the overall approach adopted by both texts, as the proposed draft focused on a more detailed enumeration

[90] McGonagle (n 85 above) 7.
[91] ICCPR, art 19.

© University of Pretoria

of the restrictive provisions while the adopted text used more precise wordings in its final provisions.[92] However, considering McGonagle's observation in describing the major difference between both texts, it showed that the draft Convention was more detailed, which not only could have helped parties to the Covenant understand the specific limits, but also stem the possibilities of violations that may arise from the wrong interpretations in the open texts in the adopted provisions. It is also important to note that parts of the first draft, which establishes the right of reply or similar corrective remedy, have been shown to be more useful today given the protracted debate on how social media platforms now play an important role and why the right of reply in such a development could have empowered more administration of justice in platform governance and regulation globally.

Despite these differences in both texts, there are also important aspects of the provisions of article 19 of the ICCPR that may be useful for determining the future of free speech through standard-setting and interpretation of the right to freedom of expression across the world. This provision of the article 19 may be interpreted in light of current legal debates on the responsibilities of the state and private actors in protecting the right to freedom of expression and information in the digital age. For example, there have been debates on who has what responsibilities under the provision. While it has been settled that states do have the primary responsibilities, private sectors could be argued to be involved by virtue of the phrase 'special duties and responsibilities.'[93]

Another key point to note is that such special duties and responsibilities have been further spelt out under the Ruggie Principles on Business and Human Rights.[94] The Principles are clear that private actors have horizontal responsibilities to protect human rights in the course of their businesses. Also, while the right to freedom of expression is closely linked to the civil and political relevance, they have been expanded through that link to include other major thematic areas of human rights like rights and welfare of children and protection of the rights of persons living with disabilities and so on.

Often, it has been argued that article 19(3) of the ICCPR should be read in conjunction with the provisions of its article 20 in limiting prohibited speech.[95] Article 20 of the ICCPR prohibits propaganda for war and advocacy of all forms of hatred that constitutes incitement to discrimination, hostility or violence. This report also notes that

[92] McGonagle (n 85 above) 16.
[93] M Nowak *UN Covenant on Civil and Political Rights: CCPR* Commentary (2005) 448; Land (n 71 above).
[94] United Nations Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework, 2011.
[95] United Nations General Assembly 'General Comment No 34, CCPR/C/GC/34' 12 September 2011 http://undocs.org/en/CCPR/C/GC/34 (accessed 26 June 2021).

© University of Pretoria

the provisions of articles 19(3) and 20 of the ICCPR are 'compatible and complement each other.'[96] The compatibility and complementary relationship referred to here is the combination of the restrictions provided for under article 19(3) and the requirement of *lex specialis* – prohibition by law. What this means, in practical terms is that in limiting the rights provided for under article 19(2) of the ICCPR on the right to freedom of expression based on article 20, the requirements under article 19(3) must be complied with.[97] Therefore, it is not enough for states to prohibit expression under article 20, such prohibition must be provided for by a law (clarity of definitions and sufficient precision), pursue legitimate aims (protect the rights of others), necessity and proportionality (apply the least intrusive means such that the restriction is commensurate to the harm sought to be neutralised). In addition to these, not only is the article a substantive provision with the sole purpose of further qualifying another right, the right to freedom of expression, it is also more proof that the right is not absolute and has internationally set limits.

Another reason for reading both provisions together is that doing so further establishes that the drafters of the Covenant envisaged the future where the interpretation of the right to freedom of expression will be stretched beyond its ordinary meaning. For example, information disorder, child pornography and hate speech have posed dangerous harms to the Internet ecosystem and while these examples have become more difficult to regulate given the dynamism of new technologies, both provisions have proven to be an important torch in navigating the complexities of protecting the right to freedom of expression in the digital age.

Perhaps the earliest defining move by the UN on the right to freedom of expression and new technologies which seems to have presented some of the most daunting challenges given the dynamism of the latter is the General Comment No 34 of the United Nations' Human Rights Committee.[98] The General Comment, which was adopted just at the turn of the decade that has witnessed the most expansive re-scoping of the right to freedom of expression, was able to set the course and formally provided more clarity on the position of the United Nations on the role of human rights and digital technologies and in particular, the right to freedom of expression and information in the digital age. The previous General Comment on Article 19, which was No 10 and adopted in 1983 could not dwell on the intricacies and complexities of

[96] United Nations (n 95) paras 50-52; United Nations General Assembly 'Expert workshops on the prohibition of incitement to national, racial or religious hatred, A/HRC/22/17/Add.4' 11 January 2011 http://undocs.org/en/A/HRC/22/17/Add.4 (accessed 1 December 2021) paras 14-19.
[97] T McGonagle 'The development of freedom of expression and information within the UN: Leaps and bounds or fits and starts?' in T McGonagle & Y Donders (eds) *The United Nations and freedom of expression and information* (2015) 21.
[98] M O'Flaherty 'Limitations on freedom of opinion and expression: Growing consensus or hidden fault lines?' (2012) 106 *Proceedings of the Annual Meeting (American Society of International Law: Confronting Complexity)* 348.

© University of Pretoria

globalised digital communications because the latter were not as prevalent as they are today.

It is important to note that even though each of the rights to freedom of opinion, expression and information are standalone rights, they have different scope of limitation, they are 'interlinked' and they can be used interchangeably.[99] The General Comment restated this link between the rights to freedom of opinion, expression and information and other human rights like privacy, association and assembly, electoral participation and others. It made a particular emphasis on how the right to hold an opinion is not a qualified right as compared to the right to freedom of expression which is qualified even while both are non-derogatory. It also clearly espoused that the rights provided for under article 19 of the ICCPR on freedom of expression, opinion and information include Internet-based modes of expression, which must be protected just like other forms of media.[100] This reference to the Internet further settled the debate on whether the ICCPR had envisaged the impacts of the Internet and other new technologies on the right to freedom of expression and information.

The General Comment also dealt lengthily with the restrictive provisions under article 19(3) of the ICCPR. This analysis was timely and necessary given the spate of misinterpretation of the content of these restrictions mostly by states and private actors. It restated that the major restrictions which must be provided for by law, proportionate to the aim sought, necessary in a democratic society and rights of others are not open-ended and disjunctive.[101] It further stated that restrictions must be construed narrowly, employ the least intrusive means and be jointly satisfied, that is they must be established by a legitimate law, not wide towards a general aim but a more specific one and be absolutely necessary in a democratic society. Article 19(3)(a) and (b) provides for these three requirements.

In carrying out restrictions on freedom of expression, the General Comment stresses that the provisions of article 19(3) must be complied with. It stated that examples of penalisation of news outlets by the state, arbitrary blocking or throttling of websites as

[99] JW Penney 'Internet access rights: A brief history and intellectual origins' (2011) 38 *William Mitchell Law Review* 23; Nowak discusses the major reason why the right to hold an opinion is regarded as an unqualified right under the ICCPR. He says it is regarded as existing within the 'realm of the mind' and therefore private as opposed to freedom of expression which involves the public and external communication. This is the main basis for distinguishing between both rights. Therefore, while the right may be often referred to as the right to freedom of expression and opinion, it must be borne in mind that such reference must include the absoluteness and non-interference with the right to freedom to hold opinions. See Nowak (n 93) 441. In another paper, Aswad draws an important distinction between both rights, right to hold an opinion without interference and right to freedom of expression, under art 19 in relation to the business models of big tech companies. See EM Aswad 'Losing the freedom to be human' (2020) 52 *Columbia Human Rights Law Review* 2, 63.

[100] General Comment (n 95 above) para 12.

[101] General Comment (n 95 above) para 22.

© University of Pretoria

a result of criticisms against the state or blanket bans on a number of websites are incompatible with the purpose of the ICCPR, and are therefore violations of not only the right to freedom of opinion, expression and information, but also to other rights that would be directly or indirectly affected by such violation.[102]

In addition to the exposition by General Comment, the Committee on the Elimination of Racial Discrimination (CERD) adopted a General Recommendation titled 'Combating racist hate speech' in August 2013. This recommendation became necessary in tracing the contours of the right to freedom of expression especially with how it relates to legitimate restrictions like prohibited speech. It is also important that given the nature of the restriction, it has a heavy focus on criminalisation of certain expressions provided for under the ICCPR and more substantively under the ICERD. It also considered the possible dynamics such kind of speech may have on non-traditional media like Internet-based platforms.[103] The General Recommendation considers other means of combating racist hate speech other than criminalisation to include civil and administrative measures. It calls for contextual factors like the position or status of the speaker, objectives of such speech, the reach, content or form or even the socio-political or socio-legal climate. It also calls for policy measures that involve education, culture, teaching and information as ways of combating racist hate speech.

2.4.2 The African Charter and recent developments on the right to freedom of expression in the digital age

The African Charter on Human and Peoples' Rights (African Charter) is the most primary regional human rights instrument in Africa. Article 9 of the African Charter provides for the right to freedom of expression and information as follows:

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*.[104]

This provision has often raised more questions than answers when it comes to the protection of the right. One of the questions it has raised is that of all the regional human rights instruments, it is the shortest and 'the weakest formulation of freedom of expression of any major international human rights document.'[105] Also, this provision, in addition to the challenges posed by inadequate protection of the right in Africa, is often interpreted by states wrongly due to the claw-back clause 'within the law.' In the past, many states have interpreted this clause to mean using domestic law to limit the

[102] General Comment (n 95 above) para 43.
[103] United Nations General Assembly 'General Comment No 35, CCPR/C/GC/35' 26 September 2013 http://undocs.org/en/CCPR/C/GC/34, (accessed 26 June 2020) para 7.
[104] African Charter on Human and Peoples' Rights, art 9.
[105] CE Welch 'The African Charter and the freedom of expression in Africa' (1998) 4 *Buffalo Human* Rights Law Review 112.

© University of Pretoria

right regardless of its effect and therefore, state parties have the powers to limit the right as they wish. However, the African Commission on Human and Peoples' Rights (African Commission), as the institution established by the African Charter to interpret the rights contained in it, has developed jurisprudence on the right to freedom of expression especially in relation to the meaning of the claw-back clause.[106] It has since stated that the meaning of the 'law', as referred to in the Charter is international law and not national or domestic law – restrictions must be by law, legitimate and necessary. It settled the erroneous claim that the state laws can limit the right without recourse to the standards set under international law and consistent with state parties' obligations under international law. In complying with international standards, state parties must show that such law does not override both constitutional and international standards, be consistent with state obligations and not permit states to apply the provisions of the Charter in such a manner that it would render the rights provided for in it meaningless.[107]

In addition to this interpretation, the African Commission also adopted the *Declaration of Principles on Freedom of Expression in Africa* in 2002, which was revised in 2019. The 2002 Declaration was followed by a resolution in 2004, which established the mandate of the Special Rapporteur on Freedom of Expression (later changed to the Special Rapporteur on Freedom of Expression and Access to Information in Africa).[108] Since 1999, the Special Rapporteurs with the exception of the African Union's Special Rapporteur have released joint declarations every year with some of them touching on the intersections of the Internet and the protection and promotion of the right to freedom of expression and information.[109] The African Union's Special Rapporteur joined other Special Rapporteurs in adopting these joint declarations in 2006 for the first time. These joint declarations and the collaboration between the Special Rapporteurs have generally improved standard-setting for the right across the region. This collaboration has given an ample opportunity for each region through the Special Rapporteur to offer their contribution to build up the expanding jurisprudence of the right as it relates to the Internet.

It is important to note two major points with respect to the major differences between the substantive provisions of the ICCPR and the African Charter on the right to

[106] *Constitutional Rights Project v Nigeria*, (2000) AHRLR 191 (ACHPR 1998) para 58; *Constitutional Rights Project, Civil Liberties Organisation and Media Rights Agenda v Nigeria* (2000) AHRLR 227 (ACHPR 1999) paras 41-42; *Egyptian Initiative for Personal Rights and INTERIGHTS v Egypt I* (2011) AHRLR 42 (ACHPR 2011) para 255.

[107] As above.

[108] African Commission 'Special Rapporteur on Freedom of Expression and Access to Information' https://www.achpr.org/specialmechanisms/detail?id=2 (accessed 15 March 2020).

[109] The Special Rapporteurs on Freedom of Expression have adopted twenty-three joint declarations since 1999. The joint declarations from 2000, 2001, 2003, 2005, 2010, 2011, 2014, 2016, 2017, 2018 & 2019 all referred to how states must regulate the right to freedom of expression on the internet. See generally Organisation for Security and Co-operation in Europe (OSCE) 'Joint Declarations' https://www.osce.org/fom/66176 (accessed 1 December 2021).

© University of Pretoria

freedom of expression and opinion. While both are interlinked rights, the scope of the rights and how they are protected are the major points of the difference. First, given the context of the ICCPR and the provisions on the right to hold an opinion as compared with the right as provided for under the African Charter, it is unqualified in the former but 'within the law' in the latter. Under the ICCPR, the right to hold an opinion is unqualified by the provisions of article 19(3) while the right to freedom of expression and opinion are qualified under article 9(2) of the Charter. Under article 9(2), the right to hold an opinion is qualified together with the right to freedom of expression while the right to access information is unqualified under article 9(1). Second, there is no equivalent provision under the African Charter on prohibited speech like there is under article 20 of the ICCPR. Rather, what exists is the catch-all provision under article 9 which may be expected to cater for such speech.

The first argument for the difference in scope of protection between both treaties is that in interpreting the rights under the Charter as provided for under article 60, the most favourable provision in both instruments would be applied to the case at hand. For example, if the right to hold an opinion is to be contested as qualified by a state party, in order to ensure that it protects the right, the African Commission or any judicial body may invoke the applicable provisions of the ICCPR under the provisions of article 60. Either way, the scope of the rights protected will be afforded the most accommodating protection of any of the binding instruments. Also, as further discussed below, the soft law instrument made pursuant to the relevant provisions of the African Charter further restated the unqualified nature of the right.

A second argument for the non-existence of substantive provisions on prohibited speech under the African Charter is that these rights may be read together with the specific duties for an individual under the African Charter, first in its preamble and later in articles 27, 28 and 29. Also, the soft laws, like the 2019 Declaration, several resolutions and press releases have also referred to the duty of state parties to apply internationally set standards on prohibited speech. For example, Principle 22(5) provides that freedom of expression may be limited on the grounds of public order or national security but there must be a 'close causal link between the risk of harm and the expression.' Also, subsection 1 of the same principle provides that 'states shall review all criminal restrictions on content to ensure that they are justifiable and compatible with international human rights law and standards.' However, while the non-binding nature of soft laws further complicates the need to mandate the adoption of these soft laws, they have provided a directional policy from the regional superstructure to national legal systems on best standards, even though practicality and implementation remains a problem.[110]

[110] B Kabumba 'Soft law and legitimacy in the African Union system: The case of the Pretoria Principles on Ending Mass Atrocities Pursuant to Article 4(h) of the Constitutive Act of the African Union' in O

© University of Pretoria

Noting the challenges that could have influenced the provision on the right to freedom of expression in the Charter, Welch[111] noted that:

> [t]he Charter was drafted on behalf of the Organization of African Unity [OAU] by persons sympathetic to governments' desires (perhaps interpreted as necessities and preconditions), and desirous of including African values. The OAU itself was far more concerned early in its history with ensuring stability for newly independent countries and self-determination for remaining colonial areas, than with protection of human rights within its member states. In order to ensure adoption by African heads of state gathered at the 1981 OAU summit and ratification by governments subsequently, the language chosen gave states wide discretion. The brief, government-cantered wording of Article 9 of the African Charter should be contrasted with the comparable, lengthier sections of the European Convention and the American Convention.

This background has also affected the protection of the right to freedom of expression and access to information in Africa. Since the adoption of the African Charter, while there have been giant strides in terms of protecting the right, however, it

> had also been characterised by many unfulfilled promises and that since 2001, year after year, 'freedom of expression, the fundamental guarantor of human rights, has been weakened and eroded in emerging and older democracies alike.[112]

In connecting the pre-2001 to post-2001 challenges on the protection of the right, it has been argued that 'the trend identified in the early years of the twenty-first century is not anecdotal or incidental but entrenched and historical in nature.'[113] This points to the fact that violations of the right are not necessarily as a result of recent technological advancements or increased state involvement in violations but due to the foundations, laid from the past and which in the context of African countries would be the colonial impact on legal systems and the particularity of post-independence instability.[114]

2.4.3 The mandates of the UN and AU Special Rapporteur on the Right to Freedom of Opinion and Expression and recent developments in the digital age

In keeping up with the challenges that may impact on the protection of the right to freedom of expression from time to time, the UN Special Rapporteur on the Promotion and Protection of Freedom of Opinion and Expression and AU Special Rapporteur on Freedom of Expression and Access to Information have provided meaningful guidance

---

Shyllon (ed) *The Model Law on Access to Information for Africa and other regional instruments: Soft law and human rights in Africa* (2018) 187-188.

[111] CE Welch *Protecting human rights in Africa: roles and strategies of non-governmental organisations* (1995) 149.

[112] A Callamard 'Accountability, transparency and freedom of expression in Africa' (2010) 77 *Social Research* 1211.

[113] As above.

[114] CE Welch 'The African Charter and the freedom of expression in Africa' (1998) 4 *Buffalo Human Rights Law Review* 105.

© University of Pretoria

on the protection of the right. In order to ensure the continued relevance of human rights treaties, given the challenges of globalisation and development, the UN Special Rapporteur together with other special procedures in the UN and the AU human rights system have carried out standard-setting or norm-setting with respect to their mandates. While there may be conceptualisations of standard-setting within other contexts, there are limited meanings to standard-setting within the context of the mandate of the Special Rapporteur. Gleaning from the mandates of the Special Rapporteur and the work that has been done since the mandate was established, it may be necessary to establish the meaning of standard-setting.

According to Mendel, standard-setting work by the Special Rapporteur has been through 'activities to both advance traditional understandings of freedom of expression and yet maintain credibility as authoritative interpretations of that right.'[115] He stated further that both needs require delicate balancing. Given the constant need for repositioning of the law for development and innovation, it has become increasingly important to keep the law relevant in the face of such developments and innovations. Therefore, standard-setting in the mandate of the Special Rapporteur on the Right to Freedom of Opinion and Expression is both a constant need as well as a dynamic effort to transform the law as a fluid concept for an ever-evolving global and regional landscape. Standard-setting under the mandate of the Special Rapporteur may then be referred to as the need to evolve and apply the formal text of applicable treaties, their purpose and application towards a demand which bears on the protection and promotion of the right to freedom of expression.[116]

For example, standard-setting under the mandate of the Special Rapporteur under the UN system has helped to achieve four key tasks on the interpretation and application of the right to freedom of expression. First, since its establishment, the mandate has facilitated a firmer, holistic and dynamic understanding of the right in relation to important areas of intersections. Through the mandate, themes like 'women's rights', 'elections', 'privacy and surveillance', 'artificial intelligence', 'administration of justice' and 'the Internet' have all been expanded on in relation to how they intersect with the right to freedom of expression and information in such a manner that it leaves no state party at a loss as to how to regulate these thematic intersections.[117]

Second, in what could have been a rigid and oft-tight application of the texts of the law on the right at the international level, the mandate has bridged the gap in jurisprudence. Borrowing from thoughts of the UN Human Rights Committee on their

[115] T Mendel 'The UN Special Rapporteur on Freedom of Opinion and Expression: Progressive development of international standards relating to freedom of expression' in T McGonagle & Y Donders (eds) *The United Nations and freedom of expression and information* (2015) 254.
[116] McGonagle (n 97) 38.
[117] See table 1 below on the various thematic focus of the Special Rapporteur on new technologies between 1998-2021.

© University of Pretoria

decisions in many petitions brought against state parties, the mandate develops clearly distilled principles of law that draw from the provisions of applicable treaties, which are finally crystallised into annual reports which serve as directions on areas of intersections with the right that would have otherwise created a gap in promoting and protecting the right.[118]

A third need for norm-setting so far by the mandate of the Special Rapporteur is that it has ensured an advanced and nuanced appreciation of international human rights law especially with respect to the protection and promotion of the right to freedom of expression and information. Perhaps, in what could have been envisaged as an impossible situation, the mandate makes it possible to attempt touching base with the basic traditional concept of the provisions under applicable regional treaties while also catering to contexts. It thereby works towards resolving the thorny challenge of international human rights law, striving for a universalism that simultaneously defers to relativism and the practice of human rights on ground by squaring the circle. Four, the mandate has further realised standard-setting through its collaboration with other regional Special Rapporteurs on the Right to Freedom of Expression and Information which now has a representative from three regions through joint declarations.[119] Each Special Rapporteur's mandate forms part of the UN and AU's special procedures and are discussed below.

A The mandate of the UN Special Rapporteur on the Right to Freedom of Opinion and Expression and recent developments in the digital age

The mandate of the UN Special Rapporteur was established by a resolution of the UN Commission on Human Rights which later became the Human Rights Council.[120] The original mandate of the Special Rapporteur requested that the Special Rapporteur gather all information on the violations of the right as provided for under the Universal Declaration, ICCPR and other mechanisms; gather relevant information on violations against professionals in the field of information as affirmed in the Universal Declaration and the ICCPR; seek and receive information from stakeholders who have knowledge of these violations, submit to the Commission a report on these tasks and offer recommendations that further help to realise the rights under the Universal Declaration and the ICCPR. This mandate was generally divided into six major tasks which include preparing annual reports to the Human Rights Council, attending meetings, producing press releases, communications, country visits and standard-setting activities.

The establishment was as a result of a report by Danilo Türk and Louis Joinet who envisaged the need to have a Special Rapporteur who has ensured the 'protection of

[118] McGonagle (n 116 above).
[119] Mendel (n 113 above) 263.
[120] Commission on Human Rights, Resolution 1993/45, 5 March 1993.

© University of Pretoria

professionals in the field of information.'[121] Also, in line with this forethought, at different times, the regional human rights systems like the Organization for Security and Co-operation in Europe (OSCE), the African Commission and the Organisation of American States (OAS) all established offices in that regard with like-mandates to ensure the protection and promotion of the right to freedom of expression and information.[122] For the first time and in a controversial Resolution in 2008 by the United Nations Human Rights Council (UNHRC), the mandate of the Special Rapporteur included 'hate speech.'[123] There were forty-four votes in support of the Resolution and of this number, eleven countries were from the African region. Of the six mandates, this section will focus on the standard-setting activities through annual reports.

Since 1994 when the first report was produced by the first Special Rapporteur and submitted to the UN Human Rights Council, it has included substantial sections addressing a thematic freedom of expression issue with conclusions and recommendations.[124] Since 1998, the annual reports by the Special Rapporteurs have in one way or the other involved considerations on the impacts of new technologies on the right to freedom of expression. It is important to note that asides the defining 2011 recommendation by the United Nations General Assembly already discussed, the Special Rapporteur submitted reports both to the General Assembly and the Human Rights Council on the exercise of the right through the Internet and on key trends and challenges to the rights of all individuals to seek, receive and impart information and ideas of all kinds through the Internet the same year. Since 1994 when the first report was presented, UN Special Rapporteurs have developed twenty-one annual reports that dealt with the right to freedom of expression and information and new technologies.[125]

---

[121] D Türk & L Joinet 'The right to freedom of opinion and expression' Final report, UN Doc No E/CN.4/Sub.2/1992/9/Add.1, 14 July 1992, para 3.

[122] The OSCE Representative on Freedom of the Media (RFoM) was established in 1997 by the Organisation for Security and Co-operation in Europe; The ACHPR's Special Rapporteur on Freedom of Expression was established by the African Commission on Human and Peoples' Rights with the adoption of Resolution 71 at the 36th Ordinary Session held in Dakar, Senegal from 23rd November to 7th December 2004;The OAS Special Rapporteur for Freedom of Expression was established by the Inter-American Human Rights Commission in October 1997 and this was approved at the Heads of States' Summit in April 1998 through the Declaration of Santiago, see here http://www.oas.org/en/iachr/expression/showarticle.asp?artID=52&lID=2#_ftn1.

[123] United Nations General Assembly 'Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/RES/7/36' 28 March 2008 HTTP://UNDOCS.ORG/EN/A/HRC/RES/7/36, (accessed 26 July 2020) para 7.

[124] See Freedom of Opinion and Expression Annual reports https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Annual.aspx (accessed 15 February 2020).

[125] As above.

© University of Pretoria

Table 1: Annual reports by the UN Special Rapporteurs on Freedom of Opinion and Expression with respect to new technologies (1998-2021)

S/N	Year	Thematic issue addressed by each report
1	1998	The right to seek and receive information, the media in countries of transition and in elections, the impact of new information technologies, national security, and women and freedom of expression.
2	1999	The right to seek and receive information, national security laws, criminal libel, new information technologies, and women and freedom of expression.
3	2000	Access to information, criminal libel and defamation, the police and the criminal justice system, and new technologies.
4	2001	Non-state actors, new technologies, and women.
5	2002	World Conference against Racism, Racial Discrimination, Xenophobia and Related Intolerance, events of 11 September, broadcasting, and the Internet.
6	2006	Internet governance and human rights, freedom of expression and defamation, and security and protection of media professionals.
7	2007	Internet governance and digital democracy, decriminalisation of defamation offences, and security and protection of media professionals.
8	2011	Key trends and challenges to the right of all individuals to seek, receive and impart information and ideas of all kinds through the Internet.
		The right to freedom of opinion and expression exercised through the Internet.
9	2012	Hate speech and incitement to hatred.
10	2013	States' surveillance of communications on the exercise of the human rights to privacy and to freedom of opinion and expression.
11	2014	The right of the child and freedom of expression
12	2015	Encryption and anonymity to exercise the rights to freedom of opinion and expression in the digital age.
13	2016	Freedom of expression, states and the private sector in the digital age.
14	2017	The role of digital access providers.
15	2018	Online content regulation.
		Artificial Intelligence technologies and implications for the information environment.

© University of Pretoria

16	2019	Surveillance and human rights.
		Online hate speech.
17	2021	Disinformation and freedom of opinion and expression.
		Gender justice and freedom of opinion and expression.

This body of annual reports has been one of the most important functions of the office of the Special Rapporteur especially in respect to the constantly evolving nature of new technologies and the threats they pose to human rights. In drawing out the elastic nature of the provisions of article 19 and other connected provisions of human rights in other treaties, these annual reports have been able to situate the dynamism of international human rights law at the centre of how new technologies not only impact the right to freedom of expression and information, but also other major thematic issues on human rights like women's rights, elections, women's rights, the criminal justice systems, non-discrimination and human dignity, communication surveillance, privacy protections, Internet access and affordability, artificial intelligence, social media platforms and others. These issues, when considered alongside the importance of protecting freedom of expression in the digital age, point not only to the centrality of the right as that which performs the nucleic function for other rights, but also demonstrates that the application of international human rights law, through an organisation like the United Nations is possible in setting new standards and norms in the area of human rights in a fast-evolving global landscape. Therefore, this chapter focuses on the themes of information disorder, online gender-based violence and hate speech and regulation of user-generated content.[126]

i. Information disorder

The report on disinformation and freedom of opinion and expression focuses on the nature, key concerns and guidance on how to regulate disinformation which it described as a form of information disorder.[127] The report highlights major concerns

[126] United Nations General Assembly, 'Contemporary challenges on freedom of expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/71/373' 6 September 2016 http://undocs.org/en/A/71/373 (accessed 26 August 2021) paras 9-49.

[127] United Nations General Assembly 'Disinformation and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/47/25' 13 April 2021 http://undocs.org/en/A/HRC/47/25 (accessed: 26 August 2021) The concept of information disorder is further discussed in section 3.3.1 below.

© University of Pretoria

and insights that could assist in combating disinformation. It notes that while information disorder infringes on the right to opinion,[128] the right to express freely is not limited to true statements.[129] It further notes that counter-speech and plurality of ideas, which is best ensured through more speech, can combat information disorder not restriction of speech. It however notes that restriction of the right is possible only in instances where the four-part test of legality, legitimacy, necessity and proportionality are complied with, for example in cases of ensuring the integrity of elections.

It further identified state-sponsored disinformation, Internet shutdowns and criminal laws as some of the major concerns in the regulation of disinformation. At this point, the reference to criminal laws as means of regulating online disinformation in the report sits at the centre of this chapter on how these laws are tell-tale signs of colonialism in most African countries. It noted that:

> Many of these "false news" laws fail to meet the three-pronged test of legality, necessity and legitimate aims set out in article 19 (3) of the International Covenant on Civil and Political Rights. They often do not define with sufficient precision what constitutes false information or what harm they seek to prevent, nor do they require the establishment of a concrete and strong nexus between the act committed and the harm caused. Words such as "false", "fake" or "biased" are used without elaboration and assertions based on a circular logic are made (for example, "a statement is false if it is false or misleading, whether wholly or in part, and whether on its own or in the context in which it appears").[130]

It also noted that the advertisement-driven model of social media companies, lack of clarity in the application of their rules, failure to provide adequate remedies, lack of context, political pressure and lack of effective oversight and access to data contribute to the difficulty of regulating disinformation.

In regulating disinformation, the report notes that governments, social media companies and other actors must commit to a multistakeholder system of dialogue. It noted that criminal laws that should only be used to limit speech that incites violence, hatred or discrimination and criminal libel pasts are legacies of colonial pasts and must be repealed. Other recommendations include the need for social media platforms to base their regulation on international human rights law, more access to public information, user remedies, transparency and accountability and so on.

This report foregrounds the need to understand disinformation both as a global challenge and also as a contextual problem especially in the digital age. It is easy to superficially focus on online disinformation, especially as many African countries are currently doing, without paying close attention to the foundational problems posed by

[128] United Nations General Assembly (n 127 above) para 36.
[129] United Nations General Assembly (n 127 above) para 38.
[130] United Nations General Assembly (n 127 above) paras 52, 53.

© University of Pretoria

colonial laws in African contexts.[131] It also sets the parameters of regulating disinformation, by extension, information disorder, within set international human rights standards.[132] These parameters include the application of the four-part test (legality, legitimacy, proportionality and necessity) and the roles of state and non-state actors in regulating information disorder. The essence of the test is to determine the direct relationship between 'the speech and harm, and the severity and immediacy of the harm' and using the least restrictive means to protect against such harm. On the roles of actors, it identifies the multistakeholder approach where actors are able to meaningfully contribute to the development of standards that regulate information disorder.

ii. Gender justice and freedom of opinion and expression

One of the reports focuses on gender justice and freedom of opinion and expression.[133] The report can be divided into three broad parts that focus on barriers women experience in exercising their right to freedom of expression online, the roles actors play in such experiences and the recommendations on how these actors can ensure more protection of women's expression online. Some of the issues identified in the report include how traditional and offline prejudices like cultural norms and laws limit women's expression, the harmful effects of online practices like manipulation of online content against women, sexist hate speech and disinformation, gender digital divide, increased attacks on female journalists and many others.[134]

The report also notes that states must reform their laws to ensure more equality and expression for women, ensure more access to information, comply with the limitative provisions of article 19(3) of the ICCPR, effective regulation of online gender-based violence using a delicate mix of criminal, legal, administrative and social responses. It also requires that in order to combat gendered disinformation, more diversity of voices and research would be necessary. For social media companies, there is a need to ensure more safety tools for women, apply context to their moderation systems and remove gender biases in their decision-making processes. It also identified the engagement-driven business model of most platforms, their lack of clarity on their remedial processes, less privacy-enhancing tools, lack of transparency and accountability and less focus on gender-sensitive environment as issues that engender online violence against women.[135]

---

[131] As above.
[132] United Nations General Assembly (n 127 above) paras 83-105.
[133] United Nations General Assembly 'Gender justice and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/76/258' 30 July 2021 http://undocs.org/en/A/76/258 (accessed 26 August 2021).
[134] United Nations General Assembly (n 133 above) paras 12-46.
[135] United Nations General Assembly (n 133 above) paras 47-99.

© University of Pretoria

It however noted that in order to remedy these issues, actors must carry out specific responsibilities but with women being the centre of designing such remedies. For example, states are required to make laws that specifically address online gender-based violence against women but such law must comply with the provisions of article 19(3) of the ICCPR. It also notes that the international community, led by human rights bodies should set the tone on gender-sensitive interpretation on the right to freedom of expression. Both social media and traditional media are required to carry out human rights-based and gender assessments of their policies, ensure more transparency and accountability and the safety of women journalists.[136]

The report points to two major issues worthy of note. One, it demonstrates that the right to freedom of expression is not enjoyed equally by all and is disproportionately limited based on gender. Two, it notes that the right to freedom of expression is required for vulnerable persons which also includes women, sexual minorities, persons living with disabilities, migrants, refugees, asylum seekers and others. This requirement shows a normative gap that needs to be filled specifically on how these groups of persons enjoy their rights to freedom of expression online.

iii. Online hate speech

The Special Rapporteur's report on online hate speech has a background in an earlier report from the same office in 2012.[137] The report which also highlights the impacts of new technologies on the right to freedom of expression alongside the extent of restrictions on the Internet opened with a global context on how hate speech has become more accentuated by several factors like, rising immigration flows, declining domestic economies, growing incidents of terrorisms have all placed certain groups under the threats of violence through speech.

It also recognised the need for laws to strike a fair balance between the need to protect the right and also protect against such harm that may be occasioned as a result of hate speech. Particularly, it considers the need to have laws that seek to strike this balance to comply with internationally set standards on the right to freedom of expression and information. Combining both the provisions of article 19 and 20 of the ICCPR with article 4 of the ICERD, it restated that the three-part test must be complied with. Drawing its inference from article 20(2) of the ICCPR specifically, it stated that:

> it is important to establish a clearer understanding of the terms to prevent any misapplication of the law. This formulation includes three key elements: first, only advocacy of hatred is covered;

[136] United Nations General Assembly (n 133 above) paras 100-122.
[137] United Nations General Assembly 'Hate speech, incitement to hatred and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/67/357' 9 October 2019 http://undocs.org/en/A/67/357 (accessed 22 August 2020).

© University of Pretoria

> second, hatred must amount to advocacy which constitutes incitement, rather than incitement alone; and third, such incitement must lead to one of the listed results, namely discrimination, hostility or violence. As such, advocacy of hatred on the basis of national, racial or religious grounds is not an offence in itself. Such advocacy becomes an offence only when it also constitutes incitement to discrimination, hostility or violence, or when the speaker seeks to provoke reactions on the part of the audience.[138]

A careful read of this part of the report shows there are three cumulative requirements for hate speech under the ICCPR. Not only must such speech advocate hatred, it must include incitement and not only incitement alone, such advocacy must lead to any of discrimination, hostility or violence. It need not be the three at once, one of the three requirements is enough for it to qualify as hate speech under international law.[139] Therefore, in the latest report on online hate speech, it is stated:

> A person who is not advocating hatred that constitutes incitement to discrimination, hostility or violence, for example, a person advocating a minority or even offensive interpretation of a religious tenet or historical event, or a person sharing examples of hatred and incitement to report on or raise awareness of the issue, is not to be silenced under article 20 (or any other provision of human rights law). Such expression is to be protected by the State, even if the State disagrees with or is offended by the expression. There is no "heckler's veto" in international human rights law.[140]

For avoidance of doubt, the report clearly defined keywords like 'hatred', 'advocacy', 'incitement' and many others in order not to allow for arbitrary construction of these words that have been used to define hate speech under international law by states.[141]

In determining the scope of hate speech under international law, the report considered six criteria by which a speech may be determined to be hateful. They are the prevalent socio-political context, the status of the speaker, intent, content and form of speech, extent and reach of speech and the likelihood or imminence of such speech becoming harmful.[142] Considering these tests, it may be argued that 'hate speech' is way beyond a feeling or an emotional displacement as a result of another's speech but rather, it is such speech that is closely associated with harm or possible harm. In establishing intent, the report focuses more on the severity of what was said together with the harm being advocated for and the means through which such speech is spread.

---

[138] United Nations General Assembly (n 137 above), para 24.
[139] United Nations General Assembly 'Online hate speech and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/74/486' 7 September 2012 http://undocs.org/en/A/74/486 (accessed 26 August 2020), para 8.
[140] United Nations General Assembly (n 139 above) para 8; The 'heckler's veto' referred to in the report is similar in approach to JS Mill's thoughts on the clear limitation of state authority in interfering with the right to freedom of expression, opinion and information.
[141] United Nations General Assembly (n 139 above) para 13.
[142] United Nations General Assembly (n 139 above) para 14.

© University of Pretoria

In essence, for a speech to be considered hateful, the harm advocated for through it must not only be intense and severe, there must be a strong link between such speech and the harm being advocated for. It is clear from that above that not only must the laws that look to establish hate speech as a criminal offence ensure that these tests are complied with, the contextual application on issues of hate speech must also be made to pass this test and not necessarily the subjective or arbitrary construction of state actors. The 2012 report may be said to have formed the basis upon which the report which further considers the dynamics of hate speech in the context of the Internet and other new technologies was carried out in 2019.

The major point of difference between the 2012 and 2019 reports on hate speech is that while the former focuses on the traditional concepts and application of hate speech under international law, the latter does the same but more within online contexts. For example, the 2019 report stated that state actions like Internet shutdowns, criminalisation of online political dissent or government criticisms are inconsistent with the provisions of international human rights law. It also pointed out that when looking to outsource regulations to online platforms through 'intermediary liability' laws, such laws must guard against the high chances of over-regulation, censorship and violation of free speech by strictly adhering to the provisions laid down under article 19(3), and article 20 of the ICCPR and article 4 of the ICERD. In addition to these, the report detailed the responsibilities of both state parties and companies involved in content regulation with respect to the limitative three-part test and how it applies to regulating hate speech online.

For state parties, the first requirement is that such regulation must be legal. Under such law, the kind of speech that is unlawful must be formulated with sufficient precision such that it is clear and unambiguous for companies or private citizens to understand.[143] Second, for state parties to comply with the necessity and proportionality principle under relevant international law treaties, it is important to note that pre-publication of likely harmful content is 'ill-advised.'[144] This note is important because of state parties who put pressure on companies to proactively remove harmful content thereby encouraging censorship. The report argues:

> Problematically, an upload filter requirement 'would enable the blocking of content without any form of due process even before it is published, reversing the well- established presumption that States, not individuals, bear the burden of justifying restrictions on freedom of expression.' Because such filters are notoriously unable to address the kind of natural language that typically constitutes hateful content, they can cause significant disproportionate outcomes. Furthermore, there is research suggesting that such filters disproportionately harm historically underrepresented communities.[145]

[143] United Nations General Assembly (n 139 above) para 14.
[144] United Nations General Assembly (n 139 above) para 35.
[145] United Nations General Assembly (n 139 above) para 34.

© University of Pretoria

Lastly, in complying with the principle of legitimacy, state parties who categorise 'hatred against the regime' or 'subversion of state power' do so unlawfully as they are inconsistent with the provisions of article 19(3) of the ICCPR.[146]

With respect to companies and their application of the limitative test, the report highlights that the use of artificial intelligence tools to identify online hate speech might be problematic in that while these tools may understand words and analysis which may be regarded as key factors in understanding patterns in hate speech, they often lack context which is also an important aspect of categorising hate speech as one.[147] Applying the legality standard to companies' content moderation policies, the report stated that most of these policies are vague and ambiguous. The report advises that companies content moderation policies may be improved by knowing who protected groups are, what speech violates their rules, the nature of hate speech they restrict and categories of whom hate speech rules may or may not apply.[148] In addition to these, it is argued that when it is possible that a company is unable to defer to international law on content moderation policies, it needs state so in advance, explain such variation and also its justification.

Also, in determining the application of the principle of necessity and proportionality, companies should focus more on the least intrusive means given a particular need for regulating content. Referring to Evelyn Aswad's paper 'the future of freedom of expression online,' the report identified three steps in ensuring such least intrusive measure which are to:

> evaluate the tools it has available to protect a legitimate objective without interfering with the speech itself; identify the tool that least intrudes on speech; and assess whether and demonstrate that the measure it selects actually achieves its goals.[149]

In remedying situations of online harms, the report advocated for several examples of mitigation of the effects of such harms including having graduated responses based on severity of violation, developing strong products that protect users, ensuring that users are educated about their policies and several others.[150]

---

[146] United Nations General Assembly (n 139 above) para 39.
[147] United Nations General Assembly (n 139 above) para 50.
[148] United Nations General Assembly (n 139 above) para 47.
[149] United Nations General Assembly (n 139 above) para 52; See EM Aswad 'The future of freedom of expression online' (2018) 17 *Duke Law and Technology Review* 26-70.
[150] Aswad (n 149 above) 27-67.

© University of Pretoria

iv. Regulation of user-generated content

The first report on user-generated content is broadly divided into two parts in its scope: state and private sector obligations in regulating user-generated content.[151] Drawing on the provisions of Principle 3 of the Guiding Principles on Business and Human Rights, the report emphasises that states must ensure an enabling environment for businesses to respect human rights while re-stating the three substantive limitative requirements under article 19(3): legality; legitimacy; necessity and proportionality.[152] At this point, it may be necessary to draw on the provision of 'special duties and responsibilities' as provided for under the ICCPR on the need to carry out such measures for online media platforms which not only includes not passing laws that further puts needless pressure on these platforms but also those that are only necessary, proportionate and legitimate.

The likely confusion of whether such 'special duties and responsibilities' may refer to private companies is therefore unfounded, as it has been expressed in the past through these reports and other academic scholars that such duties and responsibilities are those that are put in place by state parties for private actors to regulate their activities, but which must also fall within the provisions of article 19(3).[153] Identifying ways through which state's actions may affect online content regulation, the report rightly pointed out the use of vague laws to restrict freedom of expression, thereby obscuring the clear responsibilities for the private sector on how to adequately engage more pertinent issues for regulation like 'representations of child sexual abuse, direct and credible threats of harm and incitement to violence'[154] which are also required to comply with international law.

These restrictions have also been found to not only obscure the need to comply with international law, but have also introduced new and dangerous perspectives like the regulation of 'false information' into many companies' compliance requirements. Government tactics have also involved the call for extra-territorial and global content removals and removals not founded in law. Coupled with these, is the unusual pressure being placed on these private companies to remove content which often results in over-removal of permissible expression that lead to both state-initiated and private sector-enabled censorship.[155]

---

[151] United Nations General Assembly 'Online content regulation and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/38/35' 6 April 2018 http://undocs.org/en/A/HRC/38/35 (accessed 26 August 2021).
[152] United Nations General Assembly (n 151 above) para 7.
[153] United Nations General Assembly (n 151 above) para 6.
[154] United Nations General Assembly (n 151 above) para 13.
[155] United Nations General Assembly (n 151 above) para 17.

© University of Pretoria

The second part of the report focuses on the obligations of companies to follow these laid down requirements as must have been followed by the state parties under the law. According to the report, 'few companies apply human rights principles in their operations, and most that do see them as limited to how they respond to government threats and demands.'[156] It was pointed out that the acts of state parties through these dangerous laws seems to have emboldened the practice by companies to defer to local laws thereby avoiding complications with local authorities even if it means violating human rights in the process. As evidence of this claim, Facebook, a major social networking site has stated, 'if, after careful legal review, we determine that the content is illegal under local law, then we make it unavailable in the relevant country or territory.'[157] In raising more specific issues on how the activities of the private companies affect the protection of the right to freedom of expression, the report identified areas of concern on the content regulation standards by companies.

It also considered how the rules applied to content moderation on these platforms are vague, subjective and are capable of being subjected to arbitrary meanings which pose great dangers to free speech. Facebook, Twitter, YouTube and a host of other social media platforms have these vague provisions on issues of violence, extremism, incitement, hate speech and other online harms which have been used as examples in the report.[158] Perhaps, one of the major bases for platforms' compliance with state parties on content regulation has been the reason to do so due to context. The report argued that despite such claims to apply context, it has not reduced the illegal removals of content. Also, companies claim that they require more context in their community-driven regulation practices, but how this context is then achieved in the final decision-making process is unclear. Also, the report identified the issues of anonymity in carrying out their responsibilities on content regulation.

The point on anonymity was argued in the report that such a requirement by companies to demand real names may pose huge risks to the right to freedom of expression specifically and Internet freedoms in general. One of the thorny issues that the report also shone lights on with respect to the potentially dangerous activities of companies on content regulation is disinformation.[159] It acknowledged that while companies may have taken laudable steps in striking a balance between public interests and protecting free speech on their platforms, they still pose threats to alternative sources of media and reduce the ability of the larger public to make informed decisions based on available facts. Also, it recognised the ways through which state parties may be encouraged to develop the culture that companies, through

[156] United Nations General Assembly (n 151 above) para 10.
[157] United Nations General Assembly (n 151 above) para 22; Facebook 'What is a legal restriction on access to content on Facebook' https://www.facebook.com/help/1601435423440616?helpref=related (accessed 15 June 2019).
[158] United Nations (n 151 above) paras 26-31.
[159] United Nations General Assembly (n 151 above) para 31.

© University of Pretoria

their platforms and technologies they deploy, have the lasting solutions to a socially dynamic issue like disinformation.[160]

The nature of expression that is prohibited under international human rights law whether online or offline are 'child pornography; direct and public incitement to genocide; advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence; and incitement to terrorism.'[161] Therefore, while online harms include prohibited speech, not all online harms are prohibited speech. As they will be discussed in the subsequent chapter, online harms are not limited to prohibited expressions but also include information disorder and online gender-based violence that do not fall strictly within the scope of online hate speech that could be harmful.[162]

B The mandate of the Special Rapporteur on the Right to Freedom of Expression and Access to Information in Africa and recent developments in the digital age

In 2002, in one of its roles made pursuant to article 45 of the African Charter, the African Commission adopted a Resolution on article 9 referred to as the Declaration of Principles on Freedom of Expression in Africa at its 32nd Ordinary Session. This Declaration presented stakeholders in Africa who work on the protection and promotion of the right to freedom of expression and information with the opportunity to advocate with a home-grown directional policy on the scope and meaning of the right to freedom of expression. These efforts further culminated into the establishment of the mandate of the Special Rapporteur on Freedom of Expression in 2004 at the Commission's 36th Ordinary Session. The scope of the mandate of the Special Rapporteur was expanded in 2007. The mandate of the Special Rapporteur as provided for are analysis of national media regulation, fact-finding missions to member states, undertake promotional country missions, make public interventions on violations of the right, keep a record of these violations and submit reports to the Ordinary Session of the African Commission.[163]

In 2012, also through a Resolution, the African Commission revised the Declaration to include access to information.[164] The Declaration focused largely on the right to freedom of expression as a right, however, in such a manner that it cannot be easily divorceable from the right to information. In terms of its standard-setting norms on both

[160] As above.
[161] United Nations General Assembly 'The right to freedom of opinion and expression exercised through the Internet, A/HRC/66/290' 10 August 2011 http://undocs.org/en/A/66/290 paras 20-36 (accessed 1 December 2021).
[162] The classification of online harms is further discussed under section 3.5 below.
[163] African Commission (n 108 above).
[164] African Commission Resolution, ACHPR/Res.222(LI)2012, 2 May 2012.

© University of Pretoria

rights, the Declaration provided for access to information by state members to public information, private and public media ownership, print media, media plurality, broadcast and telecommunications, criminal measures on speech and many other topical issues that constituted challenges to the protection and promotion of the right.[165] Notably, just like the other regional and global mechanisms, there were no meaningful directions on the right especially in relation with new technologies until the turn of new decade in 2011. It was at this point that the plane for protection of the right, with respect to freedom of expression and new technologies experienced an upward trajectory through the General Comment 34.

The African experience, at least regionally with this upward trajectory was seen through a number of initiatives. These include joint declarations by the AU Special Rapporteur with her other regional and UN counterparts,[166] resolutions by the African Commission,[167] press releases,[168] and a Declaration revised in 2019 which has now incorporated the current realities on new technologies in the promotion, protection and interpretation of the right to freedom of expression and access to information in Africa.[169] This new Declaration replaces the 2002 version and so far, may be regarded as the most direct and binding instrument on the right to freedom of expression online in Africa. The Declaration may be regarded as a fulfilment of the hopes of correcting the weak provisions on the right under the African Charter. To Welch:

> a vague or weakly-worded treaty can be developed or interpreted over time if the political will is present. The limitations of the African Charter are striking; even more, in the case of freedom of expression, the political will to interpret the wording of the African Charter broadly has not been present.[170]

Perhaps in remedying this observation, the Declaration has provided interesting and elaborate clarity on some thorny issues on protecting the right to freedom of expression while also developing the scope of the right under the Charter. The nature of the Declaration may be assessed in two key ways.

First, it is direct because it is sourced from the provisioned mechanisms provided for in the African Charter like articles 9, 45 and 60. Article 9 as highlighted above provides for the substantive right, article 45 provides for the norm-setting responsibilities of the African Commission through the mandate of the Special Rapporteur and article 60

---

[165] See Declaration of Principles on Freedom of Expression in Africa 2002 https://www.achpr.org/presspublic/publication?id=3 (accessed 15 March 2020).
[166] OSCE (n 107 above).
[167] African Commission on Human and Peoples' Rights 'Documentation Centre' https://www.achpr.org/documentationcenter (accessed 1 December 2021).
[168] As above.
[169] African Commission on Human and Peoples' Rights 'Declaration of Principles on Freedom of Expression and Access to Information in Africa 2019' https://www.achpr.org/presspublic/publication?id=80 (accessed 1 December 2021).
[170] Welch (n 114 above) 113.

© University of Pretoria

provides an inter-connection between both the substantive right, the norm-setting responsibilities and the wider application of the international human rights treaties.

Second, the Declaration is binding because of how it ensures an active rather than passive implementation process in which soft laws are often noted for. This active implementation is evident when the cumulative provisions of Principle 43 of the Declaration mandates implementation through review of policies in order to conform with the Declaration, the Model Law on Access to Information, the Guidelines on Access to Information and Elections in Africa. Importantly, in combining implementation with effective monitoring and evaluation, it includes the provision of article 62 of the Charter which mandates the submission of periodic reports on measures taken to comply with the Declaration.

Perhaps, the most defining feature of the new Declaration is how it is able to combine the old Declaration with its new objectives of setting standards on the right to freedom of expression and information and new technologies. In demonstrating this feature, the Declaration looks to plug existing gaps in the application of the right under the African Charter by providing for corrective provisions. The corrective role of the Declaration is best understood as amendments of the substantive right to freedom of expression and information in Africa. For the corrective provisions, the Declaration provided for both the provisions of prohibited speech in the ICCPR and ICERD while also incorporating the basic principles of non-discrimination against persons living with disability and children under the CRPD and CRC respectively.[171] Also, as noted above on the scope of protection afforded the right to freedom of expression and information under both the African Charter and the ICCPR, the right to hold an opinion is qualified. However, as one of the corrective features of the Declaration, Principle 2 explicitly provided that the freedom to hold an opinion shall not be interfered with by states.[172] This provision rests the debate as to whether the right to hold an opinion under the African Charter or instruments made pursuant to it is qualified or not.

Second, Principles 22 and 23 lay down the conditions that must be met for justifiable limitation of the right to freedom of expression and information, which was not provided for under the African Charter. For example, under Principle 22, three additional provisions were made to include the requirement of states to repeal insult and false news laws; decriminalisation of defamation and libel and non-imposition of custodial sentences for defamation offences.[173] Currently, these provisions are provided for in most criminal codes and later in cybercrime and electronic communications laws in African laws when they do not comply with international human rights standards. The only identified legitimate restriction of speech is speech provided for in article 20 of the

[171] African Commission on Human and Peoples' Rights (n 169 above) Principle 3.
[172] African Commission on Human and Peoples' Rights (n 169 above) Principle 2.
[173] African Commission on Human and Peoples' Rights (n 169 above) Principle 22(2), (3) & (4).

© University of Pretoria

ICCPR, 4 of ICERD and more recently as it applies to the African human rights system, Principle 23 of the African human rights system, which expressly prohibits speech that advocates violence.[174] These two major provisions provide a basis for states to reform various laws that bear on information disorder and online violence and hate speech in the African context.

Currently, many African countries still provide for the criminal offences sedition, insult or false news laws in their legal systems and they manifest in such a manner that their foundations were laid by criminal codes with colonial backgrounds. It is this background that finds its way into other laws that currently affect freedom of expression online in Africa. Also, the explicit provision on decriminalisation of defamation will not only embolden the regional courts which have been forward-looking in their protection of the right but also provide a framework for holistic reform in the Internet rights and policy sector in Africa.[175] These new provisions, together with the previous ones have solidified the regional jurisprudence on what constitutes justifiable limitation of the right to freedom of expression and information, especially with respect to criminalisation. Also, Principle 23 addresses the lacuna of prohibited speech under Article 20 of the ICCPR which provides for the limitation of the right through hateful speech. It does not only establish the limitation; it also explains the cumulative conditions that must be fulfilled before such limitation may be applied. Principle 9 provides for a more general, elaborate and conjunctive application of justifiable limitations in national contexts with respect to the right. It does not only restate the three-part test on justifiable limitations, it further breaks down each test and how they can be practically applied.

In addition to these, the Declaration also carried out a unique standard-setting role. For example, Principle 4 clearly settles the debate of whether 'within the law' as provided for under the African Charter refers to national law and international law. It provides that where there seems to be a conflict between both systems, the international law system takes precedence on the protection of the right.[176] This precedence is so that national laws must be brought in line with the international system. Also, the Declaration introduced self-regulation and co-regulation of the media under Principle 16 which sets the standards beyond the traditional approach of regulation through the law by states. Another unique norm-setting role under the Declaration may be found in the provisions of Principle 17(4) which offers beyond co-regulation, self-regulation and traditional regulations. It made it a requirement for

[174] African Commission on Human and Peoples' Rights (n 169 above) Principle 23.
[175] African Commission on Human and Peoples' Rights (n 169 above) Principle 22(3).
[176] African Charter on Human and Peoples' Rights, arts 60 & 61.

© University of Pretoria

states to develop a multistakeholder regulatory approach for broadcast, telecommunications and the Internet regulatory framework.[177]

Perhaps, the most defining provision in the Declaration with respect to this chapter is the provisions of Principle 39 which regulates the relationship between states and Internet intermediaries. The provision protects issues like freedom of expression and access to information online, rights-respecting content moderation policies and new technologies, transparency while safeguarding human rights online and several others.

The contribution of the African human rights system to the international human rights system has been discussed in the past by scholars.[178] The discussion of how these contributions have also been underutilised by the international human rights system has also been focused on.[179] However, how the African human rights system, both from its past indigenous human rights culture and now its regional designs have been underutilised by African countries have not been adequately discussed. According to Welch, 'because Africa was subjected to a particularly strong, intense form of colonial rule, individual governments were endowed with powerful means of restraining the media and restricting freedom of expression.' Not only do these laws have colonial foundations, they are now being transplanted into laws regulating cyberspace in most African countries such that freedom of expression online is now at risk. This transplantation, viewed from the lens of postcolonial legal theory demonstrates that the violation of the right to freedom of expression online has its deep feeder taproot in colonial laws, underscoring the importance of applying critical legal studies in appraising the right in Africa.

2.5 Reliving the past through digital colonialism: The 1892 Gold Coast Criminal Code, electronic communication laws and the protection of the right to freedom expression online in Africa

The theoretical and recent normative development on the right to freedom of expression are clear that the right is not absolute and can be limited in clear instances. These limitations are also clear and provide directions to states because states have

[177] The report that reviewed multistakeholder initiatives notes that it might not have been effective according to that study but it does not mean it has not worked and upped the ante with respect to rights protection. See Institute for Multi-Stakeholder Initiative 'Not Fit-for-Purpose: the grand experiment of multi-stakeholder initiatives in corporate accountability, human rights and global governance (2020) https://www.msi-integrity.org/wp-content/uploads/2020/07/MSI_Not_Fit_For_Purpose_FORWEBSITE.FINAL_.pdf (accessed 15 July 2020).

[178] See R Murray 'International human rights: Neglect of perspectives from African institutions' (2006) 55 *The International and Comparative Law Quarterly* 193-204.

[179] As above.

© University of Pretoria

the primary responsibility to protect it. However, this clarity, particularly on the permissible limitations of the right do not feature in national contexts in Africa. Most laws and state policies that impact the right are often at odds with the requirements of international human rights law, especially on how to limit online speech. One of the reasons for this impact are colonial legacies laid through problematic laws that sought to limit expression against colonial governments and their allies. These laws have extant criminal provisions on 'publication of false information' 'blasphemy', 'libel', 'slander', 'sedition' and others. These provisions, which remained on the books even after the end of colonial rule, have now seeped into new and proposed laws that seek to limit online expression in African countries.[180]

2.5.1 The connection between colonial legacies and electronic communication laws in African national contexts

While colonial legal legacies are not the only reasons for the violation of the right in African countries, they are the main reason problematic colonial legal provisions have found their way into their cyber laws. These colonial impacts on legal systems in Africa introduced many alien concepts of criminalisation of speech through provisions on 'insults', 'sedition', 'criminal defamation' and 'publication of false information' that are not only still existing in most criminal and penal codes, but are now found in the cybercrime and electronic communications law in Africa. A point to note is that while the colonial systems that introduced these harms into the legal systems have since done away with such laws,[181] African governments still hold on to them thereby violating the right to freedom of expression and access to information in their respective contexts.

These provisions have continued to shape African states' response to the protection of the right to freedom of expression despite their transition to democratic constitutions and multiparty systems in the early 1990s. They have been used to stifle dissent and curtail the enjoyment of other human rights that are ancillary to freedom of expression. According to Howard, 'the idea that the African press should not be critical of the established government is thus a direct legacy of the colonial period'[182] and while Howard's view might not totally be the case, it is a fact that colonial laws were established to discourage dissent or criticism against colonial powers.[183] It is these codes that then inform the provisions used to regulate online expression as evident in many countries with cybercrime and electronic communication laws in Africa. These laws feed off the carcass of the colonial criminal legal system on illegitimate restriction of speech as those mentioned above. Typically, these laws replaced the statutes of

[180] United Nations General Assembly (n 127 above) para 13.
[181] M Kanna 'Furthering decolonization: Judicial review of colonial criminal laws' 70 *Duke Law Journal* 424.
[182] RE Howard *Human rights in Commonwealth Africa* (1986) 121.
[183] JA Dada 'Human rights protection in Nigeria: the past, the present and goals for role actors for the future' (2013) 14 *Journal of Law, Policy and Globalisation* 1-13.

© University of Pretoria

general application and were received into all of the British colonies in the continent save for those in North Africa. As Morris observed:

> By 1935 there was throughout the area under review (with the partial exception of Sierra Leone) a body of criminal law and procedure of very similar, or actually identical, origin. All the Criminal (or Penal) Codes, with the exception of that of the Gold Coast (derived from St. Lucia), had an original source in the Queensland Code of 1899. Nevertheless, overall the basic homogeneity of the criminal law and procedure in this large area of Africa remains. [184]

According to Morris, the sources and homogeneity still exists in today's criminal legal system in Africa especially in relation to provisions of criminal codes that impact the right to freedom of expression. There were records that these criminal codes were opposed by most of Africa's intellectuals at that time due to the arbitrary nature of its application in the various systems they were introduced into.[185] Except for Sierra Leone, the opposition was overthrown and these laws were not only introduced, it has since been in use while also setting new patterns in the restriction of freedoms in the digital age. African countries affected by these legacies lie to the South of the Sahara. The first country to have a Criminal Code was the Gold Coast (Ghana) in 1892. The project of having these colonies' Criminal Codes bear the stamp of the colonial system was mostly completed by 1935, when each country under the British colonial system had a Criminal Code or a Penal Code while a country like Nigeria had both.

In the Nigerian context, the British had introduced the Criminal Code in the Northern protectorate in 1904 and later applied the Code to the entire country when the Northern protectorate was merged with the Southern protectorate in 1914.[186] Due to conflicting provisions of the Criminal Code with Islamic law, which the Northern protectorate largely practices, a new Penal Code which was already in force in Sudan and India was introduced in the North in 1960 while the Criminal Code remained in force in the Southern parts of the country.[187] It is important to note that the Penal Codes in force in Sudan and India were also introduced by the British. Current provisions of both the Criminal and Penal Codes in Nigeria deal with publication of false news intended to cause alarm, abuse and insulting language, sedition and criminal defamation.

For South Africa, its criminal law has been influenced mainly by Roman-Dutch and British legal traditions. The Roman-Dutch influence was as a result of the arrival of Dutch colonisers in the mid-seventeenth century in South Africa while the British influence was as a result of the defeat of Dutch colonisers by British colonisers in the

---

[184] HF Morris 'A History of the adoption of codes of criminal law and procedure in British Colonial Africa, 1876-1935' (1974) 18 *Journal of African Law* 23.
[185] Morris (n above 184) 6.
[186] HF Morris 'How Nigeria got its Criminal Code' (1970) 14 *Journal of African Law* 137-154.
[187] VLK Essien 'The Northern Nigeria Penal Code: A reflection of diverse values in penal legislation' (1985) 5 *New York Law School Journal of International and Comparative Law* 89.

© University of Pretoria

early nineteenth century.[188] Due to inadequacies of Roman-Dutch legal principles in some areas like criminalisation of sedition, British laws were often used to fill in the gaps including gaps in criminal law.[189] For example, the Peace Preservation Ordinance of 1902 on sedition was based on English law.[190] The draft section 29 of the Native Administrative Act of 1927 also provided for the crime of sedition, fashioned along the 'formula of the English law.'[191] While the Peace Preservation Ordinance of 1902 and the Native Administrative Act of 1927 are no longer in force, the Criminal Procedure Act of 1977 Africa is still in force with respect to court proceedings on criminal matters. Sections 104 & 242 of the Criminal Procedure Act of 1977 provide for criminal proceedings with respect to blasphemous, seditious, obscene or defamatory matter. However, there have not been any criminal cases on blasphemy, sedition, obscenity or defamation before South African courts in recent times.

Nigeria, South Africa and other African countries have therefore been impacted directly by British colonial rule and they all have the common origin of the 1892 Gold Coast Criminal Code which laid the foundation of the formal British colonial criminal system. It is this origin that laid the foundations for violations of free speech, which has been one of the biggest limiters of democratic development in Africa in recent times.[192] The effect of this Code became easily multiplied as it spread across other British colonies and still persists till today. Not only does the foundations laid by the Code persist till today, it has taken on new ramifications through laws restricting free speech online in many African countries. These African countries may be divided into three broad categories when considering the impact of colonial laws on freedom of expression online. There are linear, semi-linear and non-linear systems.

A Linear systems

The linear systems are African countries whose legal system had direct contact with colonial laws that impact on freedom of expression and such contact continues till date. As an example of a linear system that has been impacted by colonial laws, in Kenya, despite the constitutional protection of both the right to freedom of expression and access to information, and accession to both the ICCPR and the African Charter, its Penal Code provide for illegitimate restrictions on the right to freedom of expression. Sections 52, 53, 56, 57, 132, 194, 195, of the Penal Code provide for the various

[188] M Chanock *The making of South African legal culture* (2004) 3.
[189] Vindex 'The suggested repeal of Roman-Dutch law in South Africa' (1901) 18 *South African Law Journal* 153.
[190] GG Gardiner & CW Lansdown *South African Criminal Law and Procedure* (1924) 7-8.
[191] Chanock (n 188 above) 147.
[192] AJ Jeffrey 'Media freedom in an African state: Nigerian law in its historical and constitutional context' unpublished PhD thesis, University of London, 1983 58.

© University of Pretoria

speech offences which have all been identified as posing threats to freedom of expression.[193]

While the High Court has ruled on the unconstitutionality of the provisions of section 194 on criminal libel in Kenya, the necessary amendments on the Penal Code have not been carried out by the legislature.[194] In the same vein, the provisions of sections 3, 4, 5 on slander also point to the problematic provisions on illegitimate restrictions as they do not fall under the envisaged limitations of freedom of expression under the international human rights system. In what may be referred to as the offshoot of these laws that have since been projected online is the Kenya's Computer Misuse and Cybercrimes Act of 2016. Sections 22, 23 and 27 provide for the offences of false information, publication of false information and cyber harassment with excessive punitive punishments of what should not be an offence in the first place according to international human rights law. This example is also found in Nigeria, which is one of the case studies of this thesis, Tanzania, Uganda and other legal systems in Africa.[195]

B Semi-linear systems

The semi-linear systems are those countries that while they did not have colonial laws directly introduced into their systems, or once had but reviewed their laws, their systems are still influenced by experiences from colonial systems in the framing of their laws on limitation of free speech. An example of a semi-linear country is Ethiopia.

Ethiopia demonstrates one of such countries who even though was not colonised, was largely influenced by colonial legal structures especially in illegitimate restriction of free speech. Despite its constitutional provisions and accession to applicable international human rights treaties, Ethiopia still has laws with offences like criminal defamation, insults and offences against national interests under Chapter Two of its Penal Code which deals with injuries to honour.[196] Also, the Mass Media and Freedom

---

[193] B Rickcard 'Words that started a riot: An appraisal of the law against sedition and criminal libel in Kenya' (2019) https://bit.ly/3ALym00 (accessed 19 October 2020).

[194] *Jacqueline Okuta & another v Attorney General & two others* (2017) Petition 397 of 2016 eKLR http://kenyalaw.org/caselaw/cases/view/130781/index.php?id=3479 (accessed 24 October 2020).

[195] For Nigeria, see section 5.2.1 below; for Tanzania's colonial provisions, see section 55 (seditious intention), 63B (raising discontent or ill-will for unlawful purposes), 63C (hate speech), 89 (abusive language) and 125 (insulting to religion) of the revised edition of the Penal Code of Tanzania 2019 which was first adopted in 1945. For similar provisions in Tanzania's Cybercrime Act of 2015, see sections 16 (publication of false information), 17 (racist and xenophobic material), 18 (racist and xenophobic motivated insults), 23 (cyberbullying); for Uganda's colonial provisions, see sections 39 (seditious intention), 40 (seditious offences), 50 (publication of false news), 118 (insults to religion), 179-182 (criminal defamation) of the Penal Code Act of 1950. For similar sections in Uganda's Computer Misuse Act, see sections 24 (cyberharassment), 25 (offensive communications) & 26 (cyberstalking) of the Computer Misuse Act, 2011.

[196] Penal Code of Ethiopia (1957), chapter 2.

© University of Pretoria

of Information Proclamation, in its section 41(1) links the proclamation to the Criminal Code with respect to criminal liability for defamation while (2) provides for the punishment of criminal defamation.[197]

These foundations are furthered by the provisions of sections 13 and 14 under Part 2 of the Computer Crime Proclamation No 958/2016. Section 13(3) provides for the criminalisation of defamatory statements online in Ethiopia while section 14(1) provides for a link to the provisions of the Criminal Code of Ethiopia in punishing crimes against public security.[198] Recently, Ethiopia also passed the Hate Speech and Disinformation Prevention and Suppression Proclamation (2019). Under article 2 of the Proclamation, disinformation is defined as 'speech that is false, is disseminated by a person who knew or should reasonably have known the falsity of the information and is highly likely to cause a public disturbance, riot, violence or conflict.'[199] It also defines hate speech as 'speech that promotes hatred, discrimination or attack against a person or an identifiable group based on ethnicity, religion, race, gender or disability.'[200] All these provisions are deeply problematic in that while they do not comply with internationally set standards of limiting free speech, they are also capable of being used arbitrarily by the state as has been done in the past.[201]

On the other hand, another close example of a country with a semi-linear system is Ghana. Criminal defamation, and dissemination of false information in its 1960 criminal code is still evident in its Electronic Communications Act of 2008.[202] Section 76 of the Act provides for the offence of false information, which runs contrary to the provisions of international law.[203]

C Non-linear systems

The non-linear systems are such countries who have had contact with the colonial system but have since reformed or shown signs for reforms of their laws with respect to these impacts in their laws. For the third category, South Africa, which is the other case study that this thesis focuses on, represents an example of a country of non-linear systems where even during apartheid, defamation was largely a civil matter than a criminal one.[204] Until recently, there was no law that provided for false information and this provision has since been challenged and upturned by the court.

[197] Mass Media and Freedom of Information Proclamation (1958), sec 41(1).
[198] Computer Crime Proclamation No 958/2016, arts 13 & 14.
[199] Hate Speech and Disinformation Prevention and Suppression Proclamation (2019), art 2(3).
[200] As above, 2(2).
[201] United Nations General Assembly (n 122 above).
[202] Electronic Communications Act (2008), section 76.
[203] Section 2.4 above.
[204] See D Milo *Defamation and freedom of speech* (2008).

© University of Pretoria

2.5.2 Reliving the past through a new form of colonialism

In theorising this new form of colonialism, it can be pictured as a network of rings in a circle. Colonialism is the outer and biggest circle while the next closest ring connected in the circle are colonial legal systems. The next in this ring are colonial criminal legal systems and the next after these systems is the impact of this criminal legal system on human rights. The last ring, especially as it relates to this chapter, is digital colonialism as it impacts on the right to freedom of opinion, expression and information. It suffices to state that these rings are all connected and accentuated by postcolonial legal theory which seeks to highlight the impacts of colonialism on the legal cultures of the colonised. Therefore, within the context of this chapter, digital colonialism is the colonial influences on cyber laws that seek to limit the right to freedom of expression online. These influences are found in criminal and penal codes that still provide for the offences of 'insults', 'sedition', 'criminal defamation' and 'publication of false information' which are also provided for in cybercrime and electronic communication laws that seek to regulate online speech in African countries. As it relates to African experiences, it is how colonial systems and their legacies have laid the foundations for the violations of human rights in the digital age, and in particular, with respect to the right to freedom of expression, information and opinion through laws and practices which have continued to have linear impacts on the protection of the right till the present day.[205]

The form of digital colonialism being described under this chapter is communicative, neo-colonialist and speech-focused while the mainstay concept is the sum-total of the impacts of Global North big tech companies' businesses in the Global South.[206] This form is similar to the mainstay form of digital colonialism which focuses on the expropriation of data by Global North big tech companies from the Global South – data colonialism. While this form of digital colonialism may be focused mainly on data expropriation, it is the most developed form with respect to digital colonialism studies and it is still emerging. In addition to this, the mainstay digital colonialism considers the various dimensions of impacts like governance, human rights and human development while the latter, which this chapter, just like data colonialism, seeks to establish focuses on a more specific aspect of these dimensions, its impacts on the right to freedom of expression online in Africa. It is the semi-technical application of digital colonialism as a multidimensional concept especially with how it manifests in relation to online free speech in African countries. As a result, it posits that beyond

[205] Nyabola examines the impacts of social media platforms that reinforces existing stereotypes using Kenya as an example. She points out that asides the appropriation of data of Global South citizens by many big tech companies, including social media platforms, the challenges of offline harms have been accentuated and transmuted into the online space. See N Nyabola *Digital democracy, analogue politics: How the internet era is transforming politics in Kenya* (2018) 157-178.

[206] Compare Nyabola (n 205 above) to N Couldry & UA Mejias 'Data colonialism: Rethinking big data's relation to the contemporary subject' (2019) 20(4) *Television & New Media* 339-343.

© University of Pretoria

data, digital colonialism exists in law texts and is exacerbated by platform governance as currently constituted.

It is these laws that are currently being used to regulate both offline and online speech in most African countries. It is also these laws that companies, who have been recently described by Klonick as 'new governors of speech' adhere to when applying their content moderation policies.[207] Therefore, there is a foundation laid by these colonial laws that companies who are now important stakeholders in the future of online speech refer to as 'local laws' and to which they defer.[208] What this deference means is that companies will not defer to international human rights standards in their content moderation policies especially when dealing with local contexts.[209] Most Global South countries, especially those from Africa also do not have any meaningful input to the content moderation policies of these companies who moderate globally based on limited human rights-speak.[210] This limitation also suggests that these companies do not necessarily verify whether these laws violate human rights standards. Therefore, as described above, colonial legacies, cybercrime and electronic communication laws that violate freedom of expression online and private actors triangulate above this network of rings to operationalise digital colonialism. In doing this triangulation, digital colonialism exacerbates online harms through these connected factors. For example, online harms like information disorder and targeted online violence are further amplified as a result of foundationally faulty laws and non-contextual platform governance. It paints a clear but complex picture of the future of freedom of expression in Africa, one of which relies on platform governance, the ecosystem of key stakeholders who make rules and regulations on how online content and speech are governed. It has therefore become more important to analyse the resultant impacts of this new form of colonialism, online harms and how they violate the right to freedom of expression in the region.

[207] K Klonick 'The new governors: the people, rules and processes governing online speech' (2018) 131 *Harvard Law Review* 1603.

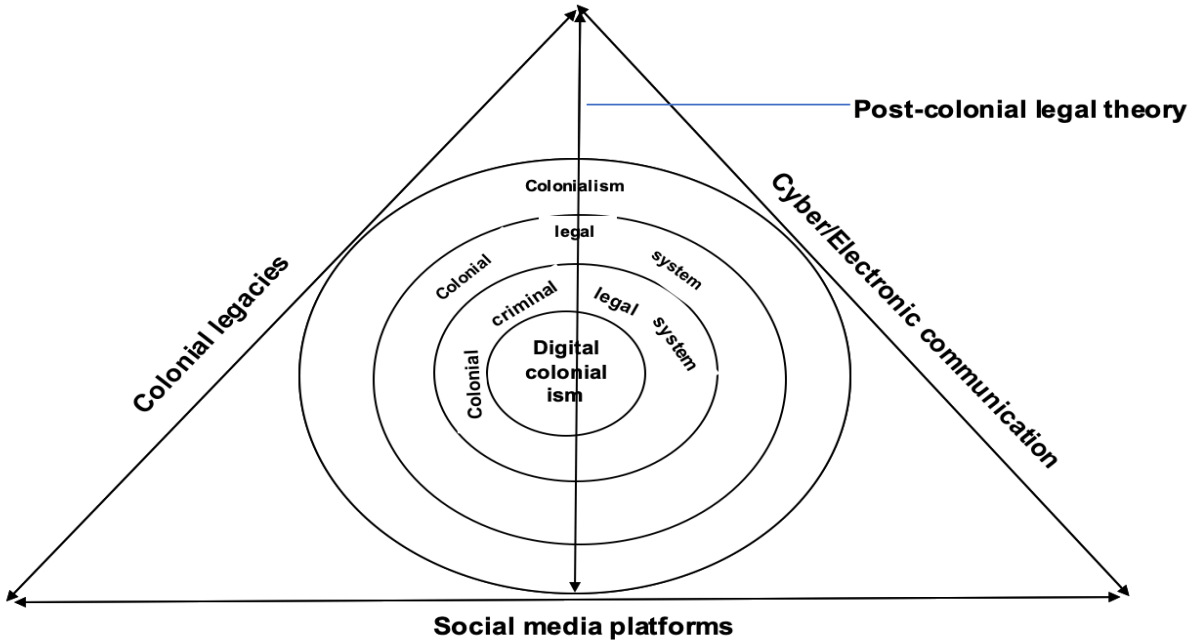
[208] Facebook (n 157 above); Twitter 'About country withheld content' https://help.twitter.com/en/rules-and-policies/tweet-withheld-by-country (accessed 12 February 2020); Facebook 'Government request to remove content' https://transparencyreport.google.com/government-removals/overview?hl=en (accessed 13 February 2020).

[209] D Kaye *Speech police: The global struggle to govern the Internet* (2018) 33-34.

[210] Klonick pointed out that most content moderation policies, especially from the likes of Facebook who has the most number of platform users are made up of Euro-American rules and that input from Global South systems were largely absent despite making up most of the users on Facebook. See K Klonick 'The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression' (2020) 129 *Yale Law Journal* 2437.

© University of Pretoria

Illustration 1: A new form of digital colonialism



2.6 Conclusion

This chapter set out to examine the extent of implementation of theoretical perspectives and recent normative developments on protecting the right to freedom of expression online in Africa. It examined the various perspectives on the right especially in indigenous Africa and later in the current African human rights system. It noted that the right to freedom of expression is not alien to African contexts – just as the right was acknowledged in the past, it is equally guaranteed now in the digital age. However, this guarantee seems to be limited and does not apply to national contexts.

It identified that despite the various developments on the protection of the right to freedom of expression online against online harms like information disorder, online violence and online hate speech under the international human rights system, some African governments still struggle with protecting the right in their various contexts. It noted that one of the reasons for this struggle is the existence of various provisions in colonial laws that violate the right to freedom of expression. These provisions which include criminalisation of insults, sedition, defamation and false information have also found their way into current laws enacted that regulate online speech.

The chapter further showed that the manner in which these provisions reach into the regulation of online expression in these countries may be attributed to a new form of colonialism. This new colonialism is an aspect of the debates on the continued repression of African systems in the digital age that is reminiscent of colonialism. It moved beyond the debates on the extraction of data by social media platforms from

© University of Pretoria

the Global South to the Global North and notes that colonialism still impacts on the right to freedom of expression online negatively in Africa. However, it noted that despite the theoretical and recent normative developments on how to protect online expression in Africa, it is still unprotected and one of the reasons for this violation are problematic colonial legal provisions which exacerbate the impacts of online harms. The impacts of these harms are examined in the next chapter.

© University of Pretoria

CHAPTER THREE: THE IMPACTS OF ONLINE HARMS ON THE RIGHT TO FREEDOM OF EXPRESSION ONLINE IN AFRICA

3.1 Introduction

In the previous chapter, the theoretical perspectives and recent normative developments on the right to freedom of expression in Africa online were discussed. It examined how various colonial foundational laws impact cyber laws and how these are the laws sought to be enforced by social media platforms. In addition, it connected the status of the right to freedom of expression online in Africa to digital colonialism. This triad – colonial laws, problematic cyberspace policies and social media companies are the major actors that engender digital colonialism which in turn makes regulating online harms difficult and pose threats to the right to freedom of expression in the region. It is these online harms, their forms, methods, classifications and impacts that this chapter seeks to examine.

Due to the dynamic nature of online activities and their increasing abilities to cause harm, it has become difficult to engage in their regulation, especially in a rights-respecting way. This regulation is especially onerous in the context of online speech which is ever-changing in context, distribution and reach. Additionally, due to the popularity of online platforms in hosting content and redefining the scope of social, economic and political frontiers which unfortunately give rise to harms, it has become important to regulate online speech. As examined in the previous chapter, the need for regulation of speech has gone beyond traditional approaches and has now included other stakeholders.[1] To Citron, 'the Internet extends the life of destructive posts,'[2] and this extension highlights the need to protect against online harms which are more organic, far-reaching and permanent in nature than traditional modes of expression.

Consequently, the goal of regulation of these harms is to strike a balance between the protection of the right to freedom of opinion, expression, access to information, and

[1] E Donahoe & FE Hampson 'Governance innovation for a connected world protecting free expression, diversity and civic engagement in the global digital ecosystem' (2018) *Centre for International Governance Innovation: Special Report* 11 https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/stanford_special_report_web.pdf (accessed 15 August 2020). Out of the three major approaches to platform governance, multi-stakeholder/user-centred approach is most favoured because it is deemed stakeholder-driven, open, transparent and consensus-based. See also K Perset *et al* 'Moving "upstream" on global platform governance' (2019) in *Models for platform governance: A CIGI essay series* https://www.cigionline.org/sites/default/files/documents/Platform-gov-WEB_VERSION.pdf 79. Some of the ways the inclusion of more stakeholders has worked include Facebook's Oversight Board and ARTICLE 19's Social Media Council.

[2] DK Citron *Hate crimes in cyberspace* (2014) 4.

© University of Pretoria

limiting credible threats to the rights of others. The need for this balance has necessitated several forms of regulatory approaches which have gone beyond the traditional model of regulation typically carried out by states. As a result, there is now an increased need for an effective regulatory approach that prevents online harms especially as they occur on social media platforms. This need is because despite states being the primary duty-bearers under international human rights law in the regulation of online speech, the online platforms, due to their increasing power, now have more responsibility to keep their platforms safe while also protecting free speech.

The harms which often manifest as information disorder and targeted online violence[3] may be broadly classified into harmful and illegal content.[4] While there have been several debates on who has what responsibilities between states and online platforms, there are currently no clearly set out and fine-tuned body of rules to regulate online harms. Also, due to the conflation of various online harms which ultimately leads to lack of clarity on what the rules are, the potential impacts of online harms especially in Africa and the possible contributory factors that accentuate them have not been critically considered. In addition, the literature on online harms is more Global North-facing than it engages alternative experiences in Global South systems. There are three major reasons for this low engagement.

First, there are inadequate academic engagements on the various impacts of digitalisation on human rights and development in regions like Africa. Second, there are limited academic literature on online harms as a concept beyond Africa and far less literature on it as an area of study in the region. Third, most African countries set policy priorities differently and as a result, this impacts the quality of debates on online harms in the region. In resolving these reasons, it is necessary to develop contextually-relevant conversations, academic, law-related or policy-related on online harms in order to address them more pointedly in Africa. In order to properly engage online harms and understand that its impacts are far-reaching due to their amplification through digital means, Wasserman argues that they need to be situated within specific contexts.[5] This foregrounds the need to address online harms within the African region and against important themes like colonialism, human rights and governance.

[3] Galtung defines violence as a 'deterioration of fundamental human needs which can be avoided, or more general, a life impairment which decreases the degree where people are able to fulfil their needs at a certain level or potential possible.' He also identified threatening as violence. In the context of this chapter, this definition moves beyond the traditional understanding of violence as limited to the physical space to include internet. See K Ho 'Structural violence as a human rights violation' (2007) 4 *Essex Human Rights Review* 1-15, citing J Galtung 'Kulturelle Gewalt' (1993) 43 *Der Burger im Staat* 106. Ho also defined structural violence as an infringement of human rights. See section 2.4.3 above for more detailed explanations of how online harms could be violent.

[4] Section 3.5 below.

[5] H Wasserman 'Fake news from Africa: Panics, politics and paradigms' 21 *Journalism* (2020) 3-5.

© University of Pretoria

Within the African context, while some online harms are more prominent than others, the less prominent ones are also beginning to require more debates especially with respect to the need for regulatory approaches. At various times, each of the online harms have been experienced in one or more African context. Some of these contexts will be analysed in turn. Given this background, this chapter seeks to answer the second sub-question of this thesis on the various forms of online harms and their impacts on the right to freedom of expression in Africa. In doing so, it breaks the question down into three sub-questions:

a. What are online harms?
b. What are their forms, methods and classifications?
c. How do these harms impact the right to freedom of expression in Africa?

In attempting these questions, this chapter is divided into seven sections. This first section introduces the chapter. The second section discusses the concept of online harms. The third section considers the various forms of online harms and how they manifest in African contexts. The fourth section focuses on how online harms are caused by actors and their methods. The fifth section examines the harm versus illegality debate on online harms while the sixth section examines the impacts of these harms on the right to freedom of expression. The seventh section concludes that these impacts may be prevented and the right to freedom of expression online may be protected if a rights-respecting approach to platform governance is considered.

3.2 The concept of online harms

Online harms can first be broken down into two components: 'online' and 'harms.' A simple dictionary meaning of 'online' means 'connected to, served by, or available through a system and especially a computer or telecommunications system (such as the Internet)' or 'also done while connected to such a system.'[6] This definition suggests that online activities require a device and the Internet to function. Taking this definition further, online activities are often carried out in a notional environment referred to as the 'cyberspace' which is often used synonymously with the Internet. The Internet has been simply defined as 'an electronic communications network that connects computer networks and organizational computer facilities around the world – used with *the*, except when being used attributively.'[7] Therefore, the term online can be used interchangeably with cyberspace and the Internet. Harm can take several forms including physical, psychological and emotional harms.[8] According to Merriam-

[6] Merriam Webster 'Online' https://www.merriam-webster.com/dictionary/online (accessed 24 July 2020).
[7] Merriam Webster 'Internet' https://www.merriam-webster.com/dictionary/Internet (accessed 24 July 2020).
[8] E Harman 'Harming as causing harm' in MA Roberts & DT Wasserman (eds) *Harming future persons* (2009) 139.

© University of Pretoria

Webster dictionary, it is 'physical or mental damage.'[9] It was not until recently that these forms became associated in effect with other non-physical environment – cyberspace.[10]

As noted above, the history of online harm is connected to the Internet ecosystem and the widespread adoption of information and communication technologies.[11] However, while online harms are often a transplantation of offline behaviours into the digital space, the specific dangers they pose did not become clear until the last five years.[12] The recent awareness of online harms became more protracted due to the impacts they have on the right to freedom of expression specifically and human well-being in general. This awareness can be seen in the 2016 Cambridge Analytica exposé which saw some of Africa's democracies witness a large-scale and unprecedented use of digital manipulation.[13] While the 2016 development caused a huge stir, the company at its helm, Cambridge Analytica is not new to Africa or its politics.[14]

Fast forward to January 2021, this manipulation grew bolder and nearly overtook the US Capitol.[15] Given this example and more across the world, online harms have not

---

[9] Merriam Webster 'Harm' https://www.merriam-webster.com/dictionary/harm (accessed 24 July 2020).
[10] I Agrafiotis *et al* 'A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate' (2018) 4 *Journal of Cybersecurity* 3.
[11] J Naughton 'The evolution of the Internet: From military experiment to general purpose technology' (2016) 1 *Journal of Cyber Policy* 5. The author describes the internet as one of the poorly understood technologies which has given rise to 'a range of new, and potentially dangerous vulnerabilities…' including the 'challenges of devising regulatory institutions which are fit for purpose in the digital age.'
[12] B Chesney & DK Citron 'Deep fakes: a looming challenge for privacy, democracy, and national security' (2019) 10 *California Law Review* 1771-1784. https://bit.ly/3ces2EX (accessed 12 October 2020). TaylorWessing 'Online harms: The regulation of internet content' October 2019 https://www.taylorwessing.com/download/article-online-harms.html (accessed 15 August 2020). The introduction of section 230 of the Communications Decency Act of 1996 provided immunity for website publishers from third-party content which also animated the debate on whether social media companies have responsibilities to protect users from harm given the protection.
[13] B Ekdale & M Tully 'African elections as a testing ground: Comparing Coverage of Cambridge Analytica in Nigerian and Kenyan Newspapers (2019) 40 *African Journalism Studies* 13; See also B Ekdale & M Tully 'Cambridge Analytica in Africa – what do we know?' 10 January 2020 *Democracy in Africa* http://democracyinafrica.org/cambridge-analytica-africa-know/ (accessed 16 August 2020). Here, both authors give one of the most suitable description of digital colonialism that causes online harms when they stated that 'it's important that African countries update their data privacy and protection laws. But… the Cambridge Analytica scandal runs deeper than access to Facebook data'; H Berghel 'Malice domestic: The Cambridge Analytica dystopia' (2018) *IEEE Computer Society* 85 https://bit.ly/3iyIcKf (accessed 16 August 2020).
[14] S Solomon 'Cambridge Analytica played roles in multiple African elections' 22 March 2018 *Voice of America* https://www.voanews.com/africa/cambridge-analytica-played-roles-multiple-african-elections (accessed 18 August 2020).
[15] R Heilweil & S Ghaffary 'How Trump's internet built and broadcast the Capitol insurrection' 8 January 2021 *VoxMedia* https://www.vox.com/recode/22221285/trump-online-capitol-riot-far-right-parler-twitter-facebook (accessed 10 January 2021).

© University of Pretoria

only gone beyond the borders of 'strong democracies', they have become a social virus that infects systems whether weak or strong. In the scandal that trailed the complicity of private social networking companies like Facebook and Twitter, a number of African countries have also been impacted by various forms of online harms. These impacts are seen in the distortion of the right to political participation online, narrowing space for political and unpopular speech, exclusion of women's right to expression, conflation of forms of online harms that lead to problematic legal and regulatory frameworks, less protection for children's rights online and the precipitation of offline violence through online hate speech.[16]

Therefore, the negative impacts of digital technologies through online harms requires more studies and effective governance approaches especially in Africa. It is also important to understand what these harms mean under the law, their impacts on human rights in general and on the right to freedom of expression specifically.[17] Due to the fact that they have just begun to garner global attention, the literature on what online harms are as a whole, is sparse and developing even though there has been research on how some of their forms manifest which has a common denominator – to distort, malign and damage.[18] Therefore, this chapter attempts a bold but important step of defining the term online harms by considering its constituent parts to arrive at a coherent and workable definition – it takes a stipulative definition approach that considers the constituent parts to regenerate meaning for a whole to arrive at a definition.[19]

Belli and Zingales attempted a definition of online harm but with more emphasis on harm as a term while also showing that harm could be as physical as it could be 'non-tangible' like:

Sidner & M Simon 'Heading 'into a buzzsaw': Why extremism experts fear the Capitol attack is just the beginning' 18 January 2021 *CNN* https://edition.cnn.com/2021/01/16/us/capitol-riots-extremism-threat-soh/index.html (accessed 19 January 2021).

[16] Section 3.6 below.

[17] Section 3.6 below.

[18] L Price 'Platform responsibility for online harms: Towards a duty of care for online hazards' (2022) 13 *Journal of Media Law* 238-261; K Wodajo 'Mapping (in)visibility and structural injustice in the digital space' (2022) 9 *Journal of Responsible Technology* 1-8; EQ Okolie 'Extent of the latitudes and limits of social media, and freedom of expression within the confines of the law in Nigeria' (2019) 83 *Journal of Law, Policy and Globalization* 162-167; D Sive & A Price 'Regulating expressions on social media' (2019) 136 *South African Law Journal* 51-83; MA Tadeg 'Freedom of expression and the media landscape in Ethiopia: Contemporary challenges' (2020) 5 *University of Baltimore Journal of Media Law and Ethics* 69-99.

[19] J Pavelka *Descriptions, prescriptions, and the limits of knowledge* (2020) 12 http://www.knowledgedefinition.com/CH6TheoryOfDefinition0416.pdf (accessed 15 September 2020). It applies the precise formalised definition using both pragmatic and technical formalisations.

© University of Pretoria

> causing a person to fear for their physical, emotional or psychological safety, experience anxiousness, limit their speech, feel intimidated in their personal or professional life or worry for their personal or professional reputation.[20]

The closest example to an application of Belli and Zingales' observation is the United Kingdom Government's White Paper which described harmful content and activities.[21] The white paper categorises 'harmful content or activities' along the clarity of definitions. For example, harmful content like child sexual exploitation and abuse, terrorist content and even online hate speech are 'clear.' Whereas, definitions for harms like cyberbullying and disinformation are 'less clear.' The last category is 'underage exposure to legal content' which is 'children accessing pornography and accessing inappropriate material.'[22] Therefore, while definitions of what is an online harm may be clear in some instances and unclear in some, their impacts are clear enough because they are not only patently harmful, but also potentially damaging. However, for the purpose of this chapter, it will be necessary to have a working definition to guide its course.

The complexity of the dynamics of online harms which are not limited to speech-related violence has made it increasingly difficult to pin down the term to a definition. Considering the definitions above, online means 'connected to…available through a system (such as the Internet)', meaning online equals the Internet. Adding the definition of harm to this, two features are common to online harms – electronic communication (Internet) and negative impact (harm). Therefore, online harms may be defined as the use of electronic communication to negatively impact on the physical and mental well-being of others. They can also be defined as technology-mediated harm or violence. In order for it to be referred to as an online harm, it must be carried out through electronic communication and be negative in impact.

In order for online harm to be communicated, the message must be received by the interpreter or recipient. Such receipt may be in any multimedia format. It does not matter whether the interpreter or recipient is online. What matters is that such harm is communicated online and received. In terms of impact, such online harm must cause either action or inaction – it suffices as an impact of online harm if it occasions either.

On negative impacts, they may include but are not limited to the various examples given by Belli and Zingales above. For example, the claims of voter's suppression in the United States have been connected to the impact of online harms in causing

---

[20] L Belli & N Zingales *Glossary of platform law and policy terms* https://cyberbrics.info/wp-content/uploads/2020/11/Glossary-on-Platform-Law-and-Policy-CONSOLIDATED17472-1.pdf 106 (accessed 10 November 2020).

[21] HM Government 'Online harms white paper', April 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf (accessed 24 July 2020).

[22] As above.

© University of Pretoria

inaction. The communication here are the various hate messages, propaganda, disinformation campaigns etc. which are communicated electronically. The negative impact is best construed as the inaction of those who chose not to vote based on these harms or those who did but admitted to have been influenced by such communication. An exacting example of these factors is the role of WhatsApp communications in elections in a number of African countries. According to Cheeseman *et al,* 'WhatsApp's disruptive capability, highly valued by opposition parties and civil society groups, also facilitates the spread of 'fake news.'[23]

3.3 Forms of online harms

Considering the context in which the term online harms is used to examine its effects, especially with respect to freedom of expression, there are specific examples of what constitute online harms. For example, for it to harm *per se*, an online activity must have been electronically communicated with an adverse and undesirable effect either on another user or a community as a whole or both. Such an effect is not limited to spaces, that is online or offline environments. Therefore, that an act is carried out online does not mean it would not have an offline effect and conversely, that an act is offline does not mean that it may not constitute an online harm. This blurred line between online and offline effects of harms is necessary to demonstrate that regardless of the environment, harms can both be online and physical. This is to say that given the ubiquity of digitalisation, the barriers between physical and cyberspace are reduced as both spaces can have the same impacts on each other. The most important feature of online harms, just like any other form of violence is to cause 'physical or mental damage.' These features are demonstrated in all the forms of online harms to be discussed subsequently under this section. This chapter considers two major forms of online harms – information disorder and targeted online violence.

3.3.1 Information disorder

Some of the major attributes of information disorder is to harm or distort the true state of facts by misrepresenting them.[24] It may be termed as the use of communications that are targeted to harm or distort the true representation of an occurrence.

---

[23] N Cheeseman *et al* 'Social media disruption: Nigeria's WhatsApp politics' (2020) 31 *Journal of Democracy* 156.

[24] C Wardle & H Derakhshan 'Information disorder: Toward an interdisciplinary framework for research and policy making' 2017 *Council of Europe Report* 5 https://rm.coe.int/information-disorder-report-2017/1680766412 (accessed 20 September 2020); C Wardle 'Understanding information disorder' October 2019 *First Draft* https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76701 (accessed 15 September 2020). This resource categorised online harms along low and high harm, information disorder carries various levels of impact. For example, misleading content, false connection and satire or parody have been categorised as low harm information disorder while fabricated content, manipulated content, imposter content and false context have all been categorised as high harm content; A Deem *et al* 'Hate speech, information disorder and conflict' *Social Science Research Council* 4-6.

© University of Pretoria

Information disorder can be divided into three broad categories: misinformation, disinformation and malinformation. These forms of information disorder are not new to the information ecosystem, rather, they have been amplified by the Internet. Several definitions have been offered to conceptualise each of these forms of information disorder. Definitions for each of these forms are considered in turn.

A Misinformation

In defining misinformation, many scholars differ in their approach to the concept. Majorly, there are three approaches used by various scholars to define misinformation. They are the broad, integrated and temporal approaches. The broad approach defines the concept as stand-alone but open. For example, Fetzer defines misinformation as 'false, mistaken, or misleading information.'[25] To Berinsky, it is the 'information that is factually unsubstantiated.'[26] These examples are broad in the sense that it is misinformation if it is false or factually unsubstantiated – there is no grey area.

Using the integrated approach, some scholars choose to see it in relation to other forms of information disorder – disinformation and other terms like misperceptions and conspiracy theories. Misperception has been defined by Nyhan and Reifler as 'cases in which peoples' beliefs about factual matters are not supported by clear evidence and expert opinion.'[27] On the other hand, according to Weeks and Garrett, conspiracy theories are 'unverified stories or information statements people share with one another.'[28] In this regard, to Wardle, it is 'information that is false, but not intended to cause harm'[29] which in sharp contrast to disinformation is 'false information that is deliberately created or disseminated with the express purpose to cause harm'[30] and is set apart by 'intentionality.'

Using the temporal approach, some scholars have chosen to define misinformation in relation to time and as a process – it is first presented as true but subsequently corrected. According to Lewandowsky *et al*, 'any piece of information that is initially

---

[25] JH Fetzer 'Disinformation: the use of false information' 14 *Minds and Machines* 231.

[26] AJ Berinsky 'Rumours and health care reform: Experiments in political misinformation' 47 *British Journal of Political Science* 241.

[27] B Nyhan & J Reifler 'When corrections fail: the persistence of political misperceptions' 32 *Political Behaviour* 305.

[28] BE Weeks & RK Garrett 'Electoral consequences of political rumours: Motivated reasoning, candidate rumours and vote choice during the 2008 US presidential election' 26 *International Journal of Public Opinion Research* 402.

[29] Wardle and Derakhshan (n 24 above).

[30] C Wardle 'Information disorder: The essential glossary' (2018) *Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy* https://firstdraftnews.org/wp-content/uploads/2018/07/infoDisorder_glossary.pdf (accessed 20 April 2020).

© University of Pretoria

processed as valid but that is subsequently retracted or corrected.'[31] To Ecker, misinformation is 'information that is initially presented as factual but subsequently corrected.'[32] In foregrounding this approach, Wittenberg and Berinsky stated that 'in this sense, information only becomes misinformation when it is first believed and later corrected, separating misinformation from other false information that goes unrebutted.'[33]

Considering these approaches, at least four cross-cutting features may be associated with misinformation. First, the approaches, together with the definitions, focus on the truth value of the information – they are concerned with whether the information has been proven or not. Second, scholars define misinformation along the formats of their presentation, especially whether or not they are designed to resemble traditional news sources. Third, in defining misinformation, the focus is intentionally about the supply of false information and not necessarily beliefs that are true or not. Lastly, the intentions of what misinformation seeks to achieve also comes into sharp focus in all of the three approaches which will be further expatiate on in the subsequent sections.[34]

A closer look at these approaches and their common features demonstrates that misinformation, though not often harmful and immediately dangerous, can be used to drive direr forms of information disorder like disinformation and malinformation.[35] However, despite this potential, it is also the form of information disorder with the most feasible chances of being addressed even though scholars have argued that it might be tough to do so.[36]

For example, some of the issues that have been raised on whether quelling misinformation is effective are continued influence and backfire effects. According to Johnson and Seifert, even after the correction of misinformation, it continues to influence peoples' attitudes and beliefs.[37] This is referred to as the continued influence effect. In buttressing this point, Wittenberg and Berinsky argued that:

---

[31] S Lewandowsky *et al* 'Misinformation and its correction: continued influence and successful debiasing' (2012) 13 *Psychological Science in the Public Interest* 124-125.
[32] UKH Ecker *et al* 'He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation' (2015) 85 *Journal of Memory and Language* 102.
[33] C Wittenberg & J Berinsky 'Misinformation and its correction' in N Persily & JA Tucker (eds) *Social media and democracy: The state of the field and prospects for reform* (2020) 166.
[34] Section 3.5 below.
[35] Wardle (n 24 above).
[36] See HM Johnson & CM Seifert 'Sources of the continued influence effect: When misinformation in memory affects later inferences' (1994) 20 *Journal of Experimental Psychology: Learning, Memory and Cognition* 1420-1436.
[37] As above.

© University of Pretoria

> Importantly, people may correctly recall a retraction yet still use outdated misinformation when reasoning about an event. From this perspective, corrections can partially reduce misperceptions but cannot fully eliminate reliance on misinformation in later judgments.[38]

There have been two potential reasons for misinformation to linger even after correction. These two reasons have been described as the dual-process theory[39] and mental model theory.[40] For the dual-process theory, it differentiates between two types of memory retrievals: automatic and strategic. Both forms of retrievals have also been associated with online processing of misinformation, where the affective connection to a piece of misinformation is stronger than any other form of effort to rebut it.[41] As the names imply, automatic retrievals are fast and less thoughtful while the strategic form of retrieval is more deliberate and planned. While automatic processing is also often not context-based, strategic retrievals due to the nature of them being deliberate retrieve specific details about a piece of information. As a result of this, misinformation is actively recalled but its correction is not.

According to the mental model theory, individuals are able to effectively construct an alternative universe in their mind such that the corrective impact of misinformation cannot be realised.[42] For this model, 'corrections are more effective when they contain alternative causal details rather than just simplified corrections.'[43] Another major attribute of this model is that even if individuals can recollect corrections, they are able to invoke a strong attachment to misinformation until a more plausible alternative correction takes place.[44]

A look at both models suggests that bias is a strong motivation for spreading misinformation even though such bias need not be actively worked up. This is because in most instances of misinformation, they spread faster because they are often well-tailored responses to behavioural tendencies and as such grow faster not because they are false but because they cater to these tendencies. Another point to be considered from both models is that they show misinformation to require more than simplified responses. In correcting misinformation, it must be plausible and deliberate. However, one of the backfire effects of correcting misinformation is that individuals

[38] Wittenberg & Berinsky (n 33 above).
[39] See UKH Ecker *et al* 'Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction' (2011) 18 *Psychonomic Bulletin & Review* 570.
[40] Johnson & Seifert (n 35 above); See B Swire & UKH Ecker 'Misinformation and its correction: Cognitive mechanisms and recommendations for mass communication' in B Southwell, EA Thorson & L Sheble (eds) *Misinformation and mass audiences* (2018) 195, 211.
[41] See M Lodge & CS Taber *The rationalising voter* (2013); E Thorson 'Identifying and correcting policy misperceptions' (2016) https://www.americanpressinstitute.org/wp-content/uploads/2015/04/Project-2-Thorson-2015-Identifying-Political-Misperceptions-UPDATED-4-24.pdf (accessed 27 September 2020).
[42] Swire & Ecker (n 40 above).
[43] Johnson & Seifert (n 36 above).
[44] Wittenberg & Berinsky (n 33 above).

© University of Pretoria

tend to hold on to patently false piece information because it fundamentally challenges their worldview.[45]

B Disinformation

According to Benkler *et al*, disinformation is 'manipulating and misleading people intentionally to achieve political ends.'[46] According to Tucker *et al*, it could be 'a broad category of information that one could encounter that could possibly lead to misperceptions about the actual state of the world.'[47] In the same work on political disinformation, the scholars referred to disinformation as 'deliberately propagated false information.'[48] Kragh and Åsberg see disinformation as 'intentionally false or inaccurate information that is spread deliberately.'[49] To Bontcheva and Posetti, disinformation is described as 'false or misleading content with potentially harmful consequences, irrespective of the underlying intentions or behaviours in producing and circulating such messages.'[50] The UK government defines disinformation as 'the deliberate creation and sharing of false and/or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain.'[51]

Ziegler and Maréchal, with a more historical and international relations perspective, see disinformation as the Soviet-era practice of *dezinformatsiya*, which is Russian connotation for 'planting false or distorted stories to influence Western public opinion.'[52] The importance of this history is that while disinformation is as old as humanity, its use beyond borders to influence local politics in another country however, became pronounced with the Cold War.[53] Rid seems to have taken on this importance as he analysed the recent history of disinformation from the Cold War which saw

[45] Wittenberg & Berinsky (n 33 above) 169.
[46] See Y Benkler et al *Network propaganda: Manipulation, disinformation, and radicalization in American politics* (2018) 24.
[47] J Tucker *et al* 'Social media, political polarisation, political disinformation: A review of scientific literature' (2018) *SSRN* 3.
[48] As above.
[49] M Kragh & S Åsberg 'Russia's strategy for influence through public diplomacy and active measures: The Swedish case' (2017) 40 *Journal of Strategic Studies* 25.
[50] K Bontcheva & J Posetti 'Balancing act: Countering digital disinformation while respecting freedom of expression' (2020) *International Telecommunications Union* 8.
[51] Digital, Culture, Media and Sport Committee 'Disinformation and 'fake news': Interim report: Government's Response to the Committee's Fifth Report of Session 2017-2019' https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1630/1630.pdf (accessed 20 September 2020).
[52] See C Ziegler 'International dimensions of electoral processes: Russia, the USA and the 2016 elections' (2016) *International Politics* 2; N Maréchal 'Networked authoritarianism and the geopolitics of information: Understanding Russian Internet policy' (2017) 5 *Media and Communication* 29-41.
[53] Citron (n 2 above) 7.

© University of Pretoria

information warfare as one of the most potent tools of international relations.[54] He created a dichotomy between democracies and authoritarian systems and the West and the East while also making copious reference to the intra-Communists struggle in the East.[55] For Bradshaw and Howard, disinformation is defined with a more applied approach towards online activities and 'computational propaganda as the use of algorithms, automation, and human curation to purposely distribute misleading information over social media networks.'[56]

Looking at the thought lines of most scholars in the field of disinformation studies, they can be categorised into using two main approaches in considering the concept – the open and closed or applied approaches. Scholars like Shu, Tucker and Barbera and Bontcheva and Posetti view disinformation more holistically without necessarily being defined solely by an activity – the term may be used in any instance where the information deliberately misleads or deceives. To Bontcheva and Posetti, disinformation is better used as a 'meta-label' to cover false content that may potentially cause societal harm. They argued further:

> It is this [meta-label] that enables a wide-ranging unpacking of responses to disinformation underway worldwide. The intent, therefore, is not to produce yet another definition of what disinformation is, but to provide for a broad umbrella conceptualisation of the field under examination and analysis.[57]

The approaches by these scholars and organisations may be referred to as the open approach wherein the conceptualisation of disinformation is not limited towards a specific scope but is underlined by its conceptual core features, deceit and potential harm.

On the other hand, Ziegler and Maréchal adopt the closed or applied approach as they see disinformation more as a political activity while Rid sees disinformation as a combination of various activities including political activities but more related to international politics and relations and history.[58] Therefore, unlike the open approach, the close or applied approach pins disinformation down to a set of motives or activities. Another example of closed or applied approach is the one used by Bradshaw and Howard which streamlines disinformation to a digital activity.[59] This seems to be the norm more recently given the constantly blurred lines between traditional and digital sources of information – what and where information is shared or received may now

[54] T Rid *Active measures: The secret history of disinformation and political warfare* (2020) 9.
[55] Rid (n 54 above) 28.
[56] S Bradshaw & PN Howard 'Challenging truth and trust: A global inventory of organized social media manipulation' (2018) *Oxford Internet Institute*, *University of Oxford* 3-21.
[57] SC Woolley & PN Howard 'Political communication, computational propaganda, and autonomous agents' (2016) 10 *International Journal of Communication* 4882 4890.
[58] Ziegler and Maréchal (n 52 above).
[59] Bradshaw & Howard (n 56 above).

© University of Pretoria

be termed as 'trado-digital' media theory.

Most of the traditional media like newspapers, broadcasting stations (both radio and television) and magazines have now transitioned into digital systems in order to reach more audiences, circumvent state censorship and cater for market needs.[60] When considered as a process, it could be termed the digital migration of news. While print and broadcast media outlets still maintain their traditional modes of information dissemination, they have also transitioned to digital media sources of information like social media platforms and websites. One reason for this could be to scale markets and avoid state overregulation.

The trado-digital theory may be traced to the libertarian media theory – it suggests that the libertarian media theory considers media as the 'self-correcting' feature of Mill's idea of a free society.[61] According to Siebert, the theory advocates for societies that provide media systems with unrestrained freedom to determine the public's right to know.[62] The relationship between the trado-digital media theory and the libertarian media theory can be gleaned from Ward's position on the libertarian media theory as 'a maximally unfettered press helping to create a liberal society over and against the forces of tradition and conservatism.'[63] Therefore, the trado-digital media theory may be defined as the interwoven and linear nature of media communications in the 21st century where both traditional and digital media are being merged in terms of how they disseminate information.

In proving the trado-digital theory, Ugangu highlighted market forces and state censorship as some of the major reasons for the exodus of many media systems into digital spaces.[64] However, he highlighted an important point that Internet access is not necessarily widespread in Kenya where he based his studies on, and therefore, such transition may not have maximum impact as traditional media. Nonetheless, Internet access is growing in Kenya and in other African countries even though it is mostly slower in the latter. Ward and Ugangu show that the movement to online spaces by traditional media is primarily to achieve unrestricted media that is free of 'tradition and

[60] T Chari 'Future prospects of the print newspaper in Zimbabwe' (2011) 3 *Journal of African Media Studies* 367-388; NZ Nkomo *et al* 'The viability of the print newspaper in the digital era in Zimbabwe: A digital strategy perspective' (2017) 5 *European Journal of Business and Innovation Research* 44.
[61] See F Siebert 'The libertarian theory of the press' in F Siebert, T Peterson & W Schramm (eds) Four theories of the press: The authoritarian, libertarian, social responsibility and soviet communist concept of what the press should be and do (1963) 43-62.
[62] As above.
[63] SJA Ward 'Classical liberal theory in a digital world' in RS Fortner & PM Fackler (eds) *The handbook on media and mass communication theory* (2014) 7.
[64] W Ugangu 'Normative media theory and the rethinking of the role of the Kenyan media in a changing social economic context' unpublished PhD thesis, University of South Africa, 2012 164 http://uir.unisa.ac.za/bitstream/handle/10500/8606/thesis_ugangu_w.pdf;sequence=1 (accessed 15 June 2019).

© University of Pretoria

conservatism.'[65]

Therefore, the trado-digital theory is more aligned towards the open approach in that it allows disinformation to be studied from multidimensional perspectives but in narrow structures that make up both traditional and digital media systems. Most legislation or policies seem to adopt the open approach rather than the closed approach in order to be able to accommodate the dynamics of false information while most practitioners tend to focus more on the closed or applied approach in order to systematically study disinformation.

However, according to Pielemeier, such approach by states in regulating disinformation presents three major challenges especially when compared to other online harms which are conceptualisation of disinformation, proving intent and the harm impacts of disinformation.[66] First, he argued that the conceptualisation of disinformation as the umbrella word for information disorder makes it difficult to regulate information disorder as an online harm as its several forms have varying impacts.[67] Second, there is also the challenge of proving intent especially on the part of the speaker. He argues, along with others, that determining the intent of the speaker is particularly difficult because 'a speaker's intent is difficult to determine and how the chilling effect may not be justification for the speaker's intent requirements.'[68] Third, determining the harm value of disinformation is notoriously difficult because regulating false statements might increase the chilling effect for free speech. In addition to this, disinformation combined with other forms of online harms are manifestly intended to cause harm through inauthentic behaviours that impact elections and affect democracies which raises the harm value for disinformation.[69] The inability to specifically measure the certitude of harm in disinformation makes it more difficult to regulate and what measure can be used to limit it that would not harm free speech.

In addition to Pielemeier's distinctions on the regulatory challenges posed by disinformation, misconceptualisation of disinformation is not limited to its immediate family of information disorder but also other forms of online harms like hate speech online and terrorist content. This is often so because while disinformation may be the means through which other forms may be carried out, they are distinct in concept, harm value and impacts. Therefore, conflation of disinformation with other forms of online harms is also problematic for regulation.

---

[65] Ward (n 63 above).
[66] J Pielemeier 'Disentangling disinformation: What makes regulating disinformation so difficult' (2020) 4 *Utah Law Review* 921.
[67] Pielemeier (n 66 above) 922.
[68] As above.
[69] Pielemeier (n 66 above) 923.

© University of Pretoria

C Malinformation or propaganda

Propaganda, in the most neutral sense, means to disseminate or promote particular ideas – in Latin, it means 'to propagate' or 'to sow.'[70] Considering information disorder as a gradient, malinformation or propaganda lies at the extreme end of the slope. What places malinformation at the far end of the slope is the degree of intent involved. Therefore, malinformation, also known as propaganda, is an organised, orchestrated campaign of distorted facts.[71] Nelson offers a more purposive definition of the term propaganda as:

> a systematic form of purposeful persuasion that attempts to influence emotions, opinions, and actors of specified target audiences for ideological, political or commercial purposes through the controlled transmission of one-sided messages (which may or may not be factual) via mass or direct media channels.[72]

An important distinction to draw from this definition from other forms of information disorder is that unlike misinformation and disinformation, malinformation need not be false.[73] To Ellul *et al*, 'truth does not separate propaganda from 'moral forms' because propaganda uses truth, half-truth, and limited truth.'[74] What seems to matter most are two major factors – purposeful persuasion which could be understood as intent and ability to convince. In supporting this, Tucker *et al* defines propaganda as 'information that can be true but is used to disparage opposing viewpoints.'[75] These two factors, combined with the railroading of information, set malinformation aside in the information disorder category as a more calculated effort towards harm.

In another definition, propaganda is defined as:

> Information, historically promulgated by state officials but today often also by political opponents, that may or may not be true, but which presents the opposing point of view in an unfavourable light in order to rally public support.[76]

This definition tailors malinformation specifically towards a political activity. This further brings its impact, along with other information disorders close to home on democratic development. Considering the various theories of free speech in the previous chapter, which includes the truth and democratic theories, the intent to harm by presenting a

---

[70] JS Jowett & V O'Donnell 'What is propaganda and how is it different from persuasion?' *Propaganda and persuasion* (2005) 2 http://www.ffri.hr/~ibrdar/komunikacija/seminari/Propaganda%20&%20persuasion%20-%20difference%20(Chapter1).pdf (accessed 12 October 2020).
[71] Woolley & Howard (n 56 above).
[72] RA Nelson *A chronology and glossary of propaganda in the United States* (1996) 336.
[73] Jowett & O'Donnell (n 69 above) 4.
[74] J Ellul et al *Propaganda: The formation of men's attitudes* (1965) xv.
[75] Tucker *et al* (n 47 above).
[76] Nelson (n 72 above).

© University of Pretoria

one-sided narrative which not only hurts access to information that assists in making informed democratic decisions but manipulates opinions, ties malinformation as a form of online harm to stunted democratic development. In the online space, malinformation is often referred to as computational propaganda.[77]

In terms of relationship with other forms of information disorder, malinformation is often associated with disinformation because the latter is often used to further spread and reinforce the former.[78] A combination of both harms is often targeted at finding a fine balance between spreading false information and manipulating and misleading content. Both forms also carry a fair amount of intent towards sharing contentious information. However, while disinformation is often required to be false to be properly so called, malinformation need not be false which makes it more difficult to identify and prevent. All that are required for malinformation is that there is a graduated intent to limit the scope of information with respect to a certain issue in order to fence off contrary information.

D The differences, similarities and features of information disorder

One of the major reasons for a conceptual clarification of information disorder is to ensure that its forms are not conflated so as to ensure the necessary policy and regulatory response. For example, while misinformation is largely permissible due to their lack of intent and low propensity to occasion harm, it is often used interchangeably with disinformation which has since been demonstrated to be more intentional and has a higher potential for harm.[79] Also, in comparison, while both misinformation and disinformation are largely defined by intent and harm, malinformation actually utilises the two.[80] Therefore, various examples of each form of information disorder and how they manifest and in turn impact human rights and democratic development will benefit how policies are formulated in order not to conflate the issues and end up with the wrong results.

Despite their differences, forms of information disorder also go through the same phases. These phases are creation, production and distribution.[81] The creation of information disorder, whether misinformation, disinformation or malinformation is the inception or conceptualisation phase where the idea to spread an item is first conceived. The production phase of information disorder often entails the tailoring of

[77] Woolley & Howard (n 57 above).
[78] D Jackson 'Distinguishing disinformation from propaganda, misinformation and "fake news"' (2018) *National Endowment for Democracy* 1 2 https://www.ned.org/wp-content/uploads/2018/06/Distinguishing-Disinformation-from-Propaganda.pdf (accessed 12 October 2020).
[79] Wardle (n 24 above).
[80] As above.
[81] As above.

© University of Pretoria

the idea or content in a communicative format that can be easily shared and spread. After this phase, comes the distribution phase where the targeted or intended audience of the platform begins to consume such information.

A point to note on the importance of identifying these phases is that different actors may be responsible for each of these phases. For example, rather than be an actor, a politician or political campaign may outsource the production of information disorder to mercenaries.[82] In addition, another actor may pick up the distribution of the content, for example trolls.[83] These phases may help in identifying the intent behind each information disorder in order to be able to identify the kind of disorder it is. For example, a deliberate conception of an idea from the creation phase to the distribution phase may easily connote intent. However, this does not mean that information disorder like misinformation, which is without intent cannot become harmful. An important point to note, while using the phases to determine intent is to consider the nature, efforts and actors involved at each stage of the phases. A political party that conceives an idea that is either false or true, produces it in media format and targets certain forms of messaging and audience will raise a red flag compared to a random user who does the same.[84]

Also, in the more granular structures of information disorder, there are three major features that come to play. Wardle described them as agent, message and interpreter.[85] The agent is the actor or actors involved in the phases of information disorder – they are often with or without motive or intent to spread false information or one-sided narrative. There are seven identified elements of agents in the agent element in order to determine the kind of information disorder at play. They are type of actors, organisation, motivation, targeted audience, automated technologies, intent to mislead and intent to harm.

The message is content of production that is distributed through information disorder. Message in information disorder may be categorised into two: patently false information or unwholesome information. The former is often associated with misinformation or disinformation while the latter is usually through malinformation. It is the content of information that is shared in digestible formats like online posts, newspapers, multimedia etc. There are currently five approaches in considering messaging in information disorder. They are: durability of the message; accuracy; legality; credibility and intended audience.

---

[82] E Woollacott 'Russian trolls outsource disinformation campaigns to Africa' 13 March 2020 *Forbes* https://www.forbes.com/sites/emmawoollacott/2020/03/13/russian-trolls-outsource-disinformation-campaigns-to-to-africa/?sh=7ec387d1a263 (accessed 13 October 2021).

[83] As above.

[84] Africa Centre for Strategic Studies 'Domestic disinformation on the rise in Africa' 6 October 2021 https://africacenter.org/spotlight/domestic-disinformation-on-the-rise-in-africa/ (accessed 13 October 2021).

[85] Wardle (n 24 above).

© University of Pretoria

The interpreter is often the consumer of information disorder in that they receive, interpret and either act or do not act on it. In considering the interpreter, at least three approaches should be focused on. They are: type of audience, that is oppositional, hegemonic or negotiated; action taken, that is ignored, shared in support or in opposition; and meta-analysis, that is the propensity of interpreters to seek facts beyond the message.

Therefore, while all forms of information disorder thrive on falsity and one-sided railroading of information, their propensity to engender harm depends on the form. For misinformation, while it does not intend to spread false information, it could be hijacked, misconstrued and distorted to carry intent to spread false information.[86] For satiric messages by agents that are intended to suggest an opposite meaning in order to fully carry out the meaning within a certain context could be easily termed misinformation. But it could graduate into construed intent if hijacked to include deception and one-sided narratives. Also, disinformation is patently harmful even though it may not be illegal. This is because a deliberate misrepresentation of facts is targeted to mislead which may be received and acted on by the interpreter as true. This could potentially lead to wrong perceptions and encourage activities that do not reflect the true state of facts which could lead to online deception causing offline harm.

With respect to malinformation, the intent to deceive graduates a notch higher to include deliberate misrepresentation of facts. The inability to wholly consider available facts in a scenario does not adequately equip the interpreter with the set of information required to fully form an opinion. While the harmful nature or legality of each of these forms will be discussed later in this chapter, it is worth noting at this point that while all forms of information disorder do have the potential to engender harm in varying degrees, in the face of the law, they are not necessarily illegal.

Organised campaigns that use information disorder fully began across the globe with 28 countries and has since grown to include African countries.[87] The term ‘fake news’ became more prominent and the value of truth or the idea of pursuing what it means took backstage as political and unpopular speech became the synonym for it. Not only did information sharing become an existential threat to exercising the right to freedom of expression online, critics became more incensed about the power of social media platforms, and rightly so.[88] The history of data analysis companies like Cambridge Analytica in Africa dates back to 1994 when it was hired by a political party in South

[86] Wardle (n 24 above).
[87] S Bradshaw & PN Howard ‘Challenging truth and trust: A global inventory of organized social media manipulation’ (2019) *Oxford Internet Institute*, *University of Oxford* 1-23.
[88] As above.

© University of Pretoria

Africa to 'mitigate election violence.'[89] In 2018, the Cambridge Analytica debacle exposed once more the various levels of information disorder that are especially present in Africa. Data analysis companies provide services to governments and political parties to influence undecided voters through appeals to their emotions.[90] Beyond the United States of America and the United Kingdom that were affected by Cambridge Analytica's activities, so far in Africa, the company has managed the electoral campaign of Kenya's current President and was also involved in Nigeria's defining 2015 elections. On what it did in Kenya and in ensuring the President and his political party's victory, the company claimed to have:

> rebranded the entire party twice, written the manifesto, done huge amounts of research, analysis, messaging. Then we'd write all the speeches and stage the whole thing. So, just about every element of his campaign.[91]

Some of the political campaigns crafted by Cambridge Analytica in Kenya included stoking fears about Al-Shabab and disease breakouts. In Nigeria, the company's major focus was to discredit the opposition party's credibility along with its flag bearer through emotive messaging online and computational propaganda.[92]

According to Bradshaw and Howard, in a report on computational propaganda around the world that analysed computational propaganda in 70 countries, between 2017 and 2019, there has been a 150% increase in countries using organised social media manipulation campaigns.[93] Eleven African countries were surveyed and all of them are taking part in social media manipulation. Out of the 26 countries categorised as authoritarian, seven were from Africa and they use social media manipulation to suppress, discredit or drown out opposing voices. This activity is carried out on various social media platforms and computational propaganda from African countries like Egypt, Eritrea, Nigeria, South Africa and others feature prominently on platforms like Facebook, Twitter and WhatsApp.

3.3.2 Targeted online violence

Most times, violence is targeted against persons or groups offline for the same reasons they are targeted online. Some of the reasons include the existing systematic and institutional discrimination which are transplanted to the online space. For example, in many local contexts, owing to several factors like colonialism, indoctrination,

[89] Solomon (n 14 above).
[90] As above.
[91] As above.
[92] As above.
[93] As above.

© University of Pretoria

miseducation, political philosophies, many groups are often treated with disdain which often leads to violence both offline and online.[94]

Most times, vulnerable groups like women, children, girls, sexual minorities, persons living with disability, migrants, refugees, are at the receiving ends of such violence but in some instances, public figures like politicians, celebrities, journalists and other persons with a certain reach and relationship with the public also become victims of these forms of violence.[95] Harmer and Lumsden have both referred to this kind of violence as 'online othering.'[96] Online othering according to them is a contestation of digital power that seeks to perpetuate offline discrimination and inequality in the online space. There are several forms and reasons for these kinds of online violence and these forms and causes are discussed in turn. It is also noteworthy that most of these forms often straddle behavioural sciences, new media studies and the role of the law in regulating a wild horse like cyberspace.

Targeted online violence may take several forms. They are often described as an online behaviour that is characterised by malicious sharing of content that aims to disparage or harass, based on protected characteristics.[97] Such characteristics are often defined along the lines of age, religion, race, nationality, sex, gender and so on. These characteristics are also noticeable in contentious social behaviours such as ableism, racism, Islamophobia, sexism, misogyny, homophobia and so on. Major forms of targeted online violence include cyberbullying and cyberaggression, online gender-based violence, online violence against children and online hate speech.[98] These forms are underpinned by harassment and intent to harm.

A Cyberstalking, cyberbullying and cyberaggression

Cyberstalking may be generally referred to as the use of communications or messages to repeatedly monitor another person through a computer or network system such that the person can reasonably fear that their life is in danger. In order to properly appreciate the idea and impacts of cyberstalking, it is important to understand stalking first. Stalking, which connotes the physical aspects of it, is usually when a person repeatedly follows or monitors another person and as such causes the person to live

[94] H Brown 'Violence against vulnerable groups' *Council of Europe* 37 42 https://www.corteidh.or.cr/tablas/r25587.pdf (accessed 15 January 2020).
[95] S Jeong *Internet of garbage* (2015) 24.
[96] E Harmer & K Lumsden 'Conclusion: researching 'online othering'—future agendas and lines of inquiry' E Harmer & K Lumsden (eds) *Online othering exploring digital violence and discrimination on the web* (2019) 380-381.
[97] SC Herring 'Cyber violence: recognizing and resisting abuse in online environments' (2002) 14 *Asian Women* 187.
[98] O Bogolyubova *et al* 'Dark personalities on Facebook: Harmful online behaviors and language' (2018) 78 *Computers in Human Behavior* 151-159.

© University of Pretoria

in fear or anxiety for their lives. Roberts aptly captures this by describing it as 'repeated intrusive pursuits that cause fear.'[99] It is these activities, when carried out online that may be referred to as cyberstalking. Typically, most cyberstalking incidents do not necessarily require the stalker and the victim to have any prior relationship.

It is important to note however, that there are at least three essential elements of cyberstalking legislation that must be considered before it can be properly called and constitute an offence. First, the communication must have been more than once, that is it must be repeated for two or more times. Second, there must be established intent for the other person to be in fear or anxiety for their life. Third, the victim must experience either physical, emotional or fear for their safety. While determination of whether cyberstalking has occurred will be left to the courts to decide, the three elements are to be considered altogether for a crime of cyberstalking to occur. The degree and extent to which they occur might then add to the final decision of the court as to whether the cyberstalking occurred and what punishment is suitable.

In terms of cyberbullying, it did not gain much prominence until recently as bullying was still considered in its traditional form.[100] The typical features of traditional bullying, limiting its concept to space and time made it somewhat difficult to think of bullying outside such parameters. According to Citron, due to contentious online behaviours that target individuals based on certain characteristics like sexuality, migration history, age, race, gender and so on, it has become necessary to relate such intentions located in traditional bullying to the online space.[101]

However, understanding these contexts do not make conceptualising cyberbullying less difficult. This may be due to two major reasons. The first reason is that the psychology associated with cyberbullying has intersecting and complex dynamics such that its meaning, even though underpinned as seeking to malign, is constantly evolving. For example, anonymity online is as much a right as it is a potential threat to victims of online bullying. The second reason is that psychology is fast evolving, just like the technologies, which are often the most basic tool used in cyberbullying. For example, the platforms used in carrying out cyberbullying were not as much as they are now and so is the reach of these platforms. Corcoran and Guckin capture such difficulty stating that:

> … attempting to operationally define cyberbullying in a world which is in constant flux, could be likened to asking time to stand still. The evolving features of the available technology only intensify the unique nature of the communication. Indeed, whilst we debate and dialogue about the defining characteristics of cyberbullying, we must remain cognisant that by the time we reach

[99] L Roberts 'Jurisdictional and definitional concerns with computer-mediated interpersonal crimes: An analysis on cyber stalking' (2008) 2 *International Journal of Cyber Criminology* 272.
[100] Agrafiotis *et al* (n 10 above).
[101] Citron (n 2 above) 14.

© University of Pretoria

> some form of consensus, children and adolescents will, in all likelihood, be using technology and social communication tools that do not yet exist.[102]

However, Besley attempted a definition of the term cyberbullying as 'the use of information and communication technologies to support deliberate, repeated, and hostile behaviour by an individual or group that is intended to harm others.'[103] To Smith *et al*, it is 'aggressive intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend him or herself.'[104] Tokunaga also considered the concept as 'any behaviour performed through electronic or digital media by individuals or groups that repeatedly communicates hostile or aggressive messages intended to inflict harm or discomfort on others.'[105]

In comparison and picking from the salient features that cut across these definitions, Langos highlighted at least four elements, which is similar to traditional bullying as applicable to cyberbullying. They are repetition, power imbalance, intention, and aggression.[106] These features are best understood when communication modes in a cyberbullying activity are analysed. In doing this, Langos categorised these modes as direct and indirect communication.[107]

Direct communication in cyberbullying refers to instances where the bully has direct communication access to a victim, through for example text messages, emails and direct messages. With this form of communication, repetition, which is an unhinged and multiple use of communication to harass others, are often rife. Also, intention to harm is clear and so is power balance. The power balance here is often demonstrated when the victim is in fear of the tool of blackmail or violence that the bully may have and may be used against them. For indirect communication, repetition is often toned down due to the reach of the means of communication. Indirect communication in cyberbullying is when the tool being used is open and accessible to others than the bully or bullies and the victims. For example, social media platforms. Here, repetition is the only feature of cyberbullying that is not as prominent as power imbalance, intention and aggression.

---

[102] L Corcoran *et al* 'Cyberbullying or cyber aggression?: A review of existing definitions of cyber-based peer-to-peer aggression' (2015) 5 *Societies* 247.
[103] B Belsey 'Cyberbullying: An emerging threat to the "always on" generation' (2005); S Bauman & A Bellmore 'New directions in cyberbullying research' (2015) *Journal of School Violence* 2.
[104] PK Smith *et al* 'The nature of cyberbullying, and an international network' (2013) *Cyberbullying through the new media: Findings from an international network* 4.
[105] RS Tokunaga 'Following you home from school: A critical review and synthesis of research on cyberbullying victimization' (2010) 26 *Computers in Human Behaviour* 279.
[106] See C Langos 'Cyberbullying: The challenge to define' (2012) 15 *Cyberpsychology, Behavior, And Social Networking* 286, 288.
[107] Langos (n 106 above) 286.

© University of Pretoria

Notwithstanding the environment, most cyberbullying occurs through peer-to-peer interactions or involves adult(s).[108] Peer-to-peer harm in online violence against children often occurs when a child engenders harm against another child or group of other children through any technological means. The adult-child harm in online violence against children is when an adult carries out violence or harm against a child through any technological means. According to Farhangpour *et al* in a study on cyberbullying and its effects in a rural area in South Africa:

> majority of participants had access to cyber technology and used Facebook frequently. More than half of the participants experienced a wide variety of cyberbullying, sexual offence being the highest. They were negatively affected both emotionally and academically to the extent that some thought of suicide.[109]

The nature of aggression, which is a major component of cyberbullying point to a need for a broader application of the term cyberbullying. In understanding cyberbullying better, it is important to turn to cyberaggression and how it furthers the research on cyberbullying. Closely related to cyberbullying in practice and definition, cyberaggression is a form of negative online behaviour that targets various victims with the sole intent to harm them. According to Bushman and Anderson, 'aggression', of which 'cyberaggression' is a subset, often involves the intention of causing harm to a targeted individual or group, as opposed to accidental or unintentional harm.'[110] According to Grigg,[111] cyberaggression is:

> an intentional harm delivered by the use of electronic means to a person or a group of people irrespective of their age, who perceive(s) such acts as offensive, derogatory, harmful, or unwanted.

The purpose of cyberaggression is clear – it is to engender hostility that leads to violence against others online and offline. The concept of cyberaggression is as a result of the need to consider cyber bullying more broadly and how its harm-related content occurs.

Adopting a broader definition, Pyżalski considers cyberaggression beyond a peer-to-peer typology that cyberbullying is often considered as.[112] This showed that

---

[108] S Shariff & D Hoff 'Cyber bullying: Clarifying legal boundaries for school supervision in cyberspace' (2007) 1 *International Journal of Cyber Criminology* 84.

[109] P Farhangpour *et al* 'Emotional and academic effects of cyberbullying on students in a rural high school in the Limpopo province, South Africa' (2019) 21 *South African Journal of Information Management* 8.

[110] BJ Bushman & CA Anderson 'Is it time to pull the plug on hostile versus instrumental aggression dichotomy?' (2001) 108 *Psychological Review* 274.

[111] DW Grigg 'Cyber-aggression: Definition and concept of cyberbullying' (2010) 20 *Australian Journal of Guidance and Counselling* 143.

[112] J Pyżalski 'From cyberbullying to electronic aggression: Typology of the phenomenon' 17 *Emotional and Behavioural Difficulties* 305, 317.

© University of Pretoria

cyberaggression includes harms targeted beyond peers. Such harms are also targeted at persons beyond a certain demography to include vulnerable individuals, celebrities, public figures and more. In achieving this, Pyżalski considers intention to harm, repetition of messages to cause harm and power imbalance in arriving at the conclusion that cyberaggression is a notch higher than cyberbullying as it applies beyond set parameters of peer-to-peer targeted violence and includes larger scenarios.[113] Applying both Grigg's and Pyżalski's views, Corcoran and Guckin further defined cyberaggression as:

> any behaviour enacted through the use of information and communication technologies that is intended to harm another person(s) that the target person(s) wants to avoid. Intent to cause harm should be judged on the basis of how a reasonable person would assess intent.[114]

Cyberbullying has often been used to categorise all forms of potentially harmful, unwanted and contentious behaviours online, targeting individuals or a group of people, but does not pay enough attention to the harm-content of such activities. Considering the literature on cyberaggression, it points to the harm in cyberbullying in order to accommodate effective and necessary policy responses needed to regulate it.

In understanding the above more, it is useful to understand the primary actors in a bullying transaction – the aggressor, the victim and an audience.[115] Oftentimes, the aggressor is the person or a group of persons who initiate a threat or actual harm on another person. The victim is often on the receiving end of a bullying incident, that is, they are the person or a group of persons targeted to be harmed by an aggressor. The audience however are those persons or groups of persons who either witnessed the violence or harm being committed or witnessed it after it was committed. These three actors are often involved in the chain of occurrences during a bullying activity.

Since the major motivation of an aggressor is to demean the victim, it is likely that he carries out his physical threat or actual infliction of harm with an audience. This does not however suggest that such threat or infliction are always carried out with an audience present. In some instances, also depending on the nature of threat or harm against the victim, the violent transaction is limited to just the aggressor and the victim. On the other hand, cyberbullying is often a one-way transaction, aided by technology tools to target a specific victim or a group of victims. Where an audience is involved, it is often after the action of harm has been carried out online in order to further spread the reach of the harm against the victim. Therefore, while traditional bullying may actively carry an audience along in its infliction of harm, most incidents of cyberbullying are often after such infliction has occurred.[116]

---

[113] As above.
[114] Corcoran *et al* (n 102 above).
[115] Grigg (n 111 above).
[116] Farhangpour *et al* (n 109 above).

© University of Pretoria

Second, in most instances of traditional bullying, the aggressor is known to the victim. Whether in a controlled setting or not, the victim can often recognise who their aggressor is. This may be due to the fact that most traditional forms of bullying are physical. However, in cyberbullying, it is not often the case that the victim recognises his or her aggressor. This may be due to the high possibility of anonymising identities, which may be used to perpetuate harm.

Third, traditional bullying may be easily corrected and handled when parties to the harmful transactions are known to each other. It is easier when these parties are in a controlled setting and an audience is present. This gives opportunities for redress for such occurrences. However, due to the fast-paced nature and amplification of content, cyberbullying may be difficult to control as it can be shared several times on several platforms and also saved on devices to be used in the future against a victim. Therefore, cyberbullying tends to inure in time when compared to traditional bullying even though they are both harmful and violent.[117]

Some of the earliest laws on cybercrimes in Africa made provisions for the offences of cyberharassment, cyberstalking and cyberbullying. Many have the provisions captured as cyberstalking while some also provide for offences of cyberharassment and cyberbullying separately. Examples include Nigeria,[118] Uganda,[119] Malawi,[120] Tanzania,[121] Ethiopia[122] and South Africa.[123]

A closer look at the various provisions of cybercrime laws in Africa shows that there is disconnection between the laws and protecting the victim.[124] There are at least three major reasons for this. First, most cybercrime laws often leave the realm of criminalising cyber fraud, protecting critical infrastructure and other major objectives a cybercrime legislature ought to serve and venture into provisions that have

[117] N Saloojee 'The prevalence of traditional bullying and cyberbullying among university students' unpublished Masters thesis, University of Witwatersrand, 2019 13.
[118] Section 24 of the Cybercrimes (Prohibition, Prevention etc) Act of 2015 provides for the offence of cyberstalking. See section 5.2.1 for more detailed analyses of the provision.
[119] Sections 24 & 26 of the Computer Misuse Act of 2011 provide for the offences of cyberharassment and cyberstalking respectively.
[120] Sections 86 & 88 of the Electronic Transactions and Cybersecurity Act of 2016 provide for the offences of cyberharassment and cyberstalking.
[121] Section 23 of the Cybercrimes Act of 2015 provides for the offence of cyberbullying.
[122] Section 13 of the Computer Crime Proclamation No 958/2016 provides for the offence of crimes against liberty or reputation of persons which includes intimidating or causing another to fear for their safety.
[123] Sections 14,15 & 16 of the Cybercrimes Act of South Africa criminalise the offences of messages that incite damage to property or violence; threatens damage to property or violence and non-consensual disclosure of intimate messages respectively.
[124] See Section 5.2 below.

© University of Pretoria

implications for protecting human rights online. Second, most laws conflate concepts like cyberbullying with cyberstalking which are completely different terms.[125] Third, the scope of criminalised actions, including speech that seek to bear on cyberbullying are broad and use vague words capable of being arbitrarily exercised by governments and governments institutions.[126]

B Online gender-based violence (GBV)

According to a report by Iyer *et al*, online gender-based violence is a specific form of harm engendered through electronic communication that targets based on sexual or gender norms and characteristics.[127] Both in law and praxis, online gender-based violence often targets women, girls and sexual minorities. This targeting is as a result of reinforcement of existing socio-cultural stereotypes against these groups of persons. In another research, online gender-based violence has been identified as technology-assisted or –related harm.[128]. Online gender-based violence includes non-consensual sharing of intimate images, use of manipulated media like deepfakes or shallowfakes, stalking, harassment, doxing, blackmail/threats, surveillance and hacking.

In some instances, non-consensual sharing of intimate images has been referred to as revenge porn.[129] Examining non-consensual sharing of intimate images, it is a broader approach required to understand the harm inflicted on its victims many of whom are women. The term revenge porn clearly assumes that the media content is always pornographic and it is done in retaliation or reprisal for an act. Non-consensual sharing of intimate images as a term is open in meaning to include pornography – an image intended to sexually arouse or any other form of images that are not intended to sexually arouse. Also, sharing of such images does not have to be limited to cases when it was done in reprisal. Non-consensual images shared as a result of

[125] As above.

[126] R Adibe *et al* 'Press freedom and Nigeria's Cybercrime Act of 2015: An assessment (2017) 52 *Africa Spectrum* 117 127; D Marari 'Of Tanzania's cybercrimes law and the threat to freedom of expression and information' 25 May 2015 *AfricLaw* https://africlaw.com/2015/05/25/of-tanzanias-cybercrimes-law-and-the-threat-to-freedom-of-expression-and-information/ (accessed 20 January 2020); ARTICLE 19 'Uganda: Government must safeguard freedom of expression after arrest and attack' 18 April 2017 https://www.article19.org/resources/uganda-government-must-safeguard-freedom-of-expression-after-arrest-and-attack/ (accessed 15 January 2018).

[127] N Iyer *et al* 'Alternate realities, alternate internets: African feminist research for a feminist Internet' (2020) *Pollicy* 10 https://www.apc.org/sites/default/files/Report_FINAL.pdf (accessed 15 October 2020); C Badenhorst 'Legal responses to cyberbullying and sexting in South Africa' *Centre for Justice and Crime Prevention* 2001; N Malhotra 'End violence: Women's rights and safety online' https://www.apc.org/sites/default/files/end_violence_malhotra_dig.pdf (accessed 15 October 2020).

[128] Malhotra (n 127 above).

[129] C McGlynn *et al* 'Beyond 'Revenge Porn': The continuum of image-based sexual abuse' (2017) 25 *Feminist Legal Studies* 26 https://link.springer.com/article/10.1007/s10691-017-9343-2 (accessed 16 October 2020).

© University of Pretoria

intoxication, hacking, stolen gadgets and so on. will also qualify as examples of non-consensual sharing of intimate images. While non-consensual sharing of intimate images is broad and envisages newer contexts where privacy could be violated online, revenge porn as a concept is narrow and limited to only instances where sexual images of a victim are shared as a form of retaliation or reprisal. Also, the operational word that sets both non-consensual sharing of intimate images and revenge porn apart, is 'non-consensual.' It suggests that lack of consent in sharing images of intimate interactions in the latter is the ground for violation while the former suggests revenge as a justifiable action and limits harm to pornography.

Previous research in seven countries, including Kenya and the Democratic Republic of Congo (DRC), showed that women who have been victims of online gender-based violence have suffered both emotional and psychological harm.[130] In another report by Plan International, 14 071 respondents made up of girls and young women in 22 countries including Nigeria stated that they have been harassed and abused online.[131] In an Africa-focused study on fighting online gender-based violence, 3 306 respondents between the ages of 18 - 65 from Ethiopia, Kenya, Senegal, South Africa and Uganda were interviewed. Most online violence took place on Facebook and WhatsApp and of those interviewed, at least 1 in 3 women had been through online gender-based violence.[132] Most of the time, women resolve to various self-help methods to protect themselves from online GBV. Some of these methods range from blocking or getting rid of the perpetrator to reporting to the authorities. An important observation to be made from these methods is that more women have had to alter their lives in order to deal with online GBV. This includes deleting their accounts, leaving an online platform, changing their phone number and reporting users to the online platform.[133] This raises an important point on the relationship between sharp offline and traditional methods of regulation of Online GBV as an online harm and the need for more effective and inclusive multistakeholder regulation of Online GBV.

In a comparative analysis of Ethiopia, Kenya, Senegal, South Africa and Uganda on fighting online gender-based violence, Nwaodike and Naidoo noted that 'legal frameworks are rarely fully representative of the practical realities in any country.'[134] This suggests that while there are laws in the various contexts examined, they do not

---

[130] The Women's Legal and Human Rights Bureau 'End violence: Women's rights and safety online from impunity to justice: Domestic legal remedies for cases of technology-related violence against women' (2015) https://www.genderit.org/sites/default/files/flow_domestic_legal_remedies_0.pdf (accessed 15 October 2019).

[131] Plan International 'Free to be online? Girls and young women's experiences of online harassment' (2020) https://plan-international.org/publications/freetobeonline (accessed 1 December 2021).

[132] Iyer *et al* (n 127 above).

[133] As above.

[134] C Nwaodike & N Naidoo 'Fighting violence against women online: A comparative analysis of legal frameworks In Ethiopia, Kenya, Senegal, South Africa, and Uganda' (2020) *Pollicy* 5.

© University of Pretoria

carry the essence of justice for victims of Online GBV because they are 'rarely fully representative of the practical realities.' The analysis shows that not only are the laws inadequate with various conflation of concepts, there is hardly any possibility that should online violence occur in these countries, they will make it to court.

C Online violence against children

According to a study conducted in 2016, at least half of the world's one billion children population between the ages of 2-17 years have experienced some form of violence.[135] The study also shows that 50% of children of the countries surveyed in Africa, put at 200 million, have also been victims of various forms of violence. The growing use of online platforms by children to network or for education further increases the risks they face.

Online violence against children may refer to technology-mediated harm targeted at children.[136] The age range for children is between 1-18 years. These forms of violence may include child sexual abuse or images, sexual or non-sexual harassment, bullying, harassment, stalking and sextortion. These forms of violence may be recurrent or one-off but how frequent they are does not necessarily reduce both the short- or long-term impacts of harm that may be caused. Some of the short-term impacts of these forms often include peer isolation, withdrawals, negative impacts on education, shaming, low self-esteem, bullying and so on.[137] The long-term impacts may be physical harm like self-harm, suicide, mental health issues, reduced social interaction and so on. These various forms may also be carried out as peer-to-peer harm or adult-child harm.

A study by the United Nations Children's Fund (UNICEF) has shown that understanding online violence against children requires many solutions, one of which is understudying the deeper trends between traditional harassment and cyberbullying.[138] In a report by the Special Representative of the UN Secretary-General on Violence against Children stated that 'a separate discussion of traditional

[135] S Hillis *et al* 'Global prevalence of past-year violence against children: a systematic review and minimum estimates' (2016) 137 *Pediatrics* 6 https://pediatrics.aappublications.org/content/pediatrics/137/3/e20154079.full.pdf (accessed 16 October 2020).
[136] As above.
[137] C Li & F Lalani 'Why so much harmful content has proliferated online - and what we can do about it' 13 January 2020 *World Economic Forum* https://www.weforum.org/agenda/2020/01/harmful-content-proliferated-online/ (accessed 17 October 2020).
[138] S Livingstone & M Bulger 'A global agenda for children's rights in the digital age recommendations for developing UNICEF's research strategy' (2013) *UNICEF* https://www.unicef-irc.org/publications/pdf/lse%20olol%20final3.pdf (accessed 27 October 2020).

© University of Pretoria

bullying and cyberbullying definitions, incidence and policy miss the deeper trend, which is to recognise the increasing connections between the two.'[139]

Online violence against children in Africa, including Child Sexual Abuse Material (CSAM), now has a sharper focus for discussion due to more dependence on various children-specific sectors like the media, educational and social justice on the Internet. The various interlinkages between each sector and how they must protect the African child in the digital age is multifarious as it is often understated. This is seen in the inadequate criminalisation and implementation of the challenge in most African countries.[140] Also, currently there are no specific reporting obligations by online platforms and intermediaries, the differentiated legal approaches, inadequate reporting mechanisms and a region-specific direction on online harms, online violence against children and non-legal approaches in African countries.[141]

In addition to this, there is a connection between violent online behaviour and violent behaviour in school among children, which may lead to experiencing and committing violent behaviour online.[142] In assessing the impact of such findings, researchers argue that:

> The consequences of violence on the internet can sometimes be even more serious than of which happened in real-life situations. In violence on the internet, there is a power of a written word,

[139] Office of the Special Representative of the Secretary-General on Violence Against Children 'Ending the torment: Tackling bullying from the schoolyard to the cyberspace' (2016) https://violenceagainstchildren.un.org/sites/violenceagainstchildren.un.org/files/documents/publications/tackling_bullying_from_schoolyard_to_cyberspace_low_res_fa.pdf (accessed 25 July 2020).

[140] Even though many African countries criminalise violence against children, this criminalisation does not include cyberviolence and its impacts on children. Criminalisation of violence against children is often carried out through criminal codes, children-specific legislation and cybercrime laws and rarely cover beyond child pornography to include cyber bullying, exposure to violent content, impacts of post-exposure to online violence etc. See Nigerian Communications Commission (NCC) 'Final report: Study on young children and digital technology' September 2021 https://www.ncc.gov.ng/accessible/documents/1005-young-children-and-digital-technology-a-survey-across-nigeria/file (accessed 1 December 2021); C Monyei 'Children's online safety in Nigeria: the government's critical role' 12 September 2018 https://blogs.lse.ac.uk/parenting4digitalfuture/2018/09/12/childrens-online-safety-in-nigeria/ (accessed 1 December 2021); J Phyfer *et al* 'South African Kids Online: Barriers, opportunities and risks. A glimpse into South African children's internet use and online activities' (2016) *Centre for Justice and Crime Prevention* http://globalkidsonline.net/wp-content/uploads/2016/06/GKO_Country-Report_South-Africa_CJCP_upload.pdf (accessed 1 December 2021).

[141] B Gacengo 'Online child sexual exploitation' 15 August 2020 *Council of Europe* https://rm.coe.int/3148-afc2018-ws9-ecpat-manifestations/16808e85b8 (accessed 12 February 2021).

[142] See D Perišin & S Opić 'Connection between exposure to Internet content and violent behaviour among students' (2013) *The 1st International conference "Research and education challenges toward the future* https://files.eric.ed.gov/fulltext/ED565463.pdf (accessed 15 June 2020).

© University of Pretoria

because the victim can re-read what the bully wrote about them, and insults in verbal form can easily be forgotten. A written word acts more concrete and realistic than the spoken one.[143]

D Online hate speech

While hate speech in its various manifestations is not new, its online version is perhaps one of the most definitive and familiar forms of online harms.[144] Understanding the context and application of online harms especially from the international human rights law perspective is also helpful as it is the only form of speech referred to as 'prohibited.'[145] While the international human rights law does not define hate speech under article 20 of the ICCPR along whether such speech is made online or offline, the United Nations has in the past stated that the rights that people have offline are also the same they have online.[146] This presupposes the argument that irrespective of a specific mention of online hate speech under the international system, the parameters used in assessing whether a speech is hateful and dangerous is the same whether online or offline.

For more definitional clarity, this chapter groups concepts of online hate speech into three approaches. They are the normative approach, the platform approach and the academic or scholarship approach. The normative approach is the various international law positions on what constitutes hate speech. As explained in the previous chapter, various treaties and conventions like the ICCPR, the ICERD, CRPD, CRC all have working definitions of what constitutes hate speech. A cross-cutting definition from these treaties are that hate speech are those forms of expression that constitute incitements that are targeted towards protected categories of vulnerable persons like persons living with disability, women and girls, sexual minorities, migrants and several others. It could also be such speech that targets violence against persons based on their sexual preferences, age, race, nationality or identification. Such speech also must connect the actual speaker to a crime or must be such that when context is applied, such speaker's speech is likely to cause violence against another.

Today, online platforms, especially the social media platforms, wield a powerful role in the dynamics of what broadly forms free speech and specifically what constitutes various forms of restricting it. As established by the international human rights treaty system, hate speech is one of such widely acceptable forms of restrictions to free speech, and online platforms have been at the forefront of defining the term under their various platform governance systems in general and specifically through their various regulatory policies like community guidelines and policies. Therefore, a platform approach to defining online hate speech is also necessary as they continue

[143] As above; Citron (n 2 above).
[144] Citron (n 2 above).
[145] Section 2.4.3 above.
[146] United Nations General Assembly 'General Comment No 34, CCPR/C/GC/34' 12 September 2011 http://undocs.org/en/CCPR/C/GC/34 (accessed 12 June 2021).

© University of Pretoria

to have an increased role in free speech policy. As a result, platform approach to online hate speech may be referred to as the several ways social media sites seek to conceptualise the term and regulate it based on such conceptualisation. A number of examples demonstrate this.

For example, Facebook defines hate speech as:

> a direct attack on people based on what we call protected characteristics—race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability.[147]

It goes further to define such an attack as 'a violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation.'

For YouTube, the definition of hate speech is denoted by an active statement to:

> remove content promoting violence or hatred against individuals or groups based on any of the following attributes: age, caste, disability, ethnicity, gender identity and expression, nationality, race, immigration status, religion, sex/gender, sexual orientation, victims of a major violent event and their kin, veteran status.[148]

Twitter also reads almost the same with YouTube's active tone but similar to Facebook's language when it couched hate speech as those tweets that:

> promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease.[149]

While these examples of platforms are not all there are to platforms' approach to online hate speech as there are many other platforms who also have definitional policies on online hate speech, they are some of the most popular with an estimated combined number of users who are on these platforms make at least 4.2 billion.[150] A closer look at these definitions also provide for a comparative value. A similarity between these three examples are various mentions of specific examples of categories that may be affected by online hate speech.

Both Facebook and Twitter make specific mention of discrimination based on 'serious disease' as one of the categories of hate speech on their platforms while YouTube

---

[147]Facebook Community Standards 'Hate Speech' https://www.facebook.com/communitystandards/hate_speech (accessed 19 November 2019).

[148] YouTube 'Hate Speech Policy' https://support.google.com/youtube/answer/2801939?hl=en (accessed 19 November 2019).

[149] Twitter 'Hateful Conduct Policy' https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy (accessed 19 November 2019).

[150] We are social 'Digital 2021: Digital overview report' January 2021 https://wearesocial-cn.s3.cn-north-1.amazonaws.com.cn/common/digital2021/digital-2021-global.pdf (accessed 13 October 2021).

© University of Pretoria

included 'veteran status' which suggests a US-specific category as the term is commonly used in the US for retired soldiers. YouTube also makes a specific mention of 'immigration status' and 'nationality' as some of the categories protected against hate speech while Twitter and Facebook use 'national origin.'

In accommodating newer realities as a result of technologies and their impacts on the right to freedom of expression, several scholars have also sought to define online hate speech. According to Onanuga, hate speech is 'any online or offline communication that expresses hatred for some group, in terms of race, ethnicity, gender, religious, sexual orientation and others defining attributes of mankind.'[151] It is a technology-mediated speech that is targeted to cause or incite violence against a person based on certain characteristics.

Arguments for its regulation can be divided into three: non-legal sanction approach, legal sanction approach and the hybrid approach. For the non-legal approach, scholars are more focused towards understanding the concept and various manifestations of hate speech as it is both a social and legal challenge, which makes it even more difficult to easily grasp.[152] The legal sanction approach scholars advocate for a direct restriction of such speech and attaching legally enforceable sanctions against them, and they are invested in the guiding policy on what may constitute lawful language.[153] The hybrid approach is more in between – they are open to understudying the various manifestations of hate speech online but caution that defining what constitutes lawful language or not through the law should be the last resort.[154]

A definition of hate speech that demonstrates the non-legal approach is:

> any form of speech that produces the harms which advocates for suppression ascribe to hate speech: loss of self-esteem, economic and social subordination, physical and mental stress, silencing of the victim, and effective exclusion from the political arena.[155]

---

[151] B Onanuga 'Roots of hate speech, remedies' (2018) Workshop on hate communication in Nigeria: Identifying its roots and remedies.

[152] CR Massey 'Hate speech, cultural diversity, and the foundational paradigms of free expression' 40 *UCLA Law Review* (1992) 103.

[153] MJ Matsuda 'Public response to racist speech: Considering the victim's story' 87 *Michigan Law Review* (1989) 2320.

[154] RA Wilson & MK Land 'Hate speech on social Media: Towards a context-specific content moderation policy' 52 *Connecticut Law Review* 47; B Parekh 'Is there a case for banning hate speech?' in M Herz & P Molnar (eds)*The content and context of hate speech*: *Rethinking regulations and responses* (2012) 46; N Jansen Reventlow *et al* 'Perspectives on harmful speech online' (2016) *Berkman Klein Center for Internet & Society Research Publication* https://dash.harvard.edu/bitstream/handle/1/33746096/2017-08_harmfulspeech.pdf?sequence=5&isAllowed=y (accessed 23 July 2019).

[155] Massey (n 152 above).

© University of Pretoria

Moran also argues that hate speech is 'complex social and cultural phenomena' and defines it as speech that is 'intended to promote hatred against traditionally disadvantaged groups.'[156]

For the second approach on regulation on hate speech, Waldron's position seems to rank prominently. He noted that:

> By "hate speech regulation," I mean regulation of the sort that can be found in Canada, Denmark, Germany, New Zealand, and the United Kingdom, prohibiting public statements that incite "hatred against any identifiable group where such incitement is likely to lead to a breach of the peace" (Canada); or statements "by which a group of people are threatened, derided or degraded because of their race, colour of skin, national or ethnic back- ground" (Denmark); or attacks on "the human dignity of others by insulting, maliciously maligning or defaming segments of the population" (Germany); or "threatening, abusive, or insulting . . . words likely to excite hostility against or bring into contempt any group of persons.. on the ground of the colour, race, or ethnic or national or ethnic origins of that group of persons" (New Zealand); or the use of "threatening, abusive or insulting words or behaviour," when these are intended "to stir up racial hatred," or when "having regard to all the circumstances racial hatred is likely to be stirred up thereby" (United Kingdom).[157]

His view is widely supported by the many international human rights treaties and also national laws across the world including African countries that have proscribed hate speech in their various provisions.[158]

In the third approach, Parekh, listing a number of examples of what may constitute hate speech made a succinct clarification between what may be offensive speech, often conflated with hate speech and what actually constitutes hate speech. He notes that:

> Hate speech expresses, encourages, stirs up, or incites *hatred* against a group of individuals distinguished by a particular feature or set of features such as race, ethnicity, gender, religion, nationality, and sexual orientation. Hatred is not the same as lack of respect or even positive disrespect, dislike, disapproval, or a demeaning view of others.[159]

In what distinguishes his approach on hate speech from others, he argued that hate speech is a problem and is unacceptable in any society. However, legal prohibition should be the last resort. He noted:

> The difficult and much-debated question is whether it should be not merely discouraged by moral and social pressure but prohibited by law. Although law must be our last resort, its intervention cannot be ruled out for several important reasons. Most obviously, assuming meaningful levels

[156] M Moran 'Talking about hate speech: A rhetorical analysis of American and Canadian approaches to the regulation of hate speech' (1994) *Wisconsin Law Review* 1428 1430.
[157] J Waldron The harm in hate speech (2012) 8.
[158] Section 2.5 above; Section 5.2.1 E below.
[159] Parekh (n 154 above) 40.

© University of Pretoria

> of enforcement and compliance, direct prohibition would reduce or eliminate speech that causes very real harm to the targets of such speech.[160]

Supporting his argument on the law being the last resort, he however noted that an outright ban on a specific form of speech like hate speech will be necessary in combating it, but that context and nuance will be important in such an instance, especially with respect to instances where authoritarian practices tend to conflate these kinds of harmful speech with political and unpopular but necessary speech in a democratic society.[161]

A look at all these definitions and arguments for how they are couched in various contexts and language point to one thing – that hate speech is undesirable in any society. That this view is shared by three of the main impactful stakeholders on online hate speech regulation is noteworthy and also suggests its harmful nature.

Most academic works on online hate speech are Global North-focused. There are limited works that analyses online hate speech within the context of online harms especially in Africa. In understanding hate speech in Africa, Azogwa and Ezeibe connected religion and ethnicity as its major drivers.[162] These drivers, according to them, were also closely linked to the various colonial legacies in many African countries. Since most hate speech is primarily so because of the likelihood of violence that may be associated with it, major political events, especially elections in Africa have witnessed a lot of challenges as a result of hate speech. These challenges range from championing ethnically-charged political speeches to real-life, offline harms in various African systems. According to Ezeibe *et al*, 'hate speech has become a strategic aspect of electioneering today, such that numerous election-related conflicts in Africa bear proximate connection to their use.'[163]

With a series of more specific examples, and testing Azogwa and Ezeibe's thesis, there have been country-level case studies on hate speech in a number of African countries and their causes. For example, in Ethiopia, according to Abraha, 'there is no doubt that hateful speech and disinformation have contributed significantly to the unfolding polarized political climate, ethnic violence and displacement in Ethiopia.'[164] Between 2018 and 2020 alone in Ethiopia, there have been various scales of violence

---

[160] Parekh (n 154 above) 46.
[161] Parekh (n 154 above) 55-56.
[162] N Asogwa & C Ezeibe 'The state, hate speech regulation and sustainable democracy in Africa: A study of Nigeria and Kenya' (2020) *African Identities* 1.
[163] CC Ezeibe & OM Ikeanyibe 'Ethnic politics, hate speech, and access to political power in Nigeria' 63 *Africa Today* 66.
[164] A Madebo 'Social media, the diaspora, and the politics of ethnicity in Ethiopia' 29 October 2020 *Democracy in Africa* http://democracyinafrica.org/social-media-the-diaspora-and-the-politics-of-ethnicity-in-ethiopia/ (accessed 23 June 2020).

© University of Pretoria

including the Amhara assassinations, Sidama riots and the murder of Hacahalu Hundessa[165], 'several other incidents, concern over the role of hate speech and disinformation/misinformation online has become a mainstream agenda item.'[166] As briefly discussed in the previous chapter, Ethiopia has a law on hate speech.[167]

In Nigeria, similar to Ethiopia in the shared characteristics of diverse ethnic groups, hate speech online is also a source of concern. Nigeria's experience with hate speech online is uniquely multifaceted. According to the Nigeria Stability and Reconciliation Programme (NSRP), using an automated process to monitor most hate speech content in Nigeria[168], it found that content is mostly shared on Facebook and Twitter.[169] 76% of online hate speech in Nigeria is shared on Facebook while the remainder messages are shared on Twitter and online articles. While the methodology does not state which of the platforms analysed or how it carried out its automated monitoring, there is a strong connection between large social media platforms like Facebook and Twitter and spreading of online hate speech in Nigeria. In its categorisation of what hate speech is and how it reflects in the survey, at least 45% of the messages were calls for discrimination, 38% for war, while 10% advocates for the killing of others. Around 75% of the responses received moderate to significant responses.[170] Nigeria is currently debating its hate speech bill before its National Assembly.[171]

In connection to both Nigeria and Ethiopia, South Africa shares a history of violence that has and continues to precipitate online hate speech. In an analysis of online hate speech regulation in South Africa, Nkrumah highlights at least ten challenges.[172] He notes the issues of over-regulation of hate content, contextual challenges of what constitutes hate speech, jurisdiction and others. In particular, he notes that the multi-racial and cultural history would pose threats to hate speech regulation in South Africa. This observation is particularly important because while domestic regulation might be alive to the contextual nuances of hate speech, external actors like other governments and social media platforms might not understand such context.

---

[165] E Chala 'How the murder of musician Hachalu Hundessa incited violence in Ethiopia: Part II' 7 August 2020 *Global Voices* https://globalvoices.org/2020/08/07/how-the-murder-of-musician-hachalu-hundessa-incited-violence-in-ethiopia-part-ii/ (accessed 23 September 2020).

[166] B Taye & J Pallero 'Ethiopia's hate speech predicament: Seeking antidotes beyond a legislative response' 27 July 2020 *Access Now* https://www.accessnow.org/open-letter-to-facebook-protect-ethiopians/ (accessed 15 August 2020).

[167] Section 2.5.1 above.

[168] NSRP 'How-to guide: Mitigating dangerous speech: Monitoring and countering dangerous speech to reduce violence' (2017) http://www.nsrp-nigeria.org/wp-content/uploads/2017/12/NSRP-How-to-Guide-Mitigating-Hate-and-Dangerous-Speech.pdf (accessed 15 September 2020).

[169] This provides more information on the methodology sheet used in the automation process. Summary Sheet https://bit.ly/3qECucH (accessed 16 September 2020).

[170] NSRP (n 168 above).

[171] Section 5.2.1 below.

[172] B Nkrumah 'Words that wound: Rethinking online hate speech in South Africa' (2018) 23 *Alternation Journal* 118-123.

© University of Pretoria

3.4 Engendering online harms

Online harms may be carried out in two major ways. The primary and secondary methods. The primary methods are the various ways through which online harms are produced. There have been three identified ways through which primary methods may occur. They are through emotive narratives and constructs, fabricated multimedia and artificial online entities. Secondary methods involve the actors that disseminate these harms at various points and the means of their dissemination. These can be broadly divided into two: actors and dissemination. The major relationship between both methods is that they are both designed to cause online harms. The major difference is that while the primary method relates to the various ways in which online harms are designed to harm, the secondary method focuses on how those designed harms are shared and spread.

3.4.1 Primary methods of online harms

A Emotive narratives and constructs

This primary method is characterised by strong emotional messaging and technology tools mixed with elements of truth.[173] Its aim is to paint a set of facts in a different light by adding outright deception or part-truths. It is usually made up of three main components: emotional content, misleading narrative and digital messaging. The emotional content in emotive narratives are aspects of the messaging which may be true in part, may be part of the content or the entire content that seek to subjectively sway an audience.

The misleading narrative is often the elements of outright deception and part-truths in an emotive narrative while the digital messaging is the form through which such narrative is carried out. Usually, it is through multimedia like texts, images, videos or a combination of these and other online communication tools. Emotive narratives and constructs weaponise emotions through online multimedia.[174] A popular example of emotive narratives and constructs is clickbait. According to Chen *et al*, 'a key variable in clickbait is emotion.'[175] This method demonstrates the relationship between society, sociology and technologies – how technologies sit at the centre of human deception.

---

[173] Bontcheva & Posetti (n 50 above) 22.

[174] S Hoffmann *et al* 'The market of disinformation' (2019) *Oxford Internet Institute, University of Oxford* 32 https://oxtec.oii.ox.ac.uk/wp-content/uploads/sites/115/2019/10/OxTEC-The-Market-of-Disinformation.pdf (accessed 28 September 2020).

[175] Y Chen *et al* 'Misleading online content: recognizing clickbait as "false news"' (2015) 17 http://dx.doi.org/10.1145/2823465.2823467 (accessed 30 July 2020).

17. M Guerini & J Staiano 'Deep feelings: A massive cross-lingual study on the relation between emotions and virality' https://arxiv.org/abs/1503.04723v1 (accessed 30 July 2020).

© University of Pretoria

Information disorder and targeted online violence online both use emotive narratives and constructs.

B Fabricated multimedia

These are various engineered and doctored information with the sole purpose of causing harm. Within the context of online harms, they are often used to create distrust. Oftentimes, the multimedia which may be text, images, videos or a combination of these and other online communication tools that may be completely manufactured, altered in contexts or deliberately de-contextualised. Examples of complete manufacture are staged multimedia like synthetic audio, Computer-Generated Images (CGI) while examples of altered contexts are manipulated media like deepfakes and shallowfakes etc.[176] Examples of deliberate de-contextualisation are the use of existing multimedia in contrasting contexts, for example sharing an old picture during the Ebola pandemic again during the COVID-19 pandemic without providing the context for the new usage. Examples of online harms that use fabricated multimedia are information disorder and targeted online violence.

C Artificial online entities

These are online entities that do not exist but are manufactured to deceive.[177] This type of method may be classified into two major types: authentic artificial online entities and inauthentic artificial online entities. Authentic artificial online entities are those who use new false identities in creating their online entities while the inauthentic artificial online entities are those who pass off the identities of other existing entities as theirs. For example, an authentic artificial online entity may be WWW.USA.GOV.US while an inauthentic artificial online entity may be WWW.CNN.ORG. Examples of online harms that may employ this method are information disorder and targeted online violence.

3.4.2 Secondary methods of online harms

These are two major aspects involved in the secondary methods of online harms. These aspects can be described as actors and dissemination. With respect to actors, there are several motivations for actors in engaging in online harms. Oftentimes, the motivations are political and in other cases economic, historical or based on several other factors.[178] In some instances, these factors combine to form these motivations for engaging in the production of online harms. These actors are broadly divided into state and non-state actors. With respect to dissemination which is often carried out by actors, it may be divided into automated and non-automated dissemination of online harms.

[176] Chesney & Citron (n 12 above) 1744 1788.
[177] Hoffmann *et al* (n 174 above).
[178] As above.

© University of Pretoria

A Actors

The state actors are often governments or government institutions who deliberately produce online harms like information disorder mostly for political gains.[179] Oftentimes, it has been found that state actors are the biggest purveyors of online harms especially information disorder given some of the reasons highlighted above.[180] Some of the prominent reasons for states getting involved in facilitating online harms are political. State actors can be broadly divided into foreign state actors and local state actors.[181] Foreign actors are those who are based outside a state but produce and ensure both covert and overt online harms in order to influence and interfere in a policy development in another country.[182] Local state actors are the in-country actors who design online harms within the context of a specific country which they reside in.[183] Many foreign actors conduct multi-country interference through online harms and often apply the same methods while most local state actors are often focused on one country but may apply different methods.[184] Examples of online harms that often employ this secondary method are information disorder, targeted and online violence.

For non-state actors, they neither belong to the state nor its institutions. Rather, they are often private companies, individuals or the general public. These actors weaponise technical expertise to construct alternate realities for the purpose of profit.[185] In other instances, governments link up with private businesses with the technical delivery that pose threats not only to the right to freedom of expression but also other interrelated rights like access to information, privacy, dignity, association, assemblies and so on.

---

[179] DA Martin *et al* 'Recent trends in online foreign influence efforts' (2019) 18 *Journal of Information Warfare* 15-48 https://scholar.princeton.edu/sites/default/files/jns/files/martin_d._shapiro_j._nedashkovskaya_m._recent_trends_in_online_foreign_influence_efforts.pdf (accessed 15 June 2020); C Ward *et al* 'Russian election meddling is back – via Ghana and Nigeria – and in your feed' 11 April 2020 *CNN* https://edition.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html (accessed 12 July 2020); J Stubbs 'French and Russian trolls wrestle for influence in Africa, Facebook says' 15 December 2020 *Reuters* https://www.reuters.com/article/facebook-africa-disinformation/french-and-russian-trolls-wrestle-for-influence-in-africa-facebook-says-idUKL8N2IV3NR?edition-redirect=uk (accessed 2 January 2021); A Essa 'China is buying African media's silence' 14 September 2018 *Foreign Policy* https://foreignpolicy.com/2018/09/14/china-is-buying-african-medias-silence/ (accessed 23 June 2020).

[180] Martin *et al* (n 179 above).

[181] Hoffmann *et al* (n 174 above).

[182] J Stubbs 'French and Russian trolls wrestle for influence in Africa, Facebook says' 15 December 2020 *Reuters* https://www.reuters.com/article/facebook-africa-disinformation/french-and-russian-trolls-wrestle-for-influence-in-africa-facebook-says-idUKL8N2IV3NR?edition-redirect=uk (accessed 10 January 2021).

[183] Hoffmann *et al* (n 174 above).

[184] As above.

[185] Chesney & Citron (n 12 above).

© University of Pretoria

These companies conduct both covert and overt mining of information at the expense of these rights in order to interfere in the business of democratisation. Oftentimes, governments, opposition parties connive with these companies to thwart the electoral process by cunningly shaping the outcome of a political process.[186] Here, we find the relationship between unethical business practices, autocratic governments and subverted technologies.

Most online platforms have as a result of various challenges ranging from lack of context to overreliance on automated systems, enabled these online harms on their platforms.[187] The debate on the level of complicity of these platforms and their actual capacity to address the issues are still ongoing and raging. However, there is a clear-cut responsibility for platforms, under international law to ensure that human rights are protected on their platforms.[188] Both states and non-states actors who produce and disseminate online harms often do it on online platforms. Therefore, online platforms can no longer be aloof.[189]

In addition, the general public also produces and shares online harms based on various motivations as explained above. While there are no fine numbers as to which of the actors share the most online harms, the general public which often includes public figures, celebrities, influencers and other online users all have capacities to sometimes produce but more often than not, share online harms. All of these non-state actors may be used for online harms like information disorder and targeted abuse online.

B Dissemination

In the dissemination of online harms, there are two major forms. They are automated and non-automated methods. Automated methods of dissemination are those carried out by trained machines to perform more human and specific tasks with less or no human supervision. They have been described as 'the backbone of techniques used to effectively 'manufacture' the amplification of disinformation.'[190] Some of these specific tasks within the context of spreading online harms are the use of bots and

[186] Ekdale & Tully (n 13 above).
[187] M Karanicolas 'The countries where democracy is most fragile are test subjects for platforms' content moderation policies' 16 November 2020 *Slate* https://slate.com/technology/2020/11/global-south-facebook-misinformation-content-moderation-policies.html (accessed 15 December 2020).
[188] Article 19(3) of the ICCPR on the right to freedom of expression requiring 'special duties and responsibilities' which have been interpreted to include such duties and responsibilities to be carried out by stakeholders. See A Callamard 'The human rights obligations of non-state actors' in RF Jørgensen (ed) *Human rights in the age of platforms* (2019) 199.
[189] T Ilori 'Content moderation is particularly hard in African countries' 21 August 2020 *Slate* https://slate.com/technology/2020/08/social-media-content-moderation-african-nations.html (accessed 21 August 2020).
[190] Wardle (n 24 above).

© University of Pretoria

algorithms to initiate coordinated behaviours like trolling, spamming, doxing etc.[191] A major form of this secondary method of dissemination is computational propaganda which has been described by Woolley and Howard as 'learning from and mimicking real people so as to manipulate public opinion across a diverse range of platforms and device networks.'[192] However, for automated methods of dissemination to spread online harms, it is often trained and as a result put in motion by a human agent – the machines learn what they are taught.

Non-automated methods, on the other hand, are the use of persons, other than machines, to spread online harms.[193] This method is often common in local contexts where the person(s) have substantial impacts on public policy. It may be a politician, public figure, celebrity etc. It is also the popular method the general public as non-state actors use in spreading online harms. In other instances, both methods of dissemination are combined to engender online harms. Information disorder, online hate speech, targeted online violence all utilize these methods of dissemination.

3.5 Harm versus illegality in classifying online harms

Depending on the forms and subforms of online harms, they all either have the propensity to harm or are patently harmful and in some instances, are both.[194] Harm here would mean either physical or emotional harm. Such propensity to harm would be that which does not immediately cause any violence or damage but may cause such in the long run. On the other hand, an online harm is patently harmful when the likelihood is high in inflicting harm on another. For example, information disorder as a form of online harm has both the propensity to harm and is patently harmful.[195] A subform of information disorder that may be categorised as having the propensity to harm is misinformation due to its lack of initial intent to cause harm but which may be hijacked on the long run to deceive, distort or malign. Conversely, disinformation and malinformation fall into a more exacting category of patent harm as the intent is often defined as seeking to distort, malign or damage. Also, targeted online violence and

---

[191] D Ordway 'Information disorder: the essential glossary' 23 July 2018 *Journalists' Resource* https://journalistsresource.org/studies/society/internet/information-disorder-glossary-fake-news/ (accessed 15 July 2020); JM MacAllister 'The doxing dilemma: Seeking a remedy for the malicious publication of personal information' *85 Fordham Law Review* 2455 https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=5370&context=flr (accessed 15 June 2020); C Jack 'Lexicon of lies: term for problematic information' (2017) *Data & Society* https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf (accessed 20 June 2020).

[192] SC Woolley & PN Howard 'Computational propaganda worldwide: Executive Summary' (2017) *Oxford Internet Institute, University of Oxford* https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf (accessed 1 July 2019).

[193] Woolley & Howard (n 192 above) 4.

[194] Wardle (n 24 above).

[195] As above.

© University of Pretoria

online hate speech as forms of online harms may be categorised as being patently harmful as there are no doubts as to their intent to cause violence or harm to another.

Together with this classification above, lies the need to ascribe liability to various categories of speech in order to be able to find balance for free speech and prohibited speech. According to the report by United Nations' High Commissioner for Human Rights on expert workshops on the prohibition of incitement to national, racial or religious bias, there are three major categories of classifying liabilities for speech:[196]

> expression that constitutes a criminal offence; expression that is not criminally punishable, but may justify a civil suit or administrative sanctions; expression that does not give rise to criminal, civil or administrative sanctions, but still raises concern in terms of tolerance, civility and respect for the rights of others.

In the classification of online harms, it is comfortable for some forms to be categorised under each of the categories of liability described above and in other instances, overlap among the categories. For example, online hate speech, cyberbullying or online gender violence, online terrorist content due to their potential to cause harm against others would strictly fall into the first category of expressions that constitute a criminal offence and therefore illegal. Information disorder except for misinformation may fall into the second category due to their various propensities to cause harm. For example, disinformation and malinformation may require civil sanctions but not criminal sanctions especially for the level of harm they may cause. Misinformation will ideally fall into the last category that raises concerns in terms of tolerance and the rights of others.

However, a closer look at this classification under the report does not give enough attention to online harms like disinformation which have varying degrees of effects, especially when combined with other forms of online harms.[197] For example, hate content spread widely through both primary and secondary methods of engendering disinformation, is not the same as deliberate false information from a Twitter account with three followers for the past six years. The liability that will apply to each of the scenarios should be different. Unless the report does not envisage information disorder as harmful, then, the last category will suffice. Unfortunately, the harm caused by information disorder cannot be understated in its negative impacts on freedom of expression and its overall democratic effects. The Nigeria Stability and Reconciliation Programme (NSRP)[198] provided a more specific and definitional clarity on how hate content can also spread disinformation by stating that hate speech 'can create a

[196] United Nations General Assembly 'Expert workshops on the prohibition of incitement to national, racial or religious hatred, A/HRC/22/17' 11 January 2011 http://undocs.org/en/ A/HRC/22/17 (accessed 1 December 2021) para 12.
[197] D Mumbere 'Fake news fuels xenophobic tensions in South Africa' 6 September 2019 *Africa News* https://www.africanews.com/2019/09/06/fake-news-fuels-xenophobic-tensions-in-south-africa// (accessed 15 July 2020).
[198] NSRP (n 168 above).

© University of Pretoria

vicious cycle as audiences convene around it and by acting as an alternative source of information that neutralises positive information.'

In addition, information disorder cannot be regarded as so harmful that what is true or false, which is always a product of facts and debate, will be determined by the law. As the Joint Declaration of Special Rapporteurs puts it, 'the human right to impart information and ideas is not limited to correct statements.'[199] Nonetheless, this raise concerns on the right to freedom of expression as there is the need to strike a delicate balance between the access to information in a broader sense, freedom to express based on free flow of information and the need to protect the rights of others. Such balance cannot be achieved through criminal sanctions alone that hang over the heads of unpopular speech or political speech.

In the same vein, it is difficult in some instances to limit the redress available against information disorder to civil or administrative claims. In the digital age, information spreads quickly whether or not it is true, therefore the impacts of that information, especially when incorrect, depending on the content and context are assuredly far-reaching and shape opinions. This impact, placed side by side with the need to ensure a democratic society, poses a lot of challenges in its regulation. Adding a more complex variation, the harm envisaged as the limits of the right to freely express oneself, especially as propagated by Mill, is limited to physical harm. As a result, it is not immediately clear whether false information and its possibility of causing psychological harm meets the threshold of 'clear and present danger.'[200]

Therefore, the classification under the report may require a more robust and nuanced view of how these classifications can be practically applied not only for online hate speech but also for other various online harms that it may be combined with to wreak havoc. This also provides an opportunity to address the challenge of regulation. While some forms of online harms are clearly illegal because they are harmful, some, despite their harm, are not illegal.

[199] Organisation for Security and Co-operation in Europe 'Joint declaration on freedom of expression and 'fake news', disinformation and propaganda' 3 March 2017 https://www.osce.org/files/f/documents/6/8/302796.pdf (accessed 24 August 2021).
[200] Section 2.3.2 above.

© University of Pretoria

Table 2: Classification of online harms

Online harms	Information disorder			Targeted online violence					
Forms of online harms	*Misinformation*	*Disinformation*	*Malinformation*	*Cyberharassment*	*Cyberbullying*	*Cyber aggression*	*Online GBV*	*Violence against children online*	*Online hate speech*
Harmful	✗	✓	✓	✓	✓	✓	✓	✓	✓
Illegal	✗	✗	✗	✓	✓	✓	✓	✓	✓

3.6 Impacts of online harms on the right to freedom of expression in Africa

In classifying the major ways information disorder impacts the right to freedom of expression, McKay and Tenove divide them into three, major, 'anti-deliberative properties.'[201] The first is epistemic cynicism which sows the seed of continuous doubt in the electoral process. The second is techno-affective polarisation which creates various online behaviour towards certain identities. The third is pervasive inauthenticity which means that the online information system becomes pervaded with deception and untruths that what is factual, logical and inclusive becomes near-impossible.

In addition to this, two more impacts can be added to McKay and Tenove's. First, depending on the impact deployed by various actors, it has been found that the reason why information disorder affects democracies is that electoral decisions are not well informed with the facts.[202] One of the important reasons the right to freedom of expression is closely tied to democratisation is because an informed citizenry will more than often beget an effective political leadership through access to information, whether for private or public purposes which should yield 'free and fair' electoral choices.[203] According to Sunstein on the impacts of manipulation on choice and applying his thoughts to manipulation in elections, a free election here would mean such a process that is not affected by fear and intimidation and is by one's choice.[204] When information disorder is deployed especially during elections or other political events, it harms the electoral process as the 'wish of the people' which ought to form the basis of the government has been surreptitiously manipulated. Information

[201] See S McKay & C Tenove 'Disinformation as a threat to deliberative democracy' (2020) *Political Research Quarterly* 703-717.

[202] Section 2.4.3 above; JR Hollyer *et al* 'Fake news is bad news for democracy' 5 April 2019 *The Washington Post* https://www.washingtonpost.com/politics/2019/04/05/fake-news-is-bad-news-democracy/ (accessed 15 December 2019).

[203] As above.

[204] See CR Sunstein 'Fifty shades of manipulation' (2016) 1 *Journal of Marketing Behaviour* 213.

© University of Pretoria

disorder as a result impacts the right to freedom of expression which in turn impacts the electoral process that ultimately impacts democracies.[205]

Second, as with most African contexts, governments use information disorder as an excuse to clamp down on the right to freedom of expression.[206] Oftentimes, other forms of online harms, especially those often deployed during elections or major political events like online hate speech, online gender-based violence, etc., are spread through information disorder therefore raising the threshold of physical harm to a victim.

In many African countries, colonial laws established false information or defamation as a crime.[207] Thereafter, these provisions developed into cyberspace policies that treat false information as pariahs, without the inclusiveness of free expression.[208] Despite this misnomer, various international human rights systems, including the African human rights system have constantly reiterated the need to review such laws and have them conform to freedom of expression standards.[209] It is this misconception laid down by colonial criminal codes that now feature in many cybercrime laws across the region, criminalising false information.[210]

It is also these laws that social media companies defer to when considering the local context on information disorder.[211] In addition, where social media companies have applied their policies,[212] especially where they removed pro-government accounts that share information disorder, they have been threatened by the government.[213] This sets off a chain reaction that is dangerous for the protection of the right to freedom of expression. The colonial law influences these laws and these laws are being used to determine who says what and how, thereby unlawfully restricting the right to freedom of expression online.

Despite being the biggest purveyors of false information, governments too began to take advantage of laws which impact the right to freedom of expression and

---

[205] Chesney & Citron (n 12 above).
[206] Adibe *et al* (n 126 above).
[207] Section 2.5 above.
[208] As above.
[209] Section 2.4.3 above.
[210] Section 2.5 above.
[211] As above.
[212] 'Facebook shuts Uganda accounts ahead of vote' 11 January 2020 *Yahoo!* https://news.yahoo.com/facebook-shuts-uganda-accounts-ahead-110022184.html?guccounter=1 (accessed 12 January 2020).
[213] Government of Uganda 'Presidency warns Facebook and Twitter' 12 January 2021 https://twitter.com/govuganda/status/1349060384490213377?s=12 (accessed 12 January 2021); The Observer 'Museveni warns Facebook ahead of elections' 12 January 2021 https://twitter.com/observerug/status/1349059958885785601?s=12 (accessed 12 January 2021).

© University of Pretoria

democratic sustenance altogether.[214] These laws, steep in colonial legal provisions are now used, through cyber laws, to inform regulation of information disorder in Africa. As a result, there is a triad of actors on information disorder – colonial laws, cyber laws and policies and online platforms where harms spread. The colonial laws laid an out-of-touch legal foundation, cyber laws in Africa built on this foundation and as a result, social media platforms' approach to information disorder are based not only on this foundation and are impacted by their moderation policies[215] hence, a challenge for protecting freedom of expression online in Africa.

On cyberbullying and cyberaggression, this impacts the right to freedom of expression by not making the various provisions on the harms sufficiently clear, narrow towards the aim and be proportional. Some of these laws include 'insult', 'annoyance', which are not only vague terms but are capable of being arbitrarily enforced and abused.[216] An important reason to be wary of these provisions in the various laws is that they are being used to silence dissent. Political and unpopular speech, both of which are central to any democratic development is constantly threatened by the whim of the state to hijack what makes for acceptable speech through these provisions.[217]

With respect to online GBV, as more women leave online platforms as a result of violence, the less inclusive the online space is for them.[218] This impacts their right to freedom of expression as they are unable to fully participate in the various political and socio-economic developments of their societies. In some instances, the law that may be used to protect women against online GBV are used against them. For example, various cybercrime legislation with provisions on cyberstalking not only conflate it, they also use it to hunt outspoken women in a bid to silence them. This connects states to the use of punitive laws to silence dissent online.[219]

In addition, there are no gender-specific provisions in many of the laws that exist in most African countries on online GBV.[220] This may be seen from a gender assessment of most colonial laws which lack perspectives on contextual experiences of women and sexual and gender minorities in former colonies.[221]

---

[214] Digital Rights Forensic Lab 'Nigerian government-aligned Twitter network targets #EndSARS protests' 20 November 2020 *Medium* https://medium.com/dfrlab/nigerian-government-aligned-twitter-network-targets-endsars-protests-5bb01a96665c (accessed 21 November 2020).
[215] Ilori (n 188 above); See Y Au *et al* 'Profiting from the pandemic, moderating COVID-19 lockdown protest, scam, and health disinformation websites' (2020) https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/127/2020/12/Profiting-from-the-Pandemic-v8-1.pdf (accessed 5 December 2020).
[216] Section 2.5 above.
[217] Rid (n 54 above).
[218] Iyer *et al* (n 127 above).
[219] Section 2.5 above.
[220] Iyer *et al* (n 127 above).
[221] Section 2.2 above.

© University of Pretoria

On online violence against children in African countries, there are currently no laws save for those on child pornography that sufficiently addresses cyberbullying of children and online violence against children differently. Here, there is also a conflation of terms as cyberbullying of children is not the same as online violence against them. For example, most cyberbullying occurs from peer-to-peer systems while online violence against children may occur both at peer-to-peer systems and with adults. Policies on these different terms are necessary and cannot be based on the problematic colonial laws and equally challenging cyberspace policies.[222] Considering this, online violence against children in Africa not only impacts their right to education and privacy, in many instances where mode of learning has to be digital, they are also exposed to hate speech and violent content online.[223] It also puts them at the cusp of both physical and emotional harm to themselves or others.

The most defining impact on line hate speech has on the right to freedom of expression is that it can be used to precipitate widespread violence and atrocities against others. As a result, such speech, considering various factors, cannot be categorised as an exercise of the right to freedom of expression as the rights of others are involved. Therefore, illegality of hate speech is one of the narrow limitations on the right to freedom of expression.[224] There are various hate speech laws in African countries and many are not specific to offline or online contexts – they are often applied broadly. These laws also have their language borrowed from colonial laws whose provisions are old, overbroad and as a result, unfit for purpose. These are the laws social media platforms are looking to enforce. This creates both a regulatory challenge and violation of the right to freedom of expression as platforms are more prone to remove speech that is legal because the local law provides for it.

From time to time, there have been various extenuating impacts of these online harms like governments shutting access to the Internet based on the prevalence of some of these harms. For example, many African governments who have shut down access have often blamed the spread of both information disorder and targeted online violence as reasons in order to forestall public disorder.[225] These impacts on the right to freedom of expression albeit not directly as it leads to a large-scale censorship by governments based on online harms.

---

[222] Livingstone & Bulger (n 138 above).
[223] S Hinduja & JW Patchin 'Offline consequences of online victimization' 6 (2007) *Journal of School Violence* 107; UNICEF 'Global kids online report' (2019) 71 https://www.unicef-irc.org/publications/pdf/GKO%20LAYOUT%20MAIN%20REPORT.pdf (accessed 15 October 2020).
[224] Section 2.4.1 A iii above.
[225] C Giles & P Mwai 'Africa internet: Where and how are governments blocking it?' 14 January 2021 *BBC News* https://www.bbc.com/news/world-africa-47734843 (accessed 15 January 2020).

© University of Pretoria

3.7 Conclusion

While it is established that online harms are majorly undesirable, it is not immediately certain if their impacts can be fully ascertained, especially in Africa. In reducing such uncertainty on the study of online harms in Africa, this chapter considered the question of what online harms are and their impacts on the right to freedom of expression in Africa. In answering the various sub-questions, for the first and second questions, it examined academic literature and reports to conceptualise online harms, its various forms, methods and classifications. In answering the third question, it considered the interplay of various online harms, digital colonialism and the right to freedom of expression in Africa. It did this by considering available literature on various forms of online harms in Africa and how they impact the right to freedom of expression in the region. In the course of answering the major question for this chapter on online harms and their impacts on the right to freedom of expression in Africa, it made five findings.

First, it found that it is logically plausible not to criminalise information disorder. In instances where such harm may be irreparable, international human rights law allows for the limitation of the right to freedom of speech but under the restrictive three-part test. Therefore, it appears that there is more free speech argument in favour of information disorder due to the philosophy of expression being at the centre of human development as opposed to other forms of online harms which are more harm-prone. In addition to this, it is important to note that when considering the impacts of information disorder and other forms of online harms, it is more remote in impact when compared to the immediacy of targeted online violence and online hate speech.[226] However, current academic and normative approaches to information disorder have yet to settle the place of information disorder that combines with other forms of online harms like the Nigerian and South African examples above.[227] This is because when this combination happens, it raises the threshold of harm from just spreading deception to using that deception to cause grievous violence. For example, the impacts of deliberately misinforming a certain demographic would still depend on the media literacy and quality of access to information in such a community as compared to the instantaneous impacts of targeted online violence and online hate speech which already puts a victim at the cusp of physical or emotional harm.

Second, political and unpopular speech are often likely to be caught up in the web of vague legal provisions on information disorder especially in Africa.[228] This patently poses harm to protecting the right to freely express oneself as it may be easily termed as disinformation or even hate speech. From philosophical free speech ideals to standardised human rights norms, there is an agreement that such views are

[226] Pielemeier (n 66 above).
[227] Mumbere (n 197 above).
[228] Jansen Reventlow (n 153 above).

© University of Pretoria

necessary for democratic development. This also gives many governments excuses to shut access to Internet services in most countries, thereby leading to various socio-political and socio-economic issues.[229]

Third, the lumping of all forms of harms as illegal also contributes to a further complexity in the application of international law in national contexts in a region like Africa. For example, criminal codes and cybercrime legislation still criminalise publication of false news and other permissible aspects of speech under international law.

Fourth, given the conceptual challenges on regulation of online harms, neither states nor private sector can effectively regulate online harms in Africa. This is because due to the grip of digital colonialism on free speech in Africa, old problematic laws are being transplanted into laws used to regulate the online space which further leads social media platforms astray. The future of effectively regulating online harms is far from being limited to just traditional approaches. Also, given that most platforms have still not demonstrated the core values that their policies are based on, private sector-regulation lacks contexts and basis for such wide applicability. The impact of the traditional approach also combined with the vague application of human rights standards of these platforms further makes such regulation far from reach. As a result, these challenges require that regulation of online harms become less binary and more inclusive in its approach beyond the state and private sector dichotomy.

Fifth, the nature of information disorder as being harmful but not illegal moves the responsibility of regulation away from the traditional regulatory approach often exercised by states under international law.[230] It means together with other forms of online harms, information disorder requires a special type of regime that is collaborative, granular and democratic in resolving the challenge of online harms such as ensuring a rights-respecting platform governance.[231]

In conclusion, circling back to the previous chapter on how digital colonialism violates the right to freedom of expression in the digital age in Africa, Belli and Zingales' description of impacts of online harms above is important as one of such impacts include limitation of speech. As a result of problematic foundational colonial laws that impact legislative policies offline and online in many African countries on free speech, the right to freedom of expression in the digital age is at risk. This transplantation of excessive limitations on the right to freedom of expression in most African countries is

[229] See DM Nyokabi *et al* 'The right to development and internet shutdowns: Assessing the role of information and communications technology in democratic development in Africa' (2019) 3 *Global Campus Human Rights Journal* 147 172.
[230] Donahoe & Hampson (n 1 above).
[231] As above.

© University of Pretoria

what informs the traditional regulatory approach of many African governments on online harms which platforms kowtow to.

When problematic colonial legal legacies morph into cyberspace policies to become what social media companies' policies enforce, they exacerbate online harms. This points to digital colonialism as one of the causes of online harms in Africa. In answering the first and second sub-questions above, defining the concept of online harm, its various forms, methods and classifications, has shown there is need for more specific application of these forms based on their meanings in legal reforms in many African contexts. With respect to the third question on the impact of online harms on the right to freedom of expression online in African countries, this chapter considered the various ways online harms hinder the right to freedom of expression online in Africa which ranges from manipulating electoral processes to actual physical offline violence on vulnerable groups like children, women, sexual and gender minorities, migrants and others. The next chapter deals with how these impacts can be prevented through a rights-respecting approach while also protecting online expression.

© University of Pretoria

CHAPTER FOUR: PLATFORM GOVERNANCE, THE PREVENTION OF ONLINE HARMS AND PROTECTION OF ONLINE EXPRESSION IN AFRICA

4.1 Introduction

Chapter three of this thesis focused on what online harms are, their various forms and their impacts on the right to freedom of expression online in Africa. It pointed out that the interplay of problematic colonial laws, cyber laws, and social media platforms' regulatory practices play a role in the African contexts which further engender online harms. In order to effectively prevent these harms in Africa, the third chapter referred to a rights-respecting approach to platform governance which would require a dynamic framing. With a history primarily tied to the Internet, platforms are often digital boundaries that mediate interactions between several actors. Therefore, platform governance arose as a result of the need for a body of rules to guide such interactions.[1]

While there are various forms of platforms that neatly cover their functions based on a set of features like purpose, design, and application, this chapter focuses on social media platforms also known as social media sites.[2] Due to the nature of platforms as allowing direct interactions between many actors,[3] speech-related online harms may be said to occur more often on social media platforms.[4]

As a result, not only are speech-related online harms that bear on the right to freedom of expression online understudied generally but there is also limited research with respect to their regulation on social media. Additionally, there is a concern of how research on online harms, their various forms, impacts and their regulation are almost non-existent in African contexts.[5] It has then become more important to consider how online harms should be regulated having considered their causes and impacts in the previous chapter and especially within a narrower context – social media platforms, separately from the general Internet ecosystem, and what opportunities are there to reduce speech-related online harms and prevent them especially within the African context.

[1] R Gorwa 'What is platform governance?' (2019) 22 *Information, Communication & Society* 2; N Suzor 'A constitutional moment: How we might reimagine platform governance' (2020) 36 *Computer Law & Security Review* 105381, 2; According to Suzor, '… because online intermediaries play such a crucial role in regulating how users behave, we should find a way to ensure that their decisions are legitimately made'; B Haggart & CI Keller 'Democratic legitimacy in global platform governance' (2021) 45 *Telecommunications Policy* 102152, 3; N Suzor *Lawless: The secret rules that govern our digital lives* (2019) 97.

[2] T Gillespie *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media* (2018) 18.

[3] See A Helmond 'The Platformization of the web: Making web data platform ready' (2015) 1 *Social Media + Society* 1-11.

[4] L DeNardis *The global war for Internet governance* (2014) 154.

[5] D Kaye *Speech police: The global struggle to govern the Internet* (2019) 94.

© University of Pretoria

Under international human rights law, governments have the primary responsibility of protecting online speech.[6] However, given the nature of the Internet as global and dynamic, such traditional responsibility is now shared with private actors including social media platforms. In the last two decades, considering the constant rise of platform power across the world, social media platforms now wield enormous resources that do not only impact the right to freedom of expression online specifically, but also have offline impacts on democratic development in general.[7] Like states, social media platforms now carry out vast and far-reaching self-imposed regulatory responsibilities that are developed and applied to govern their digital boundaries.[8] These rules and how they should apply across the world, especially with respect to balancing the protection of the right to freedom of expression online with the prevention of online harms, has been the biggest challenge of governing platforms over the past few decades.[9]

This challenge is even more protracted as most social media platforms are typically torn between a triad of influential actors, which includes themselves through their moderation practices; the users on their platforms; and governments through 'command and control, intermediary liability, and extra-legal influence,' as Land puts it.[10] These actors can be broadly described as state and non-state actors who all through their actions or inactions influence the regulation of online speech which in turn could prevent or exacerbate online harms.[11] Therefore, the goal of effective platform governance, given its current challenges should be such that holds two valid but seemingly opposing facts together: online harms can only be generatively prevented[12] and online speech governance must be underpinned by international human rights standards.[13] This is because it is in critically considering these facts that an effective platform governance can be designed – to resolve the obvious tension between the two facts that online harms can be minimised while also protecting online expression.

As a result, this chapter focuses on platform governance and its possibilities within the African context in preventing online harms and protecting online expression. It does this by answering the third sub-question of this thesis: In what way can a rights-respecting platform governance be used to prevent online harms and protect online expression? It addresses the question by considering sub-questions such as:

---

[6] M Land 'Against privatized censorship: Proposals for responsible delegation' (2019) 60 *Virginia Journal of International Law* 70, 391.
[7] M Moore *Democracy hacked: Political turmoil and information warfare in the digital age* (2018) 217, 221, 272.
[8] R Gorwa 'What is platform governance?' (2019) 22 *Information, Communication & Society* 8.
[9] DeNardis (n 4 above) 157.
[10] Land (n 6 above) 399; N Helberger *et al* 'Governing online platforms: From contested to cooperative responsibility' (2018) 34 *The Information Society* 3-5.
[11] R Gorwa 'The platform governance triangle: Conceptualising the informal regulation of online content' (2019) 8 *Internet Policy Review* 1-15.
[12] Suzor (n 1 above) 3.
[13] EM Aswad 'The future of freedom of expression online' 17 *Technology Review* 45; Kaye (n 5 above) 88; Suzor (n 1 above) 168,181.

© University of Pretoria

a. How do various stakeholders, especially state and non-state actors understand platform governance?
b. What is a human rights perspective to preventing online harms?
c. In what way can these stakeholders perform their responsibilities in order to prevent online harms and protect online expression through platform governance?

In resolving these queries, this chapter examines how to prevent online harms and promote online expression through a rights-respecting approach to platform governance. In order to examine the role of platform governance especially with respect to preventing online harms which have been established as violating the right to freedom of expression online within the African context, it would be important to examine the human rights perspective to platform governance. The reason is because as a human right, the right to freedom of expression online in its capacity is both a standalone right and the right that enables other rights.[14] In both capacities and given the increasing roles of both state and non-state actors with respect to the right, especially in the digital age, it has become necessary to understand how its regulation is carried out.[15] More specifically, in examining the roles of both actors in preventing online harms in Africa, this chapter is divided into seven sections.

The first section provides a background to the chapter while the second section considers the relationship between state and non-state actors by focusing on the background, approaches and forms of platform governance to draw up a typology of platform governance. The third section examines the various limitations of platform governance while the fourth section analyses a human rights approach to governance of online speech while the fifth section considers use of various forms of platform governance in preventing online harms. The sixth section attempts a proposal of a governance framework that is rights-respecting and not afraid to fail in reining in the new governors[16] as a result of their newly amassed far-reaching powers while also taming the old ones. The seventh part concludes that these harms can be prevented and online expression can be protected if international human rights law is applied but through a generative approach.

4.2 Platform governance and its various aspects

The commercialisation phase of the Internet marked the opening up of the Internet as a global resource which started from the early 1990s till date.[17] Even with this, the roll-

[14] Both the preambular texts and Principle 1 of the revised African Declaration refers to the right both as a right in itself and that used to realise others.
[15] Suzor (n 1 above) 97.
[16] K Klonick 'The new governors: The people, rules, and processes governing online speech' 131 *Harvard Law Review* 1670.
[17] The word 'Internet' is from the word 'Internetworking' as it was designed to function between differently configured machines and networks. See M Mueller *Ruling the root: Internet governance and the taming of the Cyberspace* (2004) 244. United States' funded computer networks It is important to

© University of Pretoria

out of platforms like Facebook, Amazon, Twitter and others as we know them today did not start till the early 2000s.[18] This was what led to the development of the World Wide Web and web software to run with it.[19] It was in the early 2000s that the innovations around Web 2.0 formed and created a distinguishing feature of 'user-generated content' and 'peer production' fuelled by its prospects to revolutionise e-commerce.[20] The initial phase of the web led to the design of the Web 2.0 which was made possible by common standards that could be enjoyed by today's 'big tech.'[21] Therefore, even though the Internet through its experimentation or closed phase set the future of platforms in motion, it was only fully realised during its commercialisation phase. In understanding the relationship between Internet governance and social media platforms, DeNardis describes three distinct areas of inquiry: normative regulation of social media platforms for different reasons, user deployment of social media and social media as privatised governance.[22]

4.2.1 Internet governance and platforms

Before considering this relationship, it is important to consider the various conceptualisations of Internet governance for two reasons: to reconsider platform governance as a specialised form of Internet governance and highlight the need for inventive interventions.[23] There have been various developments which have seen to the uptake of Internet governance conversations which can be traced to the various efforts from the World Summit on the Information Society (WSIS) to the Internet

note the theoretical disagreement between the ARPA and a consortium of four universities, Stanford Research Institute, University of Utah, University of California, Santa Barbara (USCB) and University of California, Los Angeles (UCLA) that worked on the Internet in its early stages. There is no consensus between the ARPA and the universities on why the Internet was created as there was a tussle between whether it was a military solution or a far-reaching and impactful research output. See W Kleinwächter 'Internet co-governance: Towards a multilayer multiplayer mechanism of consultation, coordination and cooperation (M3C3)' (2006) 3 *E-Learning and Digital Media* 473; W Kleinwächter 'History of Internet governance and challenges of tomorrow' (2016) http://3.15.112.233/wp-content/uploads/2018/01/History-of-Internet-Governance-and-Challenges-of-Tomorrow-Wolfgang-Kleinwächter-9-August-2016.pdf (accessed 12 February 2021); AE Marwick 'Are there limits to online free speech?' *Medium* 5 January 2017) https://points.datasociety.net/are-there-limits-to-online-free-speech-14dbb7069aec (accessed 3 February 2021).

[18] C Fuchs *Social media: A critical introduction* (2014) 29 48; See also J Goldsmith & T Wu *Who controls the Internet?: Illusions of a borderless world* (2006) 23.

[19] As above.

[20] T O'Reilly 'What is Web 2.0' (2005) www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1 accessed 12 February 2021; T O'Reilly & J Battelle 'Web squared: Web 2.0 five years on' (2009) http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper (accessed 15 February 2021).

[21] Fuchs (n 18 above).

[22] See L DeNardis 'Internet governance by social media platforms' 39 *Telecommunications Policy* 10.

[23] Haggart & Keller (n 1) 3. Haggart & Keller noted that 'as platform governance emerged as a key area of internet governance, internet governance scholarship has moved from a 'code is law' focus to further exploring the different aspects of intermediaries' private ordering.' Suzor, *Lawless* (n 1) 103, 165.

© University of Pretoria

Governance Forum (IGF) at the UN. The UN described Internet governance as a stakeholder-sourced set of rules and systems.[24]

According to DeNardis, Internet governance is about governance, not governments and is an exercise in bricolage i.e. building from many diverse aspects.[25] This forms an in-road into the relationship between the Internet, its governance and platform governance in general. It suggests that major stakeholders like state actors which include governments and non-state actors like private tech companies, Internet Service Providers (ISPs), civil society and others all play both specific and interrelated roles for the governance of the Internet, even though they are carried out by non-state actors.[26]

To Mueller, Internet governance is 'the simplest, most direct, and inclusive label for the ongoing set of disputes and deliberations over how the Internet is coordinated, managed, and shaped to reflect policies.'[27] Jasanoff has a more direct description of Internet governance as the relationship between the Internet and power, she states that technology 'embeds and is embedded in social practices, identities, norms, conventions, discourses, instruments, and institutions – in short, in all the building blocks of what we term the social.'[28] According to Dutton, also on the interdisciplinary nature of Internet governance studies, he stated that it draws on many disciplines.[29]

Given these definitions, the primary objective of Internet governance is the design, policy and administration of technologies that make the Internet as an infrastructure operational. With a more technical definition that focuses more on the governance of the Internet as a global infrastructure, Mueller *et al* further defines Internet governance as a global activity comprising many actors.[30]

In arriving at this definition, Mueller *et al* compartmentalised their approach to Internet governance divided broadly into three: technical standardisation; resource allocation and assignment; and human conduct which involves public policy setting. Technical standardisation involves reaching agreement about networking protocols, data

---

[24] C Bossey 'Report of the United Nations Working Group on Internet Governance' 2005 https://www.wgig.org/docs/WGIGREPORT.pdf (accessed 14 February 2021).
[25] DeNardis (n 4 above) 11.
[26] DeNardis (n 4 above) 11-15.
[27] M Mueller *Networks and states: The global politics of Internet governance* (2010) 9.
[28] S Jasanoff, 'The idiom of co-production' in S Jasanoff (ed) *States of knowledge: The co-production of science and social order* (2004) 3.
[29] W Dutton *Oxford handbook of Internet studies* (2013) 1.
[30] M Mueller *et al* 'The Internet and global governance: Principles and norms for a new regime' (2007) 13 *Global governance: A review of multilateralism and international organizations* 237.The broad term 'Internet governance' is often used to describe the design and administration of the technical infrastructure necessary to keep the Internet operational and the enactment of substantive policies around these technologies. Another prominent theme is the role of nation states and intergovernmental organizations in regulating or coordinating the Internet in areas as diverse as antitrust, net neutrality, computer fraud and abuse, privacy, or hate speech. (See Internet governance by social media platforms by DeNardis).

© University of Pretoria

formats and their documentation.[31] Resource allocation and assignment involve Internet identifiers such as Internet Protocol (IP) addresses and protocol numbers. The third part on human conduct focuses on defining and enforcing regulations, laws and policies. The conceptualisation by Mueller and others suggests that beyond the technicalities of the Internet infrastructure, there exists the coordination of regulations, laws and policies which defines relationships and improves trust on the Internet.

However, DeNardis offers a more comprehensive and elaborate definition of Internet governance that is divided into four distinct parameters namely, the study of Internet governance as distinct from content and usage; the Internet as a unique architecture; distributed governance; and Internet freedoms.[32] Given the largely technical aspects of the first two parameters, and because this chapter focuses more on the policy conversations on the Internet and its governance, it discusses the last two parameters more as an entry point into the relationship between Internet governance and platform governance.

For distributed governance, DeNardis divides global Internet governance systems into five, namely 'technical design decisions; private corporate policies; global institutions; national laws and policies; and international treaties.'[33] As would be discussed in the latter parts of this chapter, it is shown that these routes constitute the major aspects of platform governance, even though the efficiency of such workings depends more on other factors depending on the context. In the context of this research, one of such factors is digital colonialism, which has been explained in the previous chapters as the unique relationship between colonial laws, cyber policies and social media governance in Africa.

The last aspect which focuses on Internet freedoms is also of importance in that the impact of digital colonialism which is felt in unique ways through online harms in African countries impacts the right to freedom of expression online.[34] This is particularly necessary in that Internet governance as the main governance model envisages the problematic aspects of protecting Internet freedoms like online harms which this chapter places within the scope of platform regulation. Therefore, this chapter does not only focus on the various actors within the Internet governance or platform governance space, it also considers ways of preventing online harms as an offshoot of protecting Internet freedom. These two compartmentalisations, distributed governance and Internet freedoms offer a more critical perspective to considering the relationship between Internet governance and platform governance, especially with respect to online harms.

Referring to DeNardis' distinct areas of enquiries above, there are various reasons for the normative regulation of platforms. For example, international human rights law

[31] Mueller *et al* (n 30 above) 245-250.
[32] DeNardis (n 4 above) 19.
[33] DeNardis (n 4 above) 22.
[34] DeNardis (n 4 above) 24.

© University of Pretoria

regulates through various mechanisms in order to guarantee a human rights approach to social media governance while national governments seek to carry out regulations through laws to guarantee their various national interests. Platforms also regulate themselves in order to fulfil their corporate responsibilities and make more profits.[35] Another area of enquiry is that of users who have been described by Papacharissi as 'affective publics' with how they use social media platforms to rally for protection of human rights both online and offline.[36] Lastly, the aspects of privatised governance which is an offshoot of some of the motivations for normative regulation. Here, social media companies drive their own rules and enforce them among those who use their platforms. All of these distinct areas point to Internet governance as having a distinct relationship with how the governance of platforms is done today by linking the rules, the users and platforms. As DeNardis argued further that internet governance is a complex web of interests and actors.[37]

In essence, social media platform governance finds root in the existing Internet governance systems which were originally designed to design a global architecture through consensus.[38] Identifying the essence of the relationship between the Internet and what social media platforms do today with respect to online speech, Lessig notes that the real protection of online speech is the way the internet is built, a distributed architecture that depends on many nodes in its function.[39]

Therefore, platforms and their governance as a distinct area of inquiry requires a more nuanced approach even though there are some similarities with the wider Internet governance systems. For social media platform governance, the reach is global and so are the harms. However, national governments are also making their own rules with little or no coordination with the basic Internet governance systems let alone platforms' governance. Therefore, this makes regulating platforms more complex due to the competing interests of not only governments and social media platforms but also the international human rights systems that have the mandate to promote and protect human rights across the globe. It then becomes important to understand the various aspects of platform governance in order to further appreciate the extent of complexities involved. This understanding is necessary in order to rethink possible interventions that do not only ensure that platform governance is underpinned by human rights but also envision an effective, dynamic and generative governance system that minimise the impacts of online harms in Africa.

4.2.2 Understanding the platforms in platform governance

In understanding the concept of platform governance, it is important to state the context within which it would be used as platforms overlap not only in meaning but in

[35] Klonick (n 16 above) 1625-1627.
[36] Z Papacharissi *Affective publics: Sentiment, technology, and politics* (2015) 125.
[37] DeNardis (n 4) 222.
[38] See Suzor (n 1 above).
[39] L Lessig *Code and other laws of cyberspace* (2006) 236.

© University of Pretoria

purpose, design and application.[40] After this is established, it becomes necessary to understand platforms and their governance majorly as a departure from traditional governance that is state-centric where governments, as formal agents of the state, administer control with respect to human rights. This disaggregation of meaning is necessary to understand the motivations for the term platform governance as a complex but unique system. It is this disaggregation that this chapter focuses on in drawing on the various actors, their roles and impacts, and how these could help in reducing and eventually preventing online harms.

In defining platform governance, it is important to consider what is being governed, to what extent and why. Pointing out the centrality of platforms today in our daily lives, West, referring to platforms, notes that they are also economic systems and distributors.[41] In addition to this, DeNardis described them as Internet intermediaries 'that mediate between digital content and the humans who contribute and access this content.'[42] She gave examples of these Internet intermediaries as search engines, blogging platforms, content aggregation sites, reputation engines, financial intermediaries, transactional intermediaries, trust intermediaries, application intermediaries, locational intermediaries, advertising intermediaries and social media platforms. According to Gillespie, these companies often refer to themselves as 'platforms' which is a discursive term 'specific enough to mean something and vague enough to work across multiple venues for multiple audiences.' He further defines them as online services which are conduits for curation and sharing of users' content that is built on an infrastructure for profit.[43]

It is important to note that despite this categorisation, some platforms like Google combine the design, purpose and application of these various forms of platforms. According to Kaplan and Haenlein, a social media platform is 'a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content.'[44]

---

[40] Gillespie (n 2 above).

[41] SM West 'Thinking beyond content in the debate about moderation' (2020) 9 *Internet Policy Review: Journal of Internet Regulation* 15.

[42] DeNardis (n 4 above) 154.

[43] Gillespie groups platforms in two major categories which are social media platforms, recommendations and rating sites as social network sites. Examples include Facebook, LinkedIn, Google+, Hi5, Ning, NextDoor, and Foursquare; blogging and microblogging providers like Twitter, Tumblr, Blogger, Wordpress, and Livejournal; photo- and image-sharing sites like Instagram, Flickr, Pinterest, Photobucket, DeviantArt, and Snapchat; video-sharing sites like YouTube, Vimeo, and Dailymotion; discussion, opinion, and gossip tools like Reddit, Digg, Secret, and Whisper; dating and hookup apps like OK Cupid, Tinder, and Grindr; collaborative knowledge tools like Wikipedia, Ask, and Quora; app stores like iTunes and Google Play; live broadcasting apps like Facebook Live and Periscope. The second category comprise of recommendation and rating sites like Yelp and TripAdvisor; exchange platforms that help share goods, services, funds, or labour, like Etsy, Kickstarter, Craigslist, Airbnb, and Uber; video game worlds like League of Legends, Second Life, and Minecraft; search engines like Google, Bing, and Yahoo. See Gillespie (n 2 above).

[44] AM Kaplan & M Haenlin 'Users of the world, unite! The challenges and opportunities of Social Media' 53 *Business Horizon* (2009) 61.

© University of Pretoria

On the role that platforms play in what makes them platforms, that is content, DeNardis argues further that these platforms are the primary determinants of what stays up or goes down on their platforms.[45] To Hogan and Quan-Haase, this role is also seen as the ability to interactively exchange information with dispersed groups of recipients.[46] Within this context, platforms can also be referred to as social media and its governance. Zarsky presents at least four forms of social media or platform governance. They are by code, by contract, by law and by social norms. He stated that these forms should be 're-examined' and 'recalibrated' suggesting that while the ideas that these forms possess can only be activated based on context.[47]

Given these various perspectives, the Internet gave rise to platforms while Internet governance gave rise to various forms of governance including those of platforms. Platform governance further beamed more focus on how regulation though economic, labour and content forms the new reality of online experience today as we see platforms not only as social media platforms providing several services including core business, social impact apps, public services, social media platforms with their features as bounded, direct in interaction, comprising various actors and coordinating content without producing any. Specifically, social media platforms also have their various forms including streaming and video-based social media platforms all of which bear on the right to freedom of expression protected under international human rights law. Therefore, social media platforms are for-profit digital conduits that facilitate the sharing of user content.

4.2.3 The governance of platforms

In defining platform governance, there are at least two approaches. One, there is platform governance as content moderation or governance which considers platform governance as the ability to screen content and make decisions as to their allowance or otherwise on a platform. It is perceived as the end to the means – the means being platform governance, what it seeks to achieve. This can also be described as governance by platforms.[48] Roberts puts it more simply that it is 'the organized practice of screening user-generated content posted to Internet sites, social media, and other online outlets.'[49]

According to Gorwa on the other hand, he described platform governance as content

[45] DeNardis (n 22 above).
[46] B Hogan & A Quan-Haase 'Persistence and change in social media' 30 *Bulletin of Science, Technology and Society* (2010) 76.
[47] T Zarsky 'Social justice, social norms and the governance of social media' (2015) 35 *Pace Law Review* 171.
[48] T Gillespie 'Regulation of and by platforms' in J Burgess, T Poell, and A Marwick (eds) *SAGE handbook of social media* (2017) 12-22.
[49] ST Roberts *Behind the screen: Content moderation in the shadows of social media* (2019) 33.

© University of Pretoria

moderation.[50] Gillespie provides an additional explanation to moderation as governance of platforms by noting that they determine the propriety of content based on their rules and also design enforcement tools for these rules on their platforms.[51] Roberts' definition serves a more technical purpose for content moderation as governance in that it focuses more on the purposive meaning of content moderation as a series of activity carried out by the most proximate actor – the social media platforms. Close to Roberts' definition is Gillespie's which expressly refers to social media platforms and other information intermediaries as being at the centre of content governance. On the other hand, Gorwa offers a more elaborate definition by considering the larger processes and ecosystem involved with content moderation. For example, to Gorwa, it is how a social network governs the activity of its many actors which suggests such governance that is by a social media platform and other stakeholders like governments, civil society, users etc. However, according to Grimmelmann, content moderation serves a more specific purpose to mitigate abuse in that they are such 'governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse.'[52]

Content moderation as platform governance is an important pressure point that requires more focus. As such, it is the final output of processes of governance under discussion. This is because, whether governments, users or even civil society clamour for various forms of governing platforms, oftentimes, it is through content moderation, the real and practical governance of online speech, that social media platforms wield the most power. This power is not limited to being able to determine who stays on a platform or should go, it includes the powers to determine how to stay, which is best seen in the optimisation of different tools of new technologies to pre-determine online behaviours.[53] Therefore, in the true sense, platforms do not only hold the carrot and

---

[50] R Gorwa 'The shifting definition of platform governance' 23 October 2019 *Centre for International Governance Innovation* https://www.cigionline.org/articles/shifting-definition-platform-governance (accessed 21 March 2021).

[51] T Gillespie 'Introduction: Expanding the debate about content moderation: scholarly research agendas for the coming policy debates (2020) 9 *Internet Policy Review: Journal of Internet Regulation* 2.

[52] See J Grimmelmann 'The virtues of moderation' (2015) 17 *Yale Journal of Law & Technology* 47.

[53] D Coleman, 'Digital colonialism: The 21st century scramble for Africa through the extraction and control of user data and the limitations of data protection laws' (2019) 24 *Michigan Journal of Race & L*aw 439. She described such pre-determination from a digital colonialism perspective and argued that it 'is just as oppressive as the early colonialism from the nineteenth century. Large tech companies, typically owned and primarily operated by White men, are extracting data from uninformed users and controlling that data to profit via predictive analytics.' Ogunleye defined predictive analytics as the extraction and transformation of data to improve processes. See J Ogunleye 'The concepts of predictive analytics' (2014) 2 *International Journal of Knowledge, Innovation and Entrepreneurship* 83. The impacts of cybercrime laws in most African countries that mirror negative colonial legacies on the right to freedom of expression online like the use of vague provisions and laws are telling of the predictive nature of coloniality in African countries. The processes bettered by such predictive nature of colonial legacies belong to Western systems and perpetuated both through its laws and now including its large tech companies. Colonial laws on free speech in African countries are also predictive analytics in that they are reproducing their intended effects through violations of free speech online even many decades

© University of Pretoria

the stick – the politics of today's information power seems to be largely influenced by them as a result of content moderation.

Platforms moderate majorly in four ways: automatic or manual moderation; transparent or secret moderation; *ex-ante* or *ex-post*; and centrally or distributedly. First, moderation through automation is basically the use of machines pre-trained on a set of commands to identify a problematic online speech and make various decisions which range from leaving it in its original form, reducing its reach to other users, limiting affordances on such a platform or removing such content entirely.[54] Such automated systems include the use of software filters to identify a set of online content that the platform deems problematic.[55] Manual moderation is such where identification and decision on problematic content is made by humans and not machines. In some instances, both the automated and manual methods are combined, which is referred to as hybrid model of moderation[56] in that the automated systems flag content and humans carry out the final decisions. Grimmelmann argues that the choice for either moderation practices is often as a result of what it costs. For example, human labour in moderation is costlier compared to automated systems that carry out such tasks with even far more efficiency but lack the insights of contexts that human interventions have.[57]

Second, moderation does not end when content is decided on. What was decided, why and how also contribute to the moderation as governance.[58] For example, as Grimmelmann puts it, oftentimes, what was decided is often obvious and requires no additional disclosures but how platforms arrive at such a decision is perhaps the biggest area of contention for platform governance. He also pointed out that depending on the mode of moderation, complex automated systems mean less transparency as to why and how platforms make such decisions. In contrast, human moderation is easier to explain. Therefore, the extent of transparency or secrecy often depends on

---

after. It is also important to note that this thesis uses Western systems in the context of Global North systems including the United States, United Kingdom, France, Germany and others.

[54] Grimmelmann (n 52 above) 63.

[55] Grimmelmann (n 52 above) 64; D McCullagh 'Google's chastity belt too tight' *CNET NEWS* 23 April 2004 http://news.cnet.com/2100-1032_3-5198125.html (accessed 12 February 2021). It describes the 'Scunthorpe problem' of overzealous software filters that find false positives of prohibited terms embedded in innocent phrases; T Ilori 'Facebook's censorship of the #EndSARS protests shows the price of its content moderation errors' *Slate* https://slate.com/technology/2020/10/facebook-instagram-endsars-protests-nigeria.html (accessed 13 March 2021); M Papenfuss 'Twitter agrees to block Tweets critical of India government's COVID-19 response *HuffPost* https://www.huffpost.com/entry/india-twitter-covid-surge-cremations_n_6084cfa3e4b02e74d21a6ef2 (accessed 2 February 2021).

[56] S Singh 'Everything in moderation: An analysis of how Internet platforms are using artificial intelligence to moderate user- generated content' 22 July 2019 *New America* https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/ (accessed 15 March 2021).

[57] Grimmelmann (n 52 above) 65.

[58] J Donovan 'Navigating the tech stack: When, where and how should we moderate content' in *Models for Platform Governance CIGI Essay Series* 2019 https://www.cigionline.org/sites/default/files/documents/Platform-gov-WEB_VERSION.pdf (accessed 20 February 2021).

© University of Pretoria

which of the primary decision-makers, software filters or humans, carried out such moderation.

Third, when moderation occurs also matters in understanding the governance of online speech. For example, where a content is restricted before they are shared such moderation is said to be *ex-ante*. On the other hand, where such speech has been posted on the platform and moderation occurs, it is referred to as *ex-post*. The major difference between both modes is that one occurs before, the other after a content has been shared. The major challenge with *ex-ante* mode is that the policing of content is already done before a user already shares such, therefore a person is moderated not on what has been shared, but what is yet to be shared. This does not only police online speech, it polices thoughts and opinions even before they are fully formed. The problem with *ex-post* moderation, which is often the most popular, is also tied to the various issues associated with automated or human content moderation.

The fourth mode through which content moderation mirrors governance is the centrality or distributed nature of decision-making. For example, the decision as to whether content should stay up on a platform, or determinations of what such staying means, could be carried out just by a software filter, a human moderator or a government law that platforms claim to defer to. Such discretion is said to be centralised. However, where there are various actors involved in the decision-making process, for example, a user reports, a human moderator carries out the review with other decision-making tiers both internally and externally before making a decision, such moderation is said to be distributed. Centrality of decisions definitely throws up the concentration of powers in a single entity[59] while distributed moderation does not only take time but also involves a number of challenges including logistics, determining power-sharing models on who makes what decision and legitimacy.

In addition to platform governance as content moderation as an approach, it is also considered as a body of rules. It may also be described as governance of platforms.[60] Gorwa describes platform governance as a body of rules which means different things to different stakeholders. For example, a platform such as Facebook may consider platform governance as a complex system of regulatory interests across the world.[61] According to Klonick, the term platform governance 'does not fit neatly into any existing governance model, but it does have features of existing governance models that support its categorization as governance.' [62]

She also regards platform governance within the context of administrative law as that 'which has long implicated the motivations and systems created by private actors to

[59] D Morar & C Riley 'A guide for conceptualising the debate over Section 230' 9 April 2021 *Brookings* https://www.brookings.edu/techstream/a-guide-for-conceptualizing-the-debate-over-section-230/ (accessed 15 April 2021).
[60] Gillespie (n 48 above) 2-11.
[61] Gorwa (n 1 above) 4.
[62] Klonick (n 16 above) 1662, 1663.

© University of Pretoria

self-regulate in ways that reflect the norms of a community.'[63] Gorwa also explained further that the complexity of platform governance needs to be understood as that which involves 'the multitude rather than the few.'[64] These conceptualisations of platform governance demonstrate in the closest way, the complexities that make up the platform governance ecosystem such that finding a definition for it is as tedious as pinning down its effectiveness. Murray puts this in a more animated way when he described it as where all actors regulate simultaneously and concurrently.[65]

These various approaches should not faze ideas about platform governance as the goal is not to immediately arrive at a definition or to determine the effectiveness of such governance. The goal is to generate iterative developments where failure among actors is an opportunity to rethink approaches. Actors under platform governance can be primarily divided into two. They are state and non-state actors. State actors can be further divided. There are traditional state actors, who are made up of governments, their institutions and their traditional tools of governance. There are also quasi-state actors,[66] who are made up of international and supranational organisations. Non-state actors are proximate actors involved in platform governance that do not wield the traditional powers of state actors but are considered actors because of their role in the platform governance ecosystem. They are dominantly social media platforms, civil society and users, all of who play important roles in the governance of online speech.

4.2.4 Forms of platform governance

Given this background on approaches to platform governance and within the context of this research, it may be briefly described as a system that manages platforms. It can be used interchangeably with platform regulation and in this context, such platforms are social media platforms. This is as a result of the various roles of multi-actors who perform regulatory functions, demand more accountability through a body of rules and as a result govern platforms which includes platforms themselves. In governing platforms, whether directly or indirectly, various actors involved do so through content moderation. A broader definition may then be defined as a system

[63] Klonick (n 16 above) 1663.
[64] Gorwa (n 1 above) 12.
[65] A Murray *The regulation of cyberspace: Control in the online environment* (2007) 234.
[66] Centre for Human Rights 'Democracy, Transparency and Digital Rights Unit participates in Francophone consultation on freedom of expression and access to information' 22 October 2019 https://www.chr.up.ac.za/expression-information-and-digital-rights-news/1872-democracy-transparency-and-digital-rights-unit-participates-in-francophone-consultation-on-freedom-of-expression-and-access-to-information (accessed 15 February 2021); Centre for Human Rights 'Centre for Human Rights participates in the Southern African sub-regional consultation on the revision of the draft Declaration of Principles on Freedom of Expression and Access to Information in Africa' 3 October 2020 https://www.chr.up.ac.za/expression-information-and-digital-rights-news/1855-centre-for-human-rights-participates-in-the-southern-african-sub-regional-consultation-on-the-revision-of-the-draft-declaration-of-principles-on-freedom-of-expression-and-access-to-information-in-africa (accessed 15 February 2021).

© University of Pretoria

where state and non-state actors[67] apply various rules in order to protect particular interests with respect to online speech. These interests range from profit-making, public policy, human rights protection and many more. In understanding the interests of these various actors, it is important to consider how they make up the ecosystem of platform governance. There are primarily four forms of platform governance and they are traditional, sectoral, self-governance and multistakeholder governance.

A Traditional form of platform governance

Traditional platform governance or the traditional approach to platform governance is where states or governments are the primary actors with respect to drawing up the body of rules that regulate or govern platforms. It is state-driven, recommends platform accountability and backed by laws. For example, governments pass laws that require a platform to remove content it deems as hate speech within 24 hours.[68] According to the various aspects of platform governance described above and based on the definition given above, the government uses its powers to make policy on behalf of its citizens. The government is a state actor, the law is a form of rule, the interest being protected, according to governments, is that of the public to prevent the amplification of hate speech and ensure public order. The major examples of traditional platform governance is where governments, usually the executive or legislature, enact a specific law with respect to one or more aspects of regulating online content. Most jurisdictions outside the United States, including Europe, Asia and Africa require that platforms perform specific responsibilities with respect to online content failure of which is attached one or more legal sanctions.[69]

Due to the history of the Internet and platforms having strong ties with the United States, one of the earliest laws with respect to platform governance was section 230 of the Communications Decency Act of 1996. It provides that 'no provider or user of an interactive computer service shall be treated as the publisher or speaker of any

---

[67] See R Gorwa 'The platform governance triangle: Conceptualising the informal regulation of online content' (2019) 8 *Internet Policy Review*. Gorwa grouped them into three. Firms, NGO and State. This categorisation is repurposed into two instead of three under the chapter. State actors and firms with NGOs grouped as non-state actors.

[68] Section 8(2) of the Ethiopia's 'Hate Speech and Disinformation Prevention and Suppression Proclamation No 1185 /2020 https://chilot.me/wp-content/uploads/2020/04/HATE-SPEECH-AND-DISINFORMATION-PREVENTION-AND-SUPPRESSION-PROCLAMATION.pdf (accessed 15 June 2020).

[69] The key principles of the EU liability regime for online intermediaries are enshrined in the E-commerce Directive passed in 2000. The Directive's overarching goal was to improve the development of electronic trade in the European Union (EU). The legislation introduced a 'safe harbour' principle, under which online intermediaries who host or transmit content provided by a third party are exempt from liability unless they are aware of the illegality and are not acting adequately to stop it. See Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the internal market <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32000L0031>; Singapore's 'Protection from Online Falsehoods and Manipulation Act 2019' <https://sso.agc.gov.sg/Acts-Supp/18-2019/Published/20190625?Doc Date=20190625> (accessed 5 February 2021). For Nigeria and South Africa, see sections 5.2 & 5.3 below.

© University of Pretoria

information provided by another information content provider.'[70] It is regarded as the foundational law with respect to traditional platform governance due to the fact that the platforms with the most reach across the globe are domiciled in the United States and should ideally be regulated outside the United States. This is because the impacts of online harms on social media platforms extend to other countries outside the US.[71] The law was a response to a court decision *Stratton Oakmont, Inc v Prodigy Services Co* in 1995 where the New York Supreme Court decided that online providers could be held liable for the speech of their users.[72] As a result, according to Kosseff, the US Congress felt that companies should have the discretion of content moderation.[73]

Extolling the impacts of section 230 on the right to freedom of expression, especially in the United States, Balkin stated that section 230 had huge impacts on the vibrant culture of freedom of expression online.[74] In establishing the safe harbour principle, the US Congress shaped the fundamental perception of editorial responsibility and independence especially for social media companies. This is because not only were platforms absolved from liability if they do not moderate content, they were also relieved of any possible claims that may arise if they moderate content.[75]

Looking at the first part, editorial responsibility will be considered as non-existent for online platforms as they are 'mere conduits' of information while such independence could be considered as the blanket immunity.[76] On the second part, editorial responsibility encourages platforms to moderate content but based on choice while editorial independence is in being able to make such choice without any liability. Therefore, neither action nor inaction could make platforms liable for content under the safe harbour treatment under section 230. In the words of Klonick, 'advances in technology as well as the immunity created for Internet intermediaries under section 230 led to a new generation of cyberspace.'[77]

In terms of specific regulatory requirements, most traditional approaches to platform governance, save for the United States' regime, require some responsibilities to be performed by platform, failure of which attracts various sanctions.[78] Penalty or enforcement systems under traditional platform governance are either terms of imprisonment or fines or both.[79] Sometimes, such a penalty system may also include administrative and legal sanctions that requires the affected party to desist or carry out an action.

---

[70] J Kosseff *The twenty-six words that created the Internet* (2019) 2.
[71] Section 3.6 above.
[72] 23 Media Law Report 1794.
[73] Cox and Wyden were the US Senators that spearheaded the provision; Kosseff (n 76) 3.
[74] J Balkin 'The future of free expression in a digital age' (2009) 36 *Pepperdine Law Review* 434; Gillespie (n 2) 30.
[75] As above.
[76] Gillespie (n 2 above) 31, 40, 206.
[77] Klonick (n 16 above) 1616.
[78] Gillespie (n 48 above) 6-12.
[79] As above.

© University of Pretoria

The earliest laws on platform governance in Africa can be broadly divided into two. Traditional platform governance for trust and traditional platform governance through direct influence. The traditional governance for trust came as a result of provisions that sought to protect the integrity of transactions with the advent of online commerce in Africa.[80] These provisions are often provided for in most African countries' cybersecurity, electronic transactions or cybercrime laws.[81] Responsibility in the laws were less dependent on platforms or businesses, the objective of the provisions were to ensure that businesses through their platforms ensure the integrity of transactions and protect consumers.

Conversely, the traditional platform governance through direct influence are such laws that place all the responsibilities for keeping platforms safe on the platforms.[82] While this seems like a logical approach, questions as to how the platforms carry out such responsibility should be asked as that has been shown by several examples, in the past, to be without clearly set out rules.[83] As a major characteristic of this form of traditional governance, the state or government enacts a law that mandates platforms to regulate harmful online content within a timeframe. This has been argued to pose danger to the right to freedom of expression in that platforms could over-censor and in some other instances could also cause platforms to under-moderate harmful content. In addition, given the nature of online harms and how they are largely conflated and generally misunderstood in African countries, such enforcement will directly impact on the right to freedom of expression online.[84]

Murray described the need for some traditional regulation by governments in platform governance because the Internet could not be without rules as this could be chaotic due to harms.[85] Murray's argument is strengthened by the impacts of online harms on the right to freedom of expression online in Africa. This strength is premised on the assumption that should everyone be allowed to continuously carry out harmful activities online, expression would no longer be free. This also addresses the argument that an absolutist approach to protecting the right to freedom of expression online does not take into account various existing contextual challenges like digital colonialism, weak public institutions, democratic culture in non-Western systems. It is important to note that under international human rights law, one of the requirements in the cumulative four-part test is legality, which requires that for speech to be limited,

[80] See N Kshetri 'Cybercrime and cybersecurity in Africa' (2019) 22 *Journal of Global Information* 77-81; F Cassim 'Addressing the growing spectre of cybercrime in Africa: Evaluating measures adopted by South Africa and other regional role players' (2011) https://core.ac.uk/download/pdf/79170924.pdf (accessed 23 July 2021).
[81] Media Defence 'Module 7: Cybercrimes' December 2020 https://www.mediadefence.org/ereader/wp-content/uploads/sites/2/2020/12/Module-7-Cybercrimes.pdf (accessed 20 July 2021). See section 4.5.1 above.
[82] Gorwa (n 1 above) 12.
[83] Section 2.4.3 above.
[84] Section 3.6 above.
[85] Murray (n 65) 205.

© University of Pretoria

it must be provided for by law, clear and sufficiently precise. This requirement is in addition to three other criteria: legitimacy, proportionality and necessity.

B Sectoral platform governance

In some cases, non-state actors like social media platforms and civil society identify major aspects of platform governance that are mutually reinforcing. Under sectoral platform governance, social media companies or civil society come together as a sector to set up a body of norms or principles to be guided by. These norms could be internal and external. They are internal when the norms are meant to guide just the players within the sector, for example social media companies or other platforms involved in content governance. These norms are external when they are developed by a number of actors with similar interests but designed to guide other actors outside their sector like social media platforms or state actors.

The Global Internet Forum to Counter Terrorism (GIFCT) presents an example of a sectoral platform governance model whose body of norms are inward-looking and are meant to regulate social media platforms internally on terrorist content.[86] The GIFCT presents an interesting model in that it was brought together by four major social media companies: Facebook, Twitter, YouTube and Microsoft to be advised by stakeholders like government and civil society to design norms to assist in regulating terrorist content. It is sectoral in that the norms were designed by social media platforms and internal-facing as the advice given by other stakeholders are meant to improve these platforms' moderation practices on terrorism-related content and not external actors.

Another example of sectoral platform governance are the Manilla Principles on Intermediary Liability (the Manilla Principles)[87] and the Santa Clara Principles on Transparency and Accountability in Content Moderation (the Santa Clara Principles).[88] The Manilla Principles on Intermediary Liability was originally led by seven civil society organisations from across the globe on the rules that guide the liability and regulation of Internet intermediaries including social media platforms.[89] It formulated six rules which can be grouped into four major governance components, the extent of liability for Internet intermediaries (liability); due process (judicial oversight); compliance with the three-part test (clarity, necessity and proportionality); and transparency and accountability. This example is sectoral as it is led by civil society but external in that it is outward-facing and deals with other proximate stakeholders within the platform governance ecosystem, for example governments and social media platforms.

---

[86] Global Internet Forum to Counter Terrorism https://gifct.org (accessed 15 March 2021).

[87] Manilla principles on Internet intermediaries https://manilaprinciples.org/principles.html (accessed 15 March 2021).

[88] Santa Clara principles https://www.eff.org/files/2015/07/08/manila_principles_background_paper.pdfaccessed (accessed 15 March 2021).

[89] Manilla principles on Internet intermediaries 'Background paper' https://www.eff.org/files/2015/07/08/manila_principles_background_paper.pdf (accessed 15 March 2021).

© University of Pretoria

The Santa Clara Principles was adopted by a number of civil society actors, academic institutions and individuals to 'provide meaningful due process to impacted speakers and better ensure that the enforcement of their content guidelines is fair, unbiased, proportional, and respectful of users' rights.' It has three major rules that border on publication of numbers with respect to platforms' moderation practices, notice to users and appeal systems. It is currently being reviewed to accommodate more diverse perspectives from across the world on global platform governance. This example is sectoral as it was developed and led by non-state actors but external in the sense that it requires social media platforms to be guided by its principles.

C Self-governance

Due to social pressure on the need for social media platforms to be more accountable, some of them opt for internal mechanisms that seek to hold themselves accountable. The main feature of such governance is that the platform regulates itself through Terms of Service and a set of rules adopted by the platform often referred to as 'community guidelines.' Both the Terms of Service and community guidelines are what platforms claim to guide their decisions on moderating online speech. Therefore, self-governance or self-regulation is a form of platform governance where a social media platform carries out its own moderation according to its own set of rules. This is what Land referred to as 'privatised governance' – enforcing private rules for public spaces.[90]

In defining self-regulation, Price noted that all formalities of governance would be present just that they would be done by the entity for whom it is intended to govern.[91] Perhaps the most popular form of self-governance of platforms is Facebook's Oversight Board (the Board). Klonick describes it as 'a novel articulation of Internet governance.'[92] In a process that began in November 2018, Facebook shared its concept for a Board that could help the company make its 'toughest content decisions.'[93] In a series of activities from January 2019 to October 2020, the Board formally began operations to receive complaints from Facebook and the public and give decisions based on such complaints. The Board, assisted by a Secretariat, is said to be independent as it is funded by a Trust for its day-to-day running and it uses Facebook's community standards, values and human rights norms on free expression

[90] MK Land 'Against privatized censorship: Proposals for responsible delegation' (2019) 60 *Virginia Journal of International Law* 364, 430.

[91] ME Price & S Verhulst 'The concept of self-regulation and the internet' in J Waltermann & M Machill (eds) *Protecting our children on the internet: Towards a new culture of responsibility* (2000) 137.

[92] K Klonick **'**The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression' 129 *Yale Law Journal* 2418.

[93] J Hirsch 'Where Facebook's Oversight Board falls short' 22 October 2019 *Centre for International Governance Innovation* https://www.cigionline.org/articles/where-facebooks-oversight-board-falls-short (accessed 4 May 2021). This has been criticised by Jesse Hirsch to mean Facebook foreclosing other aspects of its activities that require regulation like privacy rights.

© University of Pretoria

to arrive at its decisions.[94] The purpose of the Board is to promote free expression by making recommendations to Facebook Company Content Policy.[95]

It proposes to have 40 members drawn from across various regions in the world with varying expertise and experiences but started with 20 members.[96] The sole responsibility of the Board is to make decisions on select cases for review and either uphold or reverse Facebook's content decision. According to Facebook, the Board is not designed as an extension of Facebook's decision-making processes but to focus more on tough cases to determine whether they are made with respect to Facebook's values and policies.[97]

The commitment by Facebook for the Board has been criticised and can be divided along those who think the Board will be independent and be able to mainstream international human rights standards into its decisions and those who think that given Facebook's conflict of interest as a result of its business model, the Board would not be independent.[98] Klonick groups them into three as realists, pessimists and optimists.[99]

The criticisms of pessimists of the Board's relevance are that Facebook would lose its incentive to regulate online content more effectively; it is a method to keep governments' regulations at bay[100] and that Facebook is outsourcing its responsibility and challenges on online speech governance to the Board. For the realists, the Board would fail as it will not be able to cope with the volume of content involved in effectively regulating online content. The number of content that are published and require review will constantly increase therefore a Board made up of 40 people and a Secretariat will not be able to handle such.[101] As pointed out by Suzor *et al*, 'the velocity at which users publish content means that it is impossible for platforms to pre-moderate, or review it all in advance.'[102] The optimists feel that the Board presents an opportunity to

---

[94] Facebook 'Oversight Board Charter' (2019) 5 https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf (accessed 1 December 2021).

[95] Oversight Board 'Ensuring respect for free expression, through independent judgement' https://oversightboard.com (accessed 15 February 2021).

[96] Oversight Board 'Meet the Board' https://www.oversightboard.com/meet-the-board/ (accessed 1 December 2021).

[97] As above.

[98] Klonick (92 above) 2488.

[99] Klonick (92 above) 2488-2492.

[100] JJ González 'Everyone on Facebook's Oversight Board should resign' *Wired* https://www.wired.com/story/opinion-everyone-on-facebooks-oversight-board-should-resign/ (accessed 13 March 2021).

[101] Oversight Board 'The Oversight Board is accepting user appeals to remove content from Facebook and Instagram' April 2021 https://oversightboard.com/news/267806285017646-the-oversight-board-is-accepting-user-appeals-to-remove-content-from-facebook-and-instagram/ (accessed 10 March 2021).

[102] T Flew *et al* 'Internet regulation as media policy: Rethinking the question of digital communication platform governance' (2019) 10 *Journal of Digital Media and Policy* 33 50.

© University of Pretoria

experiment with user-driven community speech rules and that would bring Facebook closer to the local contexts.

Since the first decision made by the Board in January 2021, there have been various analyses on the basis for the Board's decisions.[103] Generally, it seems there is a cautious optimism by most free expression advocates in that these cases are still far and in-between to determine the operational and substantive direction of the Board. However, the basis for the decisions seems to be combining both international human rights law standards and more calls for transparency and accountability from Facebook.[104]

On 28 September 2021, the Board decided an appeal brought before it by a Facebook user based in South Africa.[105] In the appeal, the user had used certain words like '[y]ou are' a 'sophisticated slave,' 'a clever black,' 'n goeie kaffir' or 'House nigger' in their post in a public group on Facebook. The general context of the post, according to the user, was to engage the state of poverty and socioeconomic challenges in South Africa. Following review by a moderator, Facebook removed the post from its platform due to a report by another user for violating their Hate Speech Community Standard. Facebook notified the user of its decision which the user chose to appeal to the Board. In deciding the appeal, Facebook considered three standards: its Community Standard, values and human rights standards. The Board upheld Facebook's decision and found that the post violated Facebook's Community Standard, values and also other international human rights standards under the provisions of Article 19(3) of the ICCPR.

There are two major issues with respect to the Board's decision especially as it relates to the application of international human rights law. One, considering the Community Guidelines which Facebook and the Board based their decisions on, while the definition of hate speech might be clear, its description of prohibited content under Tier 3 as that which 'describes or negatively targets people with slurs, where slurs are defined as words that are inherently offensive and used as insulting labels for the

---

[103] E Debré 'The Facebook Oversight Board has made its first rulings' 28 January 2021 *Slate* https://slate.com/technology/2021/01/facebook-oversight-boards-content-moderation-rulings.html (accessed 3 June 2021); E Douek 'The Facebook Oversight Board has made its first rulings.' *Slate* https://slate.com/technology/2021/01/facebook-oversight-boards-content-moderation-rulings.html (accessed 4 February 2021).

[104] So far, the Board has decided eight cases that seek to apply international human rights standards. Case decision 2020-002-FB-UA https://oversightboard.com/decision/FB-I2T6526K/; Case decision 2020-003-FB-UA https://oversightboard.com/decision/FB-QBJDASCV/ Case decision 2020-004-IG-UA https://oversightboard.com/decision/IG-7THR3SI1/ Case decision 2020-005-FB-UA https://oversightboard.com/decision/FB-2RDRCAVQ/ Case decision 2020-006-FB-FBRhttps://oversightboard.com/decision/FB-XWJQBU9A/ Case decision 2020-007-FB-FBR https://oversightboard.com/decision/FB-R9K87402/ Case decision 2021-002-FB-U https://oversightboard.com/decision/FB-S6NRTDAJ/ Case decision 2021-003-FB-UA https://oversightboard.com/decision/FB-H6OZKDS3/ (accessed 25 February 2021).

[105] Oversight Board 'Case decision 2021-011-FB-UA' https://oversightboard.com/decision/FB-TYE2766G/ (accessed 6 October 2021).

© University of Pretoria

above characteristics' does not comply with international human rights law. This non-compliance is because free speech also includes offensive, shocking and disturbing speech, especially when they do not advocate direct harm or violence against another.[106] The post in question might be offensive, shocking and disturbing but the Board failed to show how it directly constitutes violence against another even when the user had noted that the content of the post was to highlight the socioeconomic plight in South Africa and discuss the issues involved.

Two, in deciding whether content is hate speech, international human rights standards require a higher threshold to be complied with. This threshold requires six additional factors to be considered. All six factors are: the socio-political context, the status of the speaker, intent of the speaker, content of speech, reach of the speech and likelihood of harm.[107] A closer look at the Board's decision shows that while it applied the four-part test of legality, legitimacy, necessity and proportionality, it failed to consider these additional factors in deciding whether the post constituted hate speech, especially in national contexts. For example, South Africa has a unique political history that needs to be factored in to arrive at the decision on whether a particular expression is hate speech or not.

Beyond Facebook and highlighting the challenges of self-governance by platforms, Haartman pointed out that speech regulation that focuses only on traditional governance and not relationships of power is inefficient.[108] This suggests that while focus is often on government regulation of online speech, platforms are also as active as states in their interests of governing online speech. Therefore, this means that as much as states wield a formal responsibility of protecting against online harms for example, social media platforms also have the same power if not more in their combination of both political power to shape global discourses and profit motives. One of the major criticisms against this feature is that it often lacks global contexts in terms of language, history, political, social or economic backgrounds of most countries outside the US and Europe. As Edwards puts it, there is a tough balance between the platforms making their own rules, staying profitable and fulfilling their roles in hosting public speech.[109]

As a result of this public function, platforms are unable to grapple with the many nuances involved in protecting freedoms in closed societies. As Cohn puts it, 'not only does censorship remove some useful speech, as with indirect intermediaries, marginalized groups are often the first to be impacted by private censorship by direct

[106] United Nations General Assembly 'Online hate speech and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/74/486' 7 September 2012 http://undocs.org/en/A/74/486 (accessed 6 October 2021) para 13.
[107] United Nations General Assembly (n 106 above) para 14.
[108] IA Hartmann 'A new framework for online content moderation' (2020) 36 *Computer Law & Security Review* 1-10.
[109] L Edwards 'Pornography, censorship and the Internet' in L. Edwards & C. Waelde (eds) *Law and the Internet* 2009.

© University of Pretoria

intermediaries.'[110] Penalty systems under self-regulation range from disabling various affordances to permanent suspension from such a platform. In curing the possible defects of self-governance by platforms, according to Arun, platforms, especially those experimenting with self-regulatory checks can insist on complying with international human rights law. In her argument, she pointed out that social media companies like Facebook could gain more support when they are willing to shun profit for rights.[111]

D Multistakeholder platform governance

The defining feature of multistakeholderism is that actors combine their interests in the development of rules that guide the regulation of online speech.[112] As described by Gorwa, 'such models seek to provide some values of democratic accountability without making extreme changes to the status quo.'[113] Underscoring the importance of this model of platform governance, Graham and MacLellan noted that:

> Governments cannot solve the problems alone – the transborder nature of the internet and the applications that make use of it simply make that impossible. The private sector almost certainly cannot and will not solve the problem in isolation; companies do not share the incentives to do so and do not have the necessary levers to deal with the impacts of their business models. Civil society also lacks the cohesion, the levers and the experience to deal with the challenges without cooperation from the other players. All of these actors need to come together to develop a shared understanding of the problems and the possible solution space, and then to work in good faith to find the way forward.[114]

Perhaps in agreement with Graham and MacLellan, Kaye suggests:

> Wherever the companies enjoy a market presence, they should develop multi-stakeholder councils, members of which they would compensate, to help them evaluate the hardest kinds of content problems, to evaluate emerging issues, and to dissent to the highest levels of company leadership. Multistakeholder governance can foster democracy, enrich existing representative frameworks and empower citizens in our interconnected and interdependent world.[115]

Oftentimes, big social media companies with both the financial and political capital combine such powers with that of governments' monopoly of power to make laws. Borrowing a leaf from various multistakeholder interventions on Internet governance, a multistakeholder governance platform model is 'a constantly shifting balance of powers between private industry, international technical governance institutions, governments, and civil society has characterized contemporary Internet governance

---

[110] C Cohn 'Bad facts make bad law: How platform censorship has failed so far and how to ensure that the response to Neo-Nazis doesn't make it worse' (2018) 2 *George Law Technology Review* 442.
[111] C Arun 'Facebook's faces' (2021) 135 *Harvard Law Review Forum* 2-29. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3805210 (accessed 15 February 2021).
[112] DeNardis (n 4) 228.
[113] Gorwa (n 1) 11.
[114] B Graham & M MacLellan 'Overview of the challenges posed by Internet platforms: Who should address them and how' in E Donahoe & FO Hampson (eds) *Governance innovation for a connected world protecting free expression, diversity and civic engagement in the global digital ecosystem* (2018) 12.
[115] Kaye (n 5) 87.

© University of Pretoria

approaches.'[116] In order to practicalise the workings of a multistakeholder approach to platform governance, activities would need to be centralised but the goal would be to create a deliberative process of content rules and enforcements.[117] The essence of such centralisation is to have a particular stakeholder with sufficient public interest on various aspects of platform governance lead 'organization, procedural mechanisms, and even hierarchy to ensure that all stakeholder voices are heard.'[118] DeNardis divides these centralisations into three major categories, widely diffused; government-led and private-sector-led.

For a widely diffused multistakeholder approach, the various processes and powers involved in governance will be evenly distributed among a wide range of actors including both state and non-state actors. With such an approach, the various aspects of content governance, right from the conception of its rules to its deployment and enforcements will be divided along major stakeholders. For example, with respect to remedies in content governance, the powers to enforce a rule must not only be provided by platforms but must be guided by international human rights law. Such an access to justice mechanisms would be developed based on the input of various users through their experiences. Therefore, the iteration of content remedies, from its rules to its delivery for users must reflect a deliberative process as much as possible. However, the challenge with such an approach is that it is practically difficult to divide powers among stakeholders equally but the goal is not to achieve a perfect balance but a fair one that allows for more generative development of the approach.

For the second approach, governments lead the process of a governance that includes both proximate actors.[119] However, one of the benefits of such a model is that governments have the resources to ensure its success which are sourced from legitimate laws. The government drives the mechanisms and processes in a bid to ensure that all voices with respect to various issues are heard and addressed. Governments are required under this model to step back from their traditional role of a public order prefect in governing platforms and step in as a contributory stakeholder that has as much at stake and as much to contribute like other stakeholders. The challenge with such a model is that most governments are more driven by political realities and the need to maintain power than human rights protection and therefore encroach on the powers of others while overstepping theirs.

---

[116] DeNardis (n 4) 227.

[117] B Chapelle 'Multistakeholder governance: Principles and challenges of an innovative political paradigm' in W Kleinwachter (ed) *Multistakeholder Internet Dialogue Collaboratory Discussion* Paper Series No 1 September 2011 http://dl.collaboratory.de/mind /mind_02_neu.pdf (accessed 15 February 2021).

[118] DeNardis (n 4) 229.

[119] Institute for Multistakeholder Initiative Integrity (IMII) 'Not fit-for-purpose: The grand experiment of multi-stakeholder initiatives in corporate accountability, human rights and global governance' July 2020 https://www.msi-integrity.org/wp-content/uploads/2020/07/MSI_Not_Fit_For_Purpose_FORWEBSITE.FINAL_.pdf (accessed 15 June 2021) 38.

© University of Pretoria

For the third approach, platforms themselves lead the initiative that puts more than them at the table to include civil society, think thanks, academics, research and development, human rights organisations and inter-governmental organisations.[120] Just as the government-led approach, the platforms coordinate the activities but not the outcome of a deliberative process in shaping an aspect of platform governance. The challenge with this approach is two-fold. On one hand, platforms struggle with the external perception that they are averse to human rights protection in developing their products while on the other hand platforms do have real interests in how they are governed. Therefore, achieving a balanced set of interests that ensures a democratic and human rights-based system is dreamy.

The forms of platform governance that exist overlap from time to time depending on the context of application. However, what sets each form apart is the degree of involvement of a particular actor. For example, it is a traditional form of platform where governments drive the main regulatory systems while it is sectoral platform governance where it is primarily driven by a specific sector such as social media platforms or civil society. As described above, each actor, depending on the form of platform governance may be identified through their penalty or enforcement systems. While traditional platform governance is typified by law, state institutions are primarily driven by state actors, its penalty system often comprises both the force of law which indirectly requires non-state actors like platforms to carry out specific actions or desist from them. On the other hand, while self-regulation is primarily driven by social media platforms, its actions are based on two major instruments: community guidelines or rules and terms of contracts. Both documents are developed and enforced by social media companies with direct influence on moderation of content. This shows that the typology of a penalty or an enforcement system is determined by the kind of actor that primarily drives such a form of platform governance.

Table 3: Typology of platform governance

Form of platform governance	Traditional	Sectoral	Self-regulatory	Multistakeholder

[120] See Facebook 'Global feedback and input on the Facebook Oversight Board for content decisions' 27 June 2019 https://about.fb.com/news/2019/06/global-feedback-on-oversight-board/ (accessed 13 October 2021). It is important to note that while the process that led to the Board's take-off included diverse stakeholders, the form of platform governance that best fits the outcome is self-governance. This is because the Board's decisions are only limited to Facebook's activities and not other social media platforms. The kind of multi-stakeholder to be designed for platform governance would need to cater for all social media platforms.

© University of Pretoria

Actor type	Government /Government institutions	Platforms/civil society	Social media companies	Multi-actor
Mode of governance	Laws and policies	Guidelines, associational rules	Community guidelines, Terms of contracts, internal adjudicatory mechanisms	Soft and hard laws
Enforcement	Indirect	Indirect/Direct	Direct	Indirect
Penalty	Criminal sanctions or civil liability	Down ranking, filtering, take downs, temporary and permanent removal of account or content	Down ranking, filtering, take downs, temporary and permanent removal of account or content	Soft laws: Down ranking, filtering, take downs, temporary and permanent removal of account or content. Hard laws: Criminal, civil and administrative sanctions for online harms that are both harmful and illegal.

4.3 Limitations of platform governance

Due to the competing interests of various actors in the platform governance ecosystem, it becomes difficult to realise an effective model. Social media companies now influence the outcome of democratic elections and have become economic behemoths, ranking nearly as much as some governments in political power.[121] The major difference between these platforms and governments is that while the former may be powerful, the latter still retains the monopoly of making binding laws that guide

[121] Suzor (n 1 above).

© University of Pretoria

societies. This difference therefore typifies the major limitations of platform governance which have been divided broadly into four below.

4.3.1 Surveillance capitalism and economic might

One of the justifications for looking to limit the powers of social media platforms is that they have become both economically and politically powerful. This power which has been amassed by the use of advanced predictive analytics of online users has been the major argument against the might of platforms. The use of various frontier technologies to surveil online human behaviour and target individual users with advertisements based on such surveillance to generate revenue has been described as surveillance capitalism. According to Zuboff, surveillance capitalism is a market and the technological tools would not survive outside the digital space.[122]

Zuboff described surveillance capitalism as the controller that imposes an instrumentarian power through the re-engineering of behaviours.[123] Such instrumentarian power was defined as 'instrumentation and instrumentalisation of behaviour for the purposes of modification, prediction, monetization and control.' In other words, according to her, impacts of instrumentarian power is best seen in how it weaponises 'radical behaviourism' to usurp elemental rights.[124] Snowden extolled the pre-commercialisation era of the Internet which saw more to the task of creating a global community than commodifying connections. He argued that surveillance capitalism began when companies realised that human connection could be monetised.[125]

Linking Snowden's arguments to the aspect above on two tenuous but valid facts on preventing online harms, social media platforms who are currently at the heart of e-commerce – deploying various frontier technologies to pre-determine online behaviour for profit are unable to make rules that fundamentally impact their revenue.[126] The kind of rule that most social media platforms ought to defer to, going by Snowden's points, cannot be made by the same social media platforms as it affects their business model, which thrives on unreasonable inferences of human behaviour through online content.[127] Since such human behaviour is determined by both online and offline speech, platforms will need to rely on a more external form of regulation to ensure effective governance.[128]

---

[122] S Zuboff '"We make them dance": Surveillance capitalism, the rise of instrumentarian power, and the threat to human rights' in RF Jørgensen (ed) *Human rights in the age of platforms* (2019) 10.
[123] Zuboff (n 122 above) 28.
[124] Zuboff (n 122 above) 39.
[125] E Snowden *Permanent record* (2019) 4.
[126] Klonick (n 16 above) 1627.
[127] S Wachter & B Mittelstadt 'A right to reasonable inferences: Re-thinking data protection law in the age of Big Data and AI' (2019) 2 *Columbia Business Law Review* 494-620.
[128] H Bloch-Wehba 'Global platform governance: Private power in the shadow of the state' (2019) 72 *Southern Methodist University Law Review* 55.

© University of Pretoria

4.3.2 Lack of context

Platforms struggle with balance. The balance sought to be achieved is that between local contexts like the historical, social, economic aspects which influence a form of content or online speech and the moderation practices of social media companies. Often, a content may be problematic within one context but allowed in another, both of which could be permissible under international human rights standards. In particular, this occurs majorly in two ways. First, the cultures that influence various human rights on the ground including the right to freedom of expression vary from one national or local context to another. This creates a challenge for social media companies charged with moderating such content with the same set of global rules which then end up lacking in necessary nuance. For example, the moderation of nude content on social media platforms swings between the right to freely express using one's body and pornography. In some instances, it is a recognised form of expression or a means of protest in others.

However, in some national contexts, such expression whether online or offline is disallowed, owing to several contextual influences. Second, social media platforms rely a lot on automated systems which flag content based on a set of pre-taught information. Automated platform governance therefore combines several technology tools to increase, reduce or remove content based on a set of programmed rules in artificial filters. The major challenge with this is that such automated systems cannot distinguish between Southern African women who can choose to be bare-breasted as a result of a traditional ceremony or protest and others who share nude pictures for pornographic purposes. Appreciating the challenge posed by scaling global rules to specific contexts, Owen argued that it would be difficult for one single actor to govern platforms.[129]

4.3.3 Clashes of free speech laws and actors

Most popular social media platforms are domiciled in the US and are therefore governed by its laws. Section 230 which is one of the most proximate laws is backed by the First Amendment which is the most primary law on the right to freedom of expression online. Both laws have been the lodestar of most social media platforms and neither of these laws express responsibilities on preventing harm nor guard against the extra-territorial impacts of social media platforms. The inabilities of both provisions are the main tensions between social media platforms' inability to effectively police online harms and protect free speech outside the US.

While international human rights law applies broadly with specific principles on social media governance and online speech, it is largely prescriptive as there are various contexts that influence its application locally. For example, the form of speech that may constitute harm under the law in the US or United Kingdom may not be the same

[129] T Owen 'Introduction: Why platform governance?' 28 October 2019 *Centre for International Governance Innovation* https://www.cigionline.org/articles/introduction-why-platform-governance (accessed 4 February 2021).

© University of Pretoria

kind of speech that may constitute harm under the law in Ghana or India because context tempers speech. Given that social media platforms mostly use software filtering systems for content moderation, such an approach would fail in identifying the context with respect to a problematic speech. Such inability then leads to misidentifying, mislabelling, censorship by automation and possible violation of potentially problematic national laws.[130] This constitutes a major challenge to platform governance in that social media companies are generally global rather than local, profit-oriented rather than community-driven, and First Amendment-focused rather than international human rights law-based. As a result, free speech laws in many contexts continue to clash first with social media platforms' motivations and later with their software filters that do not understand contexts.

4.3.4 Impractical domestic laws

Section 230 of the CDA is inherently problematic especially with respect to non-US contexts. For example, African governments are using both colonial laws and problematic cyber laws to police online speech. The treatment of social media platforms as 'mere conduits' under section 230 therefore emboldens governments' claim that the platforms do nothing to address online harms.[131] This is the reason why some impractical domestic laws are often noticed in various provisions on criminal defamation, false news and publication, blasphemy and others to feign combating these harms when African governments only use the laws to silence dissent. In these African countries, there laws provide for criminal sanctions for disinformation which is not a basis for limiting speech except they meet the criteria set out under international human rights law.[132] These laws combined with the nature of section 230 of the CDA further make it difficult to regulate online harms and protect the right to freedom of expression on social media platforms in the African context.

In addition to this, these laws provide the pretext for African governments to block access to the Internet which neither has basis under international human rights law nor their various constitutions.[133] In some cases, these laws are modelled towards laws in France and Germany which require platforms to carry out specific moderation actions with respect to disinformation or hate speech but do not factor in the enforceability.[134] When governments enact such laws that require that platforms comply with a particular timeline in removing a problematic content, not only is enforcing such law problematic, they further put social media platforms under pressure

---

[130] Ilori (n 55 above).

[131] G De Gregorio & N Stremlau 'Internet shutdowns and the limits of the law' (2020) 14 *International Journal of Communications* 4224 4243; E Marchant & N Stremlau 'The changing landscape of internet shutdowns in Africa' (2020) 14 *International Journal of Communications* 4216-4220.

[132] Section 2.4.3 above.

[133] *Amnesty International Togo vs The Togolese Republic* (ECW/CCJ/APP/61/18) (2020) ECOWASCJ 09; Gregorio & Stremlau (n 131 above).

[134] A De Streel 'Online platforms' moderation of illegal content: Law, practices' (2020) 88-91 https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf (accessed 15 October 2020).

© University of Pretoria

to unduly limit online speech.[135] Therefore, the traditional longing by African governments to reach for legal solutions for dynamic challenges such as platform governance further puts such governance in jeopardy as online speech becomes the casualty. While the domestic laws like the United States' First Amendment is proximate to regulating social media platforms, it is a local law within a state and should not be the golden rule other states abide by. In this instance, due to the global reach of these platforms, international human rights law will provide an ideal bases for such regulation.

4.4 A human rights perspective to platform governance and online harms in Africa

Since the first Internet connection in Tunisia in 1991,[136] African countries have continued to record its various impacts. Today, more than 40% of Africans are connected to the Internet.[137] The economic prospects of Internet access have been linked to huge developmental gains for African countries.[138] Due to various challenges with protecting human rights, many Africans have also had to deploy the power of the Internet to resituate both social and political powers in Africa.[139] Today, hashtags are being combined with placards in many protests like #RhodesMustFall, #EndSARS, #Jan25, #BringBackOurGirls, to reshape public policy in various contexts in the region.[140] Given these experiences, the Internet is not just an infrastructure for

[135] United Nations General Assembly 'Online content regulation and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/38/35' 6 April 2018 http://undocs.org/en/A/HRC/38/35 (accessed 26 August 2021).

[136] Internet Society 'Full IP access timeline' https://www.internetsociety.org/wp-content/uploads/2017/09/history_internet_africa.pdf (accessed 17 February 2021); A Budree, K Fietkiewicz & E Lins 'Investigating usage of social media platforms in South Africa' (2019) 11 *The African Journal of Information Systems* 315, 333.

[137] International Telecommunications Union (ITU) 'Measuring digital development: facts and figures 2020' *ITU* https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf (accessed 18 February 2021).

[138] M Guerriero 'The impact of Internet connectivity on economic development in sub-Saharan Africa' *Economic and Private Sector* (2015) 1-20 https://assets.publishing.service.gov.uk/media/57a0899b40f0b652dd0002f4/The-impact-of-internet-connectivity-on-economic-development-in-Sub-Saharan-Africa.pdf (accessed 21 February 2021).

[139] M Dwyer & T Molony 'Mapping the study of politics and social media use in Africa' in M Dwyer & T Molony (eds) *Social media and politics: Democracy, censorship and security* (2019) 1-19.

[140] T Bosch 'Twitter activism and youth in South Africa: The case of #RhodesMustFall' *Information, Communication & Society* 2016 10; IB Ochi & KC Mark 'Effect of the #EndSARS protests on the Nigerian economy' 9 *Global Journal of Arts, Humanities and Social Sciences* 2021 2; Y Kazeem 'How a youth-led digital movement is driving Nigeria's largest protests in a decade' 13 October 2020 *QUARTZ Africa* https://qz.com/africa/1916319/how-nigerians-use-social-media-to-organize-endsars-protests/ (accessed 16 February 2021); R Salanova 'Social media and political change: The case of the 2011 revolutions in Tunisia and Egypt' 7 *ICIP Working Papers* 7 50; L Jacinto '#Jan25 hashtags resurfaces twenty years after Egypt's revolution' 25 January 2021 *FRANCE 24* https://www.france24.com/en/africa/20210125-a-hashtag-resurfaces-10-years-after-egypt-s-revolution-and-the-posts-are-bittersweet (accessed 20 February 2021); SS Ofori-Parku & D Moscato 'Hashtag activism as a form of political action: A qualitative analysis of the #BringBackOurGirls Campaign in Nigerian, UK, and US Press' 23 *International Journal of Communication* 2018 2480-2502.

© University of Pretoria

Africans, it is a tool for liberalisation. It is this same perspective that informs the use of most social media platforms in Africa.

Social media platforms today have become a major source of information for Africans. Asides this, it has also become a platform of opportunities where ideas are being connected to execution and publics are connecting with each other one phone screen scroll at a time. This power to connect diverse views and possibilities is also the bane of social media in Africa. It has become a hotbed for radicalisation, information disorder and subversive manipulations. The power of social media platforms today seems to be their weakness, liberating yet harmful. Online harms now cause offline violence, platforms have become the new governors, the old ones are deploying both the law and extra-legal means to violate online speech, all of these and many more point to the mixed impacts of social media platforms in African countries.

On 3 June 2021, Twitter restricted a tweet from the account of the Nigerian President, Mohammed Buhari for violating their rules with respect to harmful speech and the following day, the Nigerian government announced the ban of Twitter's operations in Nigeria.[141] Just months before, Ugandan government blocked Facebook for suspending pro-Government accounts which Facebook claimed were being used to spread propaganda.[142] Access to Twitter and WhatsApp were restored in April 2021 in Tanzania where they were blocked during the 2020 General Elections to stem the spread of misinformation.[143] These examples point to the potential democratising power that social media platforms possess across the world, and the threat that African governments perceive as a result. At the same time, the various online harms that are now endemic on social media platforms have increased the need to scrutinise their power. The twin-power to act as 'mere conduit' of vast amounts of content that have shaped lives and mould democracies while also being the judge of what speech is free or is not has made social media platforms crucial nodes of power in the digital age. However, since more power should mean more scrutiny and transparency in the context of regulating social media platforms, international human rights standards provide the best approach to governance.

---

[141] Some of the reasons that have been adduced for the ban are both immediate and remote. The immediate reason was the restriction of the President's tweet and the remote reason was the claim made by Nigeria's Minister for Information that Twitter's CEO, Jack Dorsey funded the #EndSARS protests in October 2020. The #EndSARS protests was a nationwide protest that called on the Nigerian government to put an end to police brutality in Nigeria. This claim has however been debunked. See J Ojo 'Did Twitter fund #EndSARS protests as Lai claimed?' 5 June 2021 *The Cable* https://www.thecable.ng/fact-check-did-twitter-fund-endsars-protests-as-lai-claimed (accessed 10 June 2021); See also J Campbell 'Nigerian President Buhari clashes with Twitter Chief Executive Dorsey' 8 July 2021 *Council on Foreign Relations* https://www.cfr.org/blog/nigerian-president-buhari-clashes-twitter-chief-executive-dorsey (accessed 16 July 2021).

[142] H Athumani 'Ugandan government restores social media sites, except Facebook' 10 February 2021 *Voice of America* https://www.voanews.com/a/africa_ugandan-government-restores-social-media-sites-except-facebook/6201864.html (accessed 16 April 2021).

[143] T Karombo 'Tanzania has blocked social media, bulk SMS as its election polls open' 28 October 2020 *QUARTZ AFRICA* https://qz.com/africa/1923616/tanzanias-magufuli-blocks-twitter-facebook-sms-on-election-eve/ (accessed 16 April 2021).

© University of Pretoria

There are various conceptions of what a human rights approach means, but they can be collectively summarised as:

> comprehensive in their consideration of the full range of indivisible, interdependent and interrelated rights: civil, cultural, economic, political and social. Rights-based approaches also focus on the development of adequate laws, policies, institutions, administrative procedures and practices, as well as on the mechanisms of redress and accountability that can deliver on entitlements, respond to denial and violations, and ensure accountability. They call for the translation of universal standards into locally determined benchmarks for measuring progress and enhancing accountability.[144]

This offers an important point of departure on regulating online harms in Africa. This is because paying attention to the last part of the quotation above, a human rights-based approach is how, in this context, international human rights law can be applied to 'locally determined benchmarks' which in this case are mostly African governments.

There are two major considerations on the application of international human rights law to platform governance. The first consideration argues that international human rights law does not offer specific solutions to broad challenges like platform governance.[145] Given that cultural relativism is one of the factors that influence the application of human rights on ground, international human rights law only offers persuasive and limited influence. In describing the challenges posed by international human rights law, Dvoskin argues that it is not a universally accepted rule, highly indeterminate and could create a façade of legitimacy for social media platforms.[146] The second consideration points out that not only does international human rights law offer the basis for regulating online speech on platforms whose impacts are now global, it offers the most feasible framework upon which such regulation can be built.

---

[144] JC Mubangizi 'A human rights-based approach to development in Africa: Opportunities and challenges (2014) 39 *Journal of Social Sciences* 67-76; M Broberg & H Sano 'Strengths and weaknesses in a human rights-based approach to international development: An analysis of a rights-based approach to development assistance based on practical experiences (2018) 22 *The International Journal of Human Rights* 664-680; D Olowu *An integrative rights-based approach to human development in Africa* (2009) 15, 16, 72. LA Abdulrauf 'The legal protection of privacy in Nigeria: Lessons from Canada and Nigeria' unpublished LLD thesis, University of Pretoria, 2015 308-359 https://repository.up.ac.za/bitstream/handle/2263/53129/Abdulrauf_Legal_2015.pdf?sequence=1&isAllowed=y (accessed 15 June 2020); F Sagasti 'A human rights approach to democratic governance and development' in UNOHRC (ed) *Realizing the Right to Development: Essays in Commemoration of 25 Years of the United Nations Declaration on the Right to Development* (2013) 125 https://www.ohchr.org/Documents/Issues/Development/RTDBook/PartIIChapter9.pdf (accessed 15 February 2021).

[145] B Dvoskin 'International human rights law is not enough to fix content moderation's legitimacy crisis' 16 September 2020 *Berkman Klein Center Collection* https://medium.com/berkman-klein-center/international-human-rights-law-is-not-enough-to-fix-content-moderations-legitimacy-crisis-a80e3ed9abbd (accessed 20 February 2021).

[146] As above.

© University of Pretoria

Aswad groups the possible arguments against international human rights law into four.[147] First, those who conflate international human rights law make it seem as if the application of international human rights law is difficult on the ground with respect to the right to freedom of expression and content governance.[148] Second, those who think that international human rights law does not offer enough guidance for social media platforms but ignore the wealth of jurisprudence and standard-setting that various UN mechanisms have been involved with over the decades like the UN Special Rapporteur on the Right to Freedom of Expression and Opinion and others.[149] Third, those who consider the US First Amendment which section 230 is modelled after as the ideal basis of regulating online speech which does not incorporate the speech codes of other contexts and therefore cannot apply.[150] Four, those who may argue that international human rights law does not apply to governance at scale beyond the United States. Aswad argued that the goal of international human rights law is to anchor speech codes to ensure 'home-grown' solutions but not necessarily to carry out such solutions.

In regulating online harms through international human rights law, Aswad pointed out that not only do social media platforms already seek a universal basis for anchoring their speech codes due to the challenge of scale but it presents an opportunity for companies to push back against authoritarian systems. In a rather tacit agreement with Aswad, but in a different way, Dvoskin suggested the involvement of the public other than international human rights bodies in making speech rules.[151]

Given these positions, a recent paper by the Business for Social Responsibility (BSR) demonstrates how international human rights law can be practically applied towards platform governance on online speech.[152] The BSR is a group of experts that work with companies on sustainability. A human rights-based approach to platform governance is categorised into four parts: content policy; content policy implementation; product development; and tracking and transparency. The features of the four categories in the approach are that they allow for human rights due diligence of content moderation and a basis for contextual application of international human rights law.

The proposal identifies the key role of governments and social media platforms, argues for an international law approach to platform governance and the responsibilities of social media companies under the UN Guiding Principles on

---

[147] Aswad (n 13 above) 57-64. Kaye also addresses the criticisms of the international human rights system as basis for platform governance in his book, *Speech Police: The global struggle to govern the Internet* with similar arguments to those of Aswad's; See D Kaye *Speech police: The global struggle to govern the Internet* (2019).

[148] Aswad (n 13 above) 57-59.

[149] Section 2.4.3 above.

[150] Aswad (n 13 above) 60.

[151] Dvoskin (n 145 above).

[152] Business for Social Responsibility (BSR) 'A human rights-based approach to content governance' March 2021 https://www.bsr.org/reports/A_Human_Rights-Based_Approach_to_Content_Governance.pdf (accessed 2 April 2021).

© University of Pretoria

Business and Human Rights (UNGPs).[153] The paper was informed by the works of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Expression, international digital rights organisations like Access Now, Global Network Initiative, Global Partners Digital, and various scholars in the fields of human rights and content governance.

4.4.1 Content policy implementation

In developing content policies, social media platforms are advised to use international human rights principles as standards in order to have a counterweight argument against any illegal demand for censorship. In pushing back against illegal requests from both state and non-state actors, social media platforms have both legal and ethical reasons to rely on international human rights law for two reasons. One is that governments have the obligation under international human rights law to protect human rights and therefore, social media platforms relying on such as opposed to local laws that violate human rights are legally substantiated. Second is that international human rights standards also require social media platforms to apply a human rights-based approach to its governance and therefore has a duty to respect human rights and the direct responsibility to protect it. With respect to such duty and direct responsibility, Callamard argued that:

> a textual analysis of Article 19 and a review of the *travaux préparatoires* further suggest that duties and responsibilities attached to these non-state actors are not one and the same but 'specific.' They are distinct from the obligations imposed upon states, and they are distinct from one non-state actor to another.[154]

In doing this, some of the ways identified are that content policy should encompass all human rights, be informed by stakeholder engagement and be able to distinguish between paid and organic speech.

In implementing content policies, the BSR paper identified three challenges social media platforms need to surmount as the scale of content to be reviewed; making timely decisions; and ensuring contextual nuances.[155] In resolving these and achieving effective implementation, it suggests that such policies should be informed by the stakeholders and experts who understand the context. It also suggests that content policies should apply the joint three-part test by considering whether their moderation is the least intrusive, poses clear threats to human rights and cannot be achieved by other means. Particularly, it noted that isolated decisions that need to be combined with other decisions in order to determine their human rights impact requires special attention. For example, permanently suspending a bot that spreads hate speech online might not be enough until it includes investigations into its coordinated behaviour and reach. Social media platforms are also advised to choose international human rights principles when conflicted with other local laws.

---

[153] As above.
[154] A Callamard 'The human rights obligations of non-state actors' in RF Jørgensen (ed) *Human rights in the age of platforms* (2019) 199.
[155] BSR (n 152 above) 4-7.

© University of Pretoria

In addition to these, in implementing content policies, social media companies must prioritise decisions based on the severity of the case, because of the high volume of content requiring attention, priority would need to be given to the ones that are dire and have easily foreseeable far-reaching negative effects, especially offline. Conflict-affected areas are also required to be prioritised given the nature of how online harms could increase offline harm through 'heightened' and 'enhanced' due diligence. One of the ways to achieve this is for social media platforms to hire qualified professionals with both cultural knowledge and content moderation skills for specific contexts. Importantly, it connects the responsibility of social media platforms to ensure effective appeal systems to the rules of 'legitimacy, accessibility, predictability, equitability, and transparency' under the UNGPs. It appears that what is referred to as legitimacy here is the authorial, legal and formal basis of international human rights law as the basis for appeal mechanisms on social media platforms. Accessibility would mean the ability of the average user to register their concern with respect to a content moderation decision and receive timely feedback. Such ability to access the mechanism would include the literacy of such an average user of the social media platform.

For example, an appeal mechanism on a social media platform cannot be said to be accessible if users do not know it exists. Predictability in this context would be that the process of administering such an appeal mechanism if given the same set of facts would deliver the same set of results. In ensuring equitability in content moderation appeals, social media platforms must understand the context in rendering justice. For example, if a sexual minority Facebook page in Uganda is removed due to reports that such page promotes Satanism and violates a local law on same-sex relations, equity demands that the platform understands the plight of sexual minorities in Uganda in its appeal decision. Lastly, transparency is required to ensure an effective appeal system on social media platforms. This is because transparency is not only about seeing how appeal mechanisms work but also how to further improve the challenges social media platforms face along the way. Social media companies are also advised to ensure effective remedy in that such a decision should restore a victim to its original or near-original place before the harm occurred. It identified five ways such restoration can occur. They are satisfaction, restitution, non-repetition, rehabilitation and compensation.[156]

4.4.2 Product development

Just as privacy by design exists,[157] the BSR paper proposes that human rights be designed into social media platforms' products.[158] It notes that social media platforms should carry out human rights-assessment of new features or affordances on their platforms before their launch. It also points out that platforms should consider that their products may from time to time have unintended consequences and be well-poised to

[156] BSR (152 above) 6.
[157] A Cavoukian 'Privacy by design: The 7 foundational principles' 2011 https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf (accessed 5 February 2021).
[158] BSR (n 152 above) 7.

© University of Pretoria

address them. It also notes that platforms should design privacy and consent-based features into their products.

4.4.3 Tracking and transparency

In monitoring human rights compliance, social media platforms should measure their performance through more qualitative and quantitative indicators.[159] Qualitative indicators could be by paying attention to diverse and multi-disciplinary scholarship on how to improve social media transparency and the welfare of its human moderators, while quantitative indicators could be the various response times on content moderation decisions and appeal systems, error rates etc. It also suggests publication of annual reports by social media companies on how they have complied with internationally set standards. They are also enjoined to be transparent on their various motivations behind major decisions. As steps that apply all through the four segments, social media companies' content policies should be stakeholder-inclusive, pay more attention to human rights defenders and adopt a structural approach to vulnerability.

It is ironic, however, that a paper that seeks to provide a contextual application of international human rights law does not list a Global South scholar as one of its experts. A Global South scholar would be such an academic or civil society that is focused on platform governance impacts in systems outside Western countries. While this does not foreclose that possibility of Global South participation through the various international digital rights organisations consulted, having an expert who not only has the expertise on platform governance but also experiences its impacts from the Global South would have contributed more to the shaping of the paper. However, this does not mean that the paper does not have useful insights for the discourse on human rights and platform governance.

Social media companies perform a governance function when they enact censorship (either requested by a government or of their own volition) or carry out a law enforcement function.[160] Even more so, they wield significant governance power when they use their discretionary authority to *not* carry out censorship requests. This differential illustrates the discretionary regulatory role private companies assume.[161] According to DeNardis, social media platforms are now actively involved in preventing online harms.[162] This underscores the major need for social media platforms to adopt a more deliberate human rights approach to their regulatory actions on online speech. Not only have the arguments that international human rights neither applies to social media platforms nor proffer a viable platform for regulating online harms, they have

[159] BSR (n 152 above) 8-9.
[160] DeNardis (n 4 above) 159.
[161] As above.
[162] DeNardis (n 4 above) 156.

© University of Pretoria

been described as the most authoritative and viable basis for regulating online speech on social media platforms.[163]

However, given these positions, Sanders further subdivided a human rights approach into three, which includes: the application of substantive international human rights law to online speech governance process referred to as the substantive dimension; a process dimension that enhances transparency and regulatory oversight; and a procedural dimension that guarantees just remedies and mechanisms for human rights protection.[164] According to Sanders, these three rights-based dimensions of social media regulation are not enough and should be expanded. In order to effectively regulate online speech, Sanders argues that such governance cannot be pigeonholed into human rights-based approaches but to also include other thematic aspects of governance which includes data protection law, electoral and advertising regulation, and antitrust regulations. What this suggests is that, in thinking through an effective platform governance model, it must be such that it is not limited by narrow perspectives but that which benefits from the various intersections of online speech governance.

4.5 Platform governance and online harms in Africa

Understanding the important role of a human rights-based approach to online speech governance requires that such approach is applied to specific online harms as has been discussed in chapter three. The extent of success of existing forms of platform governance with respect to regulating online harms is important in order to further underscore the need for a creative governance system within international human rights standards.

4.5.1 Information disorder and platform governance in Africa

Currently, in many African countries, false information online is criminalised. This is often the case in countries whose legal systems have been influenced by colonial laws. For example, Kenya, Uganda, Nigeria and a number of other African countries have laws that criminalise the spread of false information online which may be found in most cybercrime or electronic communications or computer misuse laws that criminalise not only 'false information' but 'insults' or 'annoying' speeches.[165]

---

[163] Global Network Initiative (GNI) 'Content regulation and human rights' September 2020 https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf (accessed 23 February 2021).

[164] B Sanders 'Freedom of expression in the age of online platforms: The promise and pitfalls of a human rights-based approach to content moderation' (2020) 43 *Fordham International Law Journal* 955-1004.

[165] Kenya's section 22 of Computer Misuse and Cybercrime Act, 2018; Tanzania's section 16 of the Cybercrime Act, 2015; Nigeria's section 24 of the Cybercrime (Prohibition, Prevention etc) Act, 2015. Section 2.4.3 above.

© University of Pretoria

The right to freedom of expression is not limited to just sharing the truth.[166] A position that was also restated by Kaye, a law professor and former UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Expression and Opinion that 'lies are speech and speech is protected.'[167] The reason for such a position is that the nature of what is true from time to time depends on context. For example, satirical, sarcastic, cartoons are all forms of artistic expression which when considered superficially are untruths. However, these untruths serve social and political purposes in that they shape social behaviours while also strengthening political speech. Consequently, these are the laws that most social media platforms aim to defer to while responding to moderation requests either by governments or individuals.

This is one of the major criticisms against traditional platform governance in that the laws provided for by governments violate freedom of expression online. In addition to this, regulating information disorder as an online harm requires more than laws. As a subject of regulation, information disorder requires more creative governance because they change depending on context and limitations on the right to freedom of expression through laws or policies that do not reflect such flexibility will be inimical. Therefore, it requires a system that first incorporates the expertise, experience and impacts of various stakeholders into its making. Not solely regulated by platforms nor groups of platforms or one or two other proximate stakeholders, such a system will include not only multi-disciplinary expertise and experiences but also other stakeholders including users impacted by problematic few actor-led initiatives. For example, journalists, human rights practitioners, lawyers, linguists should have as much presence in such a system as software programmers, techno-policy experts, engineers and others. At the stakeholder level, it would include state actors like representatives of various arms of government and non-state actors like civil society, social media companies, academia etc.

4.5.2 Targeted online violence and platform governance in Africa

A combination of one or two forms of platform governance could be effective in some other aspects of targeted online harm including hate speech online, cyberharassment, online gender violence, non-consensual sharing of intimate images etc. This is because beyond the punitive function of laws, laws should also encourage alternative methods of regulation which include creating more awareness, re-modelling educational curricula, media empowerment and so on. Except in grave instances of online harmful speech for example, as provided for under the African Declaration, laws

[166] Joint declaration on freedom of expression and 'fake news', disinformation and propaganda https://www.osce.org/files/f/documents/6/8/302796.pdf (accessed 24 March 2021). The declaration stressed that 'the human right to impart information and ideas is not limited to "correct" statements, that the right also protects information and ideas that may shock, offend and disturb, and that prohibitions on disinformation may violate international human rights standards, while, at the same time, this does not justify the dissemination of knowingly or recklessly false statements by official or State actors.'
[167] Kaye (n 5 above) 81.

© University of Pretoria

on online harms should be drafted with more alternative approaches to regulation that are less adversarial and punitive.[168]

Another benefit of such form of governance is that where it allows for diverse input in its processes and comply with international human rights standards on online harms, it assists with more chances of success mainly because it is driven by specific actors – they have more authoritative access to resources that could make such governance smooth and effective more than any other actor.[169] A challenge with the forms of platform governance referred to above is that it could lead to over-censoring online expression or under-regulating online harms by platforms. Over-censorship in this context is not limited to the number of moderation decisions but also the proportionality in the response of platforms in such decisions. Under-regulation means instances when platforms refuse to take action due to either under-reporting or system challenges. This is because platforms and their systems do not often understand the various contexts in non-Western systems and as a result, they are prone to acceding to government laws which have been proven to be deeply problematic with respect to human rights protection or they do not take any action at all.[170] An offshoot of this challenge, as Suzor *et al* puts it, is that 'national media policies also cause difficulties as states again seek to impose their content regulation rules globally.'[171] Critically considering the various roles of governments in regulation of online harms, Land argued that governments need to do more than outsourcing governance through impractical laws.[172]

With a more practical argument, Land and Hamilton argue that in mitigating the harms caused by hate speech online for example, approaches to its regulation must be more nuanced and inclusive of proximate actors.[173] On the question of governance, DeNardis argues that 'the appropriate question involves determining what is the most effective form of governance in each specific context.'[174]

4.6 A generative approach to platform governance, the prevention of online harms and protection of online expression in Africa

Given the various aspects of platform governance that have been considered above to foreground an understanding of it especially from an African perspective, it is important to attempt a rethink of both its theoretical and applied aspects. This is because in preventing online harms and proposing a human rights approach to

[168] Section 2.4.2 above.
[169] Balkin (n 74 above).
[170] Access Now 'Open letter to Facebook on violence-inciting speech: Act now to protect Ethiopians' 27 July 2020 *Access Now* https://www.accessnow.org/open-letter-to-facebook-protect-ethiopians/ (accessed 15 April 2021).
[171] Flew *et al* (n 102 above) 46.
[172] Land (n 6 above) 409-410.
[173] MK Land & RJ Hamilton 'Beyond takedown: Expanding the toolkit for responding to online hate' in P Dojčinović (ed) *Propaganda, war crimes trials and international law: From cognition to criminality* (2020) 14.
[174] DeNardis (n 4 above).

© University of Pretoria

platform governance as a viable alternative, it should not only be logical but also be demonstrable within the African contexts.

The theoretical aspects take off from Murray's concept of the regulatory matrix for cyberspace. He argued that:

> First, regulators must produce a dynamic model of the regulatory matrix surrounding the action they wish to regulate (including a map of the communications networks already in place). From this they may design a regulatory intervention intended to harnesses the natural communications flow by offering to the subsystems, or nodes, within the matrix, a positive communication that encourages them to support the regulatory intervention. Finally, they must monitor the feedback that follows this intervention. If the intervention is initially unsuccessful they should consider modifying it slightly and continuing to monitor the feedback in the hope of producing constant improvements. If successful, the positive feedback generated will reinforce the regulatory intervention, making it much more likely to succeed.[175]

Here, Murray's matrix rests on three major principles:

a. Regulators working on a dynamic model
b. Opening up mode of communication between subsystems
c. Using the feedback loop to improve the dynamic model

The nature of governance that would advance a complex interest such as speech, does not only need to be dynamic in nature, but also allow for more practical ideas through diversity of experiences. It is neither governments' sole responsibility nor is it just a social media platform-driven initiative – everyone must be able to contribute to its development. While this model can be adopted anywhere across the globe, what makes African countries' case unique is that there is a history of colonial laws that have breathed into current cyber laws.[176] Such dynamism should include the need to rethink the foundations of laws that bear on the right to not only speech, but expression in the online context.[177] Not only would these laws and policies be required to be reviewed in line with international human rights law, their essence for facilitating democratic development must also be ensured.

Considering the proliferation of information disorder online today, it is going to be impracticable to effectively prevent all of it, mostly due to its dynamic and evolving nature. In order to effectively prevent information disorder online, regulatory solutions must not only be dynamic, but rooted in the human rights approach. For example, the point of departure should not be the enactment of laws but rather, the last step only when arguments for such policies are based on the human rights approach. This is because information disorder is not a basis for limiting expression, especially as protected under international human rights law. Even though they could be unpopular,

---

[175] Murray (n 65 above) 250.
[176] Section 2.5 above.
[177] WJ Vollenhoven 'The right to freedom of expression: The mother of our democracy' (2015) 18 *Potchefstroom Electronic Law Journal* 2302.

© University of Pretoria

they should still be free until found violent.[178] Sharing unpopular views does not mean being pro-violence or pro-harm and therefore should not be the basis of limiting a right as important as freedom of expression. As argued by Popper, conjectures and refutations are some of the surest ways a society advances itself through expression. Particularly, the conjectures – opinions based on incomplete information are not only regarded as necessary for development but also demonstrates how systems learn.[179] Therefore, that a speech or content is unjustified, unjustifiable, assumed or incomplete should not be the motivation to limit it. Rather, a more scientific approach is subjecting such conjectures to criticisms that are critical – a systematic use of refutations to identify the illogic of such conjectures.

In regulating such a dynamic system, Murray argues that it is grounded in a set of standards that is adaptive to the fleeting changes of several constellations of online harms. At the heart of this, he argues, is communication, which feeds the entire loop of rule-making in the system with experience as feedback. Based on this, an actor, whether state or non-state, that is truly committed to solving a challenge as online harm should be invested in more alternative methods to regulation other than the law. It is these alternative methods like education, proactive disclosure of public information, awareness and sensitisation that a policy on regulation should be placed on. Regulating information disorder with laws is an example of placing something on nothing and expecting it to stand. Therefore, not only must African countries repeal their policies on information disorder, they must commit to ensuring a diverse system that thinks up governance based on the respect for human rights.

In drawing the relationship between the role of governments who have the primary obligation to protect human rights online given the complexity involved with large-scale moderation and feasible governance systems that actively prevents online harms, Land argues three main opportunities: differentiated liability for social media platforms; specificity and guidance; and accountability mechanisms and moderation by design. On differentiated liability, Land argued that strict liability is more appropriate for the regulation of certain content like targeted online violence with higher propensity of offline harm but low risk to freedom of expression. She notes that:

> in this way, intermediary regulation could preserve the freedom of expression of their users while tamping down on the virality of speech that magnifies its harm. Thus, this duty of care would be aimed at transforming online hate speech into something that looks more like its offline equivalent – individuals who make hateful comments in small conversations or shout invectives on the street, whose speech quickly fades into oblivion.[180]

Land's argument on differentiated liability improves the solution towards regulating online harms in two major ways. First, it proposes a deconstruction of the harm level of various online harms in order to effectively regulate them. Secondly, it sets a more purposive alternative method of regulation to traditional platform governance –

[178] M Pohjonen 'A comparative approach to social media extreme speech: Online hate speech as media commentary' (2019) 13 *International Journal of Communication* 3091.
[179] See KR Popper *Conjectures and refutations: The growth of scientific knowledge* (1989) 215.
[180] Land (n 6) 425.

© University of Pretoria

maximising the impacts of traditional governance while also acknowledging its limitations.

With respect to specificity and guidance, Land points out that in order for governments to maximise their protective responsibility of online free speech without violating it, liability should be clear and sufficiently precise. This draws out the need for states to abandon the use of common terms used in the limitation of speech for more direct, clear and sufficient self-explanatory terms that leaves no room for arbitrary construction. With respect to accountability mechanisms, governments are advised to consider a variety of accountability systems, one of which is the use of legislation to ensure more user-focused procedures that assist with ensuring justice and protecting against violations. The third proposal, moderation by design addresses the motives of social media platforms for profit, which are carried out through 'personalized-advertising, algorithm-fuelled, maximized-engagement-at-any-cost business model that has played a large role in creating a poisonous online environment.' According to her, such motives should be critically considered always in favour of protecting human rights principles.

Both Land and Murray offer at least three perspectives to platform governance that are important in the reframing of platform governance that works. First, both see the need to identify the actions that actors need to regulate. Second, both scholars see functional models of platform governance as generative – a series of experiments where feedback improves experience. Third, both argue that in order to achieve governance that works, governments must adjust along with the needs for protecting online speech and this means formulating policies together with more stakeholders.

Therefore, the generative approach to platform governance is the use of policies beyond their traditional role and more inclusion to drive alternative methods of regulations other than criminal or permanent sanctions. For example, such methods may include civil and administrative sanctions by governments and down ranking, filtering, take downs, temporary removal of account or content by social media platforms. An important point to note on the generative approach is that it does not exclude criminal or permanent sanctions, rather, it focuses on other least intrusive means.

It is important to note that such a method may include laws, regulations, and guidelines on specific aspects of online harms with criminal sanctions or permanent sanctions which are only reserved for serious cases of violence. With this approach, the focus is on understanding the dynamics of various speech-related online harms, and making policies that are effective in preventing such harms. The main reason for this is because given the complex nature of communications infrastructure across the world today, isolated, mono-themed and singular approaches to platform governance will not solve platform governance challenges like online harms. For example, a United States' safe harbour and approach to platform governance cannot continue to be neutral under section 230 to the nuances of the global environment. This is because its 2016 elections were alleged to have been marred by foreign manipulation from

© University of Pretoria

Russia using various means. By that token, the Ethiopian government cannot protect the right to freedom of expression online by criminalising disinformation as expressions are not limited to only true statements.[181]

As Murray's regulatory matrix suggests, communication is the most important aspect of such a regulatory approach. Such alternative methods are also social in nature in that they look to build a foundational basis upon which legitimate claims of criminal sanctions and permanent sanctions can be made on a society. This is even more pertinent given that most actors are regarded as being complicit in engendering online harms. Putting it simply, laws and policies must provide more access to content online and according to the three-part test limitation under international human rights law. That way, it would earn the trust of the society on its commitment to the right to freedom of expression rather than censor content online with the full blunt force of the law.

A practical example of this approach is seen in the paper by BSR on a human rights-based approach to content governance discussed above. In looking to close that gap between context and scale that is the main challenge for regulating social media companies even for international human rights law,[182] the paper proposes how the law can address such a challenge. The BSR paper demonstrates the generative approach in two major ways. It is an alternative method of regulating online harms and requires multi-layered level of inputs (communication) all through the steps of its application. The BSR recognised not only the power, but the responsibility for such power in developing such alternative methods.[183]

In demonstrating the need for communication as the most constant value in the course of building such approach which foregrounds Popper's thoughts on conjectures and refutations, the BSR stated that 'the paper has been written to inform discussion and debate, and we welcome comments to amend, improve, and build on this approach.' Admitting the limitations of the proposal and identifying the need for more development of the approach, the BSR added 'as these limitations indicate, the content governance debate has some distance still to travel, and there are many elements in need of deeper exploration.' In order to build towards a consensus-based platform governance through the human rights approach, the BSR suggests that:

> Stakeholder-inclusive approaches should be taken throughout all elements of content governance. A human rights-based approach implies placing the interests of those whose rights are affected at the center, and for this reason, it is essential that platform policies are developed through meaningful consultation. In the social media industry, user "personas" are often created to inform policy and product development, and from a human right point of view, it is important that these personas are drawn from a range of different vulnerable groups.[184]

[181] Joint Declaration (n 166 above).
[182] Keller & Haggart (n 1 above).
[183] BSR (n 152 above) 10.
[184] BSR (n 152 above) 9.

© University of Pretoria

Reimagining such governance as a new form of constitutionalism, Suzor described the kind of platform governance that would work as such that it requires some imagination and invention especially with a diverse set of actors.[185] In imagining such an invention, it is important to redistribute decision-making powers that are underpinned by human rights but informed by the dynamism required to adapt to the complex challenge of online speech governance. The ultimate goal of the redistribution is to ensure a generative system of norms towards such a form of platform governance where every stakeholder is able to shape the future of online speech and reduce online harms.

Douek makes a stronger argument for the generative approach when she noted that the:

> fundamental issue of "We need a way to think about content on the Internet and the rules that govern them"— I don't think that's ever going away, and so we need to start developing the norms around that. We need to acknowledge that we are balancing societal interests, we are going to have error rates. Once we can even be on the same page about that, that's when we can start having a proper conversation about what we're doing.[186]

This is in agreement with what Suzor reckons would be the role of international human rights law – to form the anchor upon which such inventions and new norms would mean for protecting online speech from harm.[187] However, Murray identifies one major flaw in the generative approach is that accepting it is accepting the limits of our knowledge.[188] This therefore presents an opportunity to consider an experiment currently underway, which is led by ARTICLE 19, a frontline civil society organisation whose mission is to 'work for a world where all people everywhere can freely express themselves and actively engage in public life without fear of discrimination.'[189]

4.6.1 Applied generative model of platform governance

In 2019, ARTICLE 19 began a consultation on its recent proposal on content moderation governance based on international human rights law. The proposed governance framework is referred to as Social Media Councils (SMC), which has since received the backing of the UN Special Rapporteur on Promotion and Protection of the Right to Freedom of Expression.[190] According to its consultation paper, the SMCs seeks to be a mechanism that is open, transparent, accountable and participatory to address content moderation based on international human rights standards on freedom of expression.[191]

---

[185] Suzor (n 1 above) 3.
[186] G Edelman 'On social media, American-style free speech is dead' 27 April 2021 *Wired* https://www.wired.com/story/on-social-media-american-style-free-speech-is-dead/ (accessed 29 April 2021).
[187] Suzor (n 1 above) 168, 171.
[188] Murray (n 65 above) 257.
[189] ARTICLE 19 'Our mission' https://www.article19.org/about-us/ (accessed 15 June 2020).
[190] United Nations (n 135) para 58.
[191] ARTICLE 19 'Social Media Councils: Consultations' 11 June 2019 https://www.article19.org/resources/social-media-councils-consultation/ (accessed 15 June 2020).

© University of Pretoria

What the SMCs' model offers is a non-binding, soft law and generative governance framework where 'the participating social media companies would commit to executing the Council's decisions in good faith.' It also confines itself to non-pecuniary remedies such as 'a right of reply, the publication of an apology (if, for instance, some content was removed by mistake), the publication of a decision, or the re-upload of suppressed content.' In its overall motive, it states that, 'the voluntary approach we advocate for in the SMC project is only intended to focus on the accessibility, visibility and findability of content on social media platforms.'

The SMCs model proposes two core features which are the role of the SMC and national laws and participation. The SMCs propose not to be involved with reviewing government requests with respect to online content. Rather, this would be done indirectly by reviewing the decisions of social media companies based on international human rights law. It intends to leave the direct governance of speech by national governments to their impartial and independent judicial system who also decide based on the three-tier test. On the nature of speech that clashes due to changing contexts, the proposal suggests that balance would have to be struck between margins of appreciation in such context, community guidelines of social media platforms and international human rights standards. In terms of participation in the SMCs, the proposal suggests a 'multistakeholder' approach that is driven by actual decision makers that are transparent, provide public access, rely on consensus in their outcomes and creates a win-win for a broader spectrum of stakeholders.[192]

It also highlights the conditions that would make the SMCs' effective which are that they are independent, consultative, democratic, representative, robust complaint mechanism and accountability. In addition to this and as a basis for its consultations, it focuses on five core aspects for its governance which includes substantive standards; functions of the SMCs; territorial jurisdiction; subject-matter jurisdiction and other technical questions. On substantive standards, it proposes two major avenues which include the direct application of international human rights standards and adoption of a code of human rights principles. The challenge with the first proposition is that contexts temper free speech and therefore vary in their mode of compliance with international human rights standards. In addition, the basis for such variance would be a challenge if the standards are directly applied. The adoption of a code seems more feasible as it is not only more specific to online speech governance, it also zooms in on various aspects of content governance like hate speech, online violence against children, misinformation and others.[193]

In terms of its structure, it proposes whether the SMCs should be an advisory body, an appeal mechanism or both. As an advisory body, it would provide guidance on social media platforms' twin constitutions, Terms of Service and Community

[192] ARTICLE 19 (n 191 above).
[193] ARTICLE 19 'Social Media Councils: Consultation paper' June 2019 25 https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf (accessed 16 June 2019).

© University of Pretoria

Guidelines based on international human rights law. As an appeal mechanism, it would adjudicate disputes which social media companies would be bound to follow and would be open and accessible to all. It also proposes both structures. Considering a generative approach, both structures can be adopted in that the SMCs can actively shape the 'law and norms' that guide the social media companies based on international human rights law while also providing the avenue to attain it – interpret these 'laws and norms' as an appeal mechanism.

It proposes that in terms of territorial jurisdiction, the SMCs could be focused at any of the global, regional or national levels. Highlighting the various strengths of each level, the proposal suggests that any of the three levels could be useful in attaining the goals of effective platform governance. The global level could be useful for its possibility to apply universal norms, the national level has strengths in experts who not only understand the subject matter but also know the socio-cultural, socio-political and historical perspectives to online speech. Its proposal on having a regional system also has strength in that it allows for more inclusion of civil society organisations in the process. It considers a possibility between the global and national levels of the SMCs. For a generative approach, it could combine the three tiers in its governance. That way, the weaknesses of each of the tiers is overcome by the strengths of the others with which it is combined.

With respect to subject-matter jurisdiction, it could focus on either a more general scope of content moderation to include hate speech, privacy, the use of frontier technologies on platforms or on a more specialised scope like information disorder or targeted online violence with an opportunity to broaden its scope in the future. The latter approach for specialised scope agrees more with the generative approach in that through a diverse and improved communications system, the narrow scope would be focused on, improved upon, and its experience will be used to design other areas. On the other hand, should the SMCs take on a general scope, it might take more than it could chew and therefore have stakeholders lose interest in the process.

Lastly on technical issues, the SMC model proposes that its membership will be made up of governments, civil society, academics, large social media platforms, journalists, media houses, the advertising industry and other proximate stakeholders. It particularly notes with respect to government participation that 'governments might be associated with the creation of the SMC as observers, with no decision-making power. This participation might contribute to alleviating concerns among policymakers.' The SMC would be bound by 'the Charter' which will provide for the mandate; composition, powers and roles of its organs; rules of procedure; basis of authority; functions; and commitments of members. Each of these aspects evince the generative approach in that it puts the shaping of the SMC model in the hands of the public like a truly democratic process.

However, both the BSR paper and the SMC model need to involve more strategic state actors. Any model that would ensure effective governance of social media platforms must include as many relevant actors as possible. Taking African countries

© University of Pretoria

for example, the laws have to change to apply multistakeholderism. Using Nigeria as a case study, it will be practically impossible to apply the BSR paper and the SMC's model due to problematic legal and regulatory provisions that violate the right to freedom of expression online.[194] State actors play a key role in reforming these provisions. Not including state actors in such a context would be replicating systems that do not understand these contexts but make policies for these contexts. The position that governments or social media platforms might negatively affect these ideas and as a result, their involvement should be played down might be counter-productive in the African context. While this position is understandable because of governments' censorious disposition to free expression, it would be necessary to think of state actors beyond the traditional actors like the executive, legislature and the judiciary to include National Human Rights Institutions (NHRIs) in applying the BSR paper and the SMC model.

Considering this background and the various forms of governance already analysed above, the generative approach provides an opportunity to reframe the responsibilities of both state and non-state actors in the prevention of online harms through four major ways. First, both state and non-state actors should adopt a model that combines all forms of governance which caters to various interests. Governments, social media platforms, academics, civil society and most importantly users are designed to be part of the consultative process. Second, both categories of actors should encourage an iteration towards a system that is generative – 'open to the participation of a broad range of stakeholders' and incremental processes towards norm developments.'[195] Third, both actors should re-commit to international human rights standards as the basis for such a model. Fourth, each of the actors should work with the goal that in adapting the generative approach to various contexts, there are bound to be failures but more importantly, they are opportunities for lessons and re-evaluation. Therefore, not only does the generative approach offer a promising opportunity for African governments in preventing online harms and protecting online speech, a more inclusive approach to the BSR paper and the SMC model presents an opportunity to test such an approach.

4.6.2 Justifications for a generative model of platform governance

There are a number of factors that should motivate African governments to adopt a generative approach to platform governance. These factors foreground the importance of protecting online speech in the age of digital democracy in Africa. Therefore, while platform governance may be dominated by few Western social media companies and countries, there is a need for African countries to optimise their systems to make the most of it. In thinking through a more global approach beyond Western institutions to platform governance, Bloch-Wehba has argued that addressing platform governance through global governance assists in expanding the meaning of

[194] Section 5.2.1 H.
[195] ARTICLE 19 (n 191 above) 9.

© University of Pretoria

traditional concepts of governance and crowdsourcing for solutions on legitimacy and accountability.[196]

By adopting the generative approach to platform governance, it gives ample opportunities for African countries to understudy the peculiar challenges of online harms and use the international human rights system to design dynamic systems to reduce them. Two examples of such dynamic systems are the BSR's paper for social media platforms on content moderation and the SMCs' proposed framework. Not only do both systems improve and bring the protection of online speech closer to reality, they could help reduce online harms if not completely prevent them. For example, using the various provisions under the African Declaration,[197] African countries can design alternative methods to prevent online harms through digital literacy, education on the dangers of online harms, mainstreaming online speech protection into public policies, encouraging more diverse stakeholders in the shaping of their soft and hard laws that impact on the right to freedom of expression online.

Asides the clear connection between the BSR paper, SMCs and how online speech can be effectively governed based on international human rights law, both the BSR paper and SMCs' model seems to draw their strength from the annual report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression.[198] The report which focuses on online content regulation provides a basis upon which both state and non-state actors can deploy international human rights law.[199] Referring to a statement by the founder of Facebook on the process of scaling different contexts and the need for compliance with international human rights law than national laws, the report notes that the:

> process, and the relevant standards, can be found in human rights law. Private norms, which vary according to each company's business model and vague assertions of community interests, have created unstable, unpredictable and unsafe environments for users and intensified government scrutiny. National laws are inappropriate for companies that seek common norms for their geographically and culturally diverse user base. But human rights standards, if implemented transparently and consistently with meaningful user and civil society input, provide a framework for holding both States and companies accountable to users across national borders.[200]

The same report also elaborates on specific areas that social media platforms could focus on when applying such principles. Such areas include the standards for content moderation, responses to government requests, rule-making and product development, rule enforcement, and decisional transparency. Each of these areas have either been addressed by the BSR or the proposed frameworks of the SMCs therefore providing a fertile ground for a generative approach to platform governance to grow.

---

[196] Bloch-Wehba (n 128 above) 70.
[197] Section 2.4.3 above.
[198] United Nations (n 135 above).
[199] United Nations (n 135 above) paras 41-43.
[200] United Nations (n 135) para 41.

© University of Pretoria

Given the exponential increase in Internet access in Africa in the last few decades, Africa might be the next frontier for the Internet.[201] What this means is that not only will there be more opportunities for economic development, there will be a more dire need to ensure the integrity of online communications. In doing that, actors do not only need to mainstream a human rights approach to online speech, they should start now to minimise the error rates and improve the iteration of an effective platform governance in Africa. Elections are already being shaped by social media platforms as much as social power is also beginning to have new meanings. The growth of the Internet in Africa and as a result an increase in social media adoption presents the urgency for approaching the latter's governance with more patience than fear. This is because social media platforms play an important role in institutionalising the digital civic space and are therefore an important stakeholder in Africa's democratic development. Taking this responsibility seriously requires social media platforms to rethink their various structures, privileges and powers in order to design people-focused rules underpinned by international human rights law.

Given the nature of online harms with how they change based on contexts and are regulated through the wrong frameworks, considering a generative approach to platform governance would give opportunities to various stakeholders to invent effective solutions. For example, as suggested above, stakeholders could map a number of online harms and decide based on the three-part test which of them would be regarded as legal and illegal content in various contexts. This would then see to designing policies to that effect and as a result ensure prevention of online harms on social media platforms.

4.7 Conclusion

This chapter set out to show the way a rights-respecting platform governance can be used to prevent online harms and protect freedom of expression online. In answering the question, it found that governments, social media platforms and civil society seek to govern online speech for various reasons. Governments do so in order to protect public interests but end up violating the right to freedom of expression online instead. Social media platforms also regulate online content in order to keep their platforms safe and generate revenue while the civil society often engages in online speech governance in order to bring it in line with international human rights standards.

These analyses also showed that no single actor can effectively regulate social media platforms in that the power distribution has been made more complex due to the Internet, the basis for platforms as a distributed architecture. Therefore, it considered the argument as to whether a human rights approach to platform governance is feasible especially in African countries. It found that not only is such an approach

[201] M Tuerk 'Africa is the next frontier for the Internet' 9 June 2020 *Forbes* https://www.forbes.com/sites/miriamtuerk/2020/06/09/africa-is-the-next-frontier-for-the-internet/?sh=f1e169749001 (accessed 10 August 2021).

© University of Pretoria

possible, it could be complemented with an approach that is open to the redistribution of powers referred to as the generative approach to platform governance.

It suggests that both state and non-state actors can perform the necessary responsibilities of preventing online harms if this approach is utilised. It urged stakeholders like government, civil society, social media platforms and other stakeholders to consider a more inclusive system in governing online speech, commit to international human rights standards on protecting the right to freedom of expression online and other human rights and should be adaptable to contexts. These could be generatively achieved given the BSR's paper and the proposed SMCs' model. The next chapter examines the role of actors in applying this generative model in Nigeria and South Africa.

© University of Pretoria

CHAPTER FIVE: THE ROLES OF STATE AND NON-STATE ACTORS IN ENSURING A RIGHTS-RESPECTING APPROACH TO PLATFORM GOVERNANCE IN NIGERIA AND SOUTH AFRICA

5.1 Introduction

The previous chapter focused on the various aspects of platform governance and their possibilities of preventing online harms and promoting the right to freedom of expression online. It discussed the historical basis of platform governance through internet governance, the meaning of platforms, their approaches and forms. Understanding that platform governance is nascent and not fool-proof, it considered limitations that might militate against an effective platform governance model especially in African countries. To circumvent these limitations, the chapter considered a human rights approach to platform governance. It found that such an approach, while desirable, requires a more practical application to be effective in African countries. By combining regulatory approaches, it proposes a generative approach to platform governance. Such an approach would be iterative, dynamic and rights-respecting especially to prevent online harms. For a more practical application of what such would look like, it considered the BSR report on a human rights approach to content governance and the SMC model.

In applying this generative approach in light of African contexts, this chapter examines the roles of both internal and external state and non-state actors in regulating online harms through a rights-respecting approach in Nigeria and South Africa and how a generative approach can be applied. In examining these roles, the major objective of this chapter is further broken down into three sub-questions:

a. What are the legal and regulatory provisions of online harms in Nigeria and South Africa?
b. In what ways, can these legal and regulatory provisions be brought in line with international human rights standards?
c. What are the roles of state and non-state actors, internally and externally in ensuring an effective rights-respecting approach to platform governance in order to prevent online harms and protect the right to freedom of expression online?

In attempting responses to these questions, this chapter is divided into four parts. The first part links the chapter to the previous ones, lays out its objectives and how it contributes to the thesis as a whole. The second part considers the legal and regulatory provisions on online harms in Nigeria and South Africa. This part assesses these provisions based on international human rights standards. It also carries out a micro-analysis of both countries' laws and their provisions with respect to online

© University of Pretoria

harms. The third part of this chapter focuses on framing the roles of state and non-state actors in ensuring a rights-respecting platform governance in Nigeria and South Africa. It considers the application of the human rights-based approach to platform governance in both countries and the roles each actor in the governance ecosystem must play in preventing online harms and promoting the right to freedom of expression online. The fourth part concludes that social media platform governance in Nigeria and South Africa is possible. However, for it to be possible, it must be anchored on international human rights law and operationalised by various actors performing specific responsibilities in ensuring that online harms are prevented and the right to freedom of expression online is protected in both countries.

5.2 An international human rights law analysis of the legal and regulatory provisions of online harms in Nigeria and South Africa

This part considers the legal and regulatory provisions of online harms in Nigeria and South Africa. In considering these harms, their assessments have to be based on at least one obligatory human rights standard.[1] This standard-based assessment is necessary in order to measure the impacts of online harms on the right to freedom of expression in order to objectively arrive at practical solutions. The two major standards to be considered within the context of this thesis are the constitutions of Nigeria and South Africa (constitutionalism) and international human rights law (internationalism) provisions that impact on the right to freedom of expression online in Nigeria and South Africa.[2] Even though both standards have their weaknesses and strengths, as argued by Benvenisti and Harel, the focus should not be on whether any of the standards are desirable, but it should focus on which of them is most effective.[3]

For example, the weaknesses of constitutions may include possible state subversion of the constitutional powers and the tyranny of the majority.[4] The strengths of constitutions may also lie in the fact that they are often the most proximate and primary legal document that regulate power relations in a democratic state and they are the means through which international human rights law can be applied in national contexts.[5] For international human rights law, its main weakness is that its

[1] Open Rights Group 'ORG policy responses to Online Harms White Paper', May 2019 https://modx.openrightsgroup.org/assets/files/reports/report_pdfs/ORG_Policy_Lines_Online_Harms_WP.pdf (accessed 13 July 2021).

[2] 'Individual human rights are secured by at least two different legal sources: constitutional law and international law.' See E Benvenisti & A Harel 'Embracing the tension between national and international human rights law: The case for discordant parity' (2017) 15 *International Journal of Constitutional Law* 37.

[3] Benvenisti & Harel (n 2 above) 51-57. See Section 2.2 above on Silungwe's description of African legal theory which postulates legal pluralism as the law that best serves the end of justice.

[4] Benvenisti & Harel (n 2 above) 50.

[5] Benvenisti & Harel (n 2 above) 48.

© University of Pretoria

implementation often involves a complex web of actors.[6] An additional challenge is that international human rights law is often perceived as aspirational, proliferated and lacking enforcement.[7] In terms of its strengths, international human rights law often provides anchor with respect to interpretation of human rights standards for states and it also derives its legitimacy from states.[8]

Both Benvenisti and Harel further referred to the need to embrace the tension between international human rights law and constitutions as the 'discordant parity paradigm'. The discordant parity paradigm is the choice of standard made by state actors in order to ensure the promotion and protection of human rights, which is their primary responsibility, that is they must consider the standard that is most robust and serves the end of human rights development.[9]

In this case, this part, along with this thesis, argues that in preventing online harms and protecting the right to freedom of expression online through a rights-respecting platform governance in Nigeria and South Africa, more attention needs to be paid to internationalism by various actors. This is because national laws, including those made by African governments, often play catch up with technologies.[10] As a result of this slow development, there is limited analysis that examines the human rights impacts of technology-related laws particularly on how they relate to the right to freedom of expression online. However, as discussed in the second chapter of this thesis, there are ample examples of how international human rights standards could fill in the gaps for legal frameworks that respect, promote and protect the right to freedom of expression online.[11] This standard provides that while States are allowed to limit the right, they are only allowed to do so within set parameters of the four-part test for limitations which may be said to be provided for under Nigeria and South Africa's constitutions.[12]

---

[6] F Viljoen 'Contemporary challenges to international human rights law and the role of human rights education' (2011) 44 *De Jure* 209-220.

[7] Benvenisti & Harel (n 2 above) 46-47; Viljoen (n 6 above).

[8] Benvenisti & Harel (n 2 above) 41-46; F Viljoen *International human rights law in Africa* (2012) 34, 146.

[9] As above. This argument by Benvenisti & Harel compares with Principle 4 of the revised Declaration of Principles on Freedom of Expression and Access to Information in Africa on the 'Most Favourable Principle.'

[10] T Timan *et al* 'Surveillance theory and its implications for law' in R Brownsword, E Scotford & K Yeung (eds) *The Oxford handbook of law, regulation and technology* (2017) 749.

[11] Section 2.4 above.

[12] Constitutionalism can also be desirable for a rights-respecting platform governance in two major ways. One, as the most organic law in most African systems, the fundamental rights provided for in constitutions are most likely the first port of call with respect to the protection of online expression. Therefore, constitutionalism in this regard plays a key role in receiving international human rights law developments into local contexts. In this instance, such a relationship creates a value chain where constitutions play the 'middleman' role between international human rights law as the 'wholesaler' and the specific laws and individual rights as the 'retailer.' Two, constitutions carry the force and certitude

© University of Pretoria

For example, section 39 of the 1999 Nigerian Constitution (as amended) provides for the right to freedom of expression, opinions and information as follows:

> 1. Every person shall be entitled to freedom of expression, including freedom to hold opinions and to receive and impart ideas and information without interference.
> (2) Without prejudice to the generality of subsection (1) of this section, every person shall be entitled to own, establish and operate any medium for the dissemination of information, ideas and opinions:
> Provided that no person, other than the Government of the Federation or of a State or any other person or body authorised by the President on the fulfilment of conditions laid down by an Act of the National Assembly, shall own, establish or operate a television or wireless broadcasting station for, any purpose whatsoever.

While its subsections (3), (a) & (b) also provides for internal limitations on the right to freedom of expression as follows:

> (3) Nothing in this section shall invalidate any law that is reasonably justifiable in a democratic society –
> (a) for the purpose of preventing the disclosure. of information received in confidence, maintaining the authority and independence of courts or regulating telephony, wireless broadcasting, television or the exhibition of cinematograph films; or
> (b) imposing restrictions upon persons holding office under the Government of the Federation or of a State, members of the armed forces of the Federation or members of the Nigeria Police Force or other Government security services or agencies established by law.

Section 45(1) then provides for external limitation on the right to include the protection of the rights of others and the need to protect public interest as follows:

> 45. (1) Nothing in sections 37, 38, 39, 40 and 41 of this Constitution shall invalidate any law that is reasonably justifiable in a democratic society
> (a) in the interest of defence, public safety, public order, public morality or public health; or
> (b) for the purpose of protecting the rights and freedom or other persons…

South Africa also has similar provisions on internal and external limitations. Section 16(2) of the Constitution of South Africa, 1996 provides for internal limitations in that the right does not include 'propaganda for war; incitement to imminent violence; advocacy based on hatred is based on race, ethnicity, gender or religion, and that constitutes incitement to cause harm.' An external limitation of the right is provided for under section 36 of the Constitution and requires that limitation of the rights, including the right to freedom of expression must state the nature of the right, the legitimacy of

---

of responsibilities – it is the most fundamentally enforceable legal document available for human rights protection. Therefore, while a rights-respecting law on platform governance must be inspired by international human rights standards, its enforcement should be anchored to the fundamental rights provided for under the constitutions.

© University of Pretoria

limitation; the nature and extent of such limitation; the relationship between the limitation and the purpose, and the least intrusive means to achieve such purpose.

It is important to note that Nigeria adopts a dualist approach with respect to domestication of international human rights law while South Africa adopts both dualist and monist approaches to international human rights law.[13] For example, section 12 of the Constitution requires the National Assembly to pass a law to give effect to international law treaties in Nigeria. For South Africa, while section 231 of the Constitution allows for direct application of customary international law (monist approach), section 232 requires the Parliament to enact laws to give effect to treaties made under international law (dualist approach).

However, given the normative advancement on online speech governance when compared to constitutional developments on the right to freedom of expression, as explained in previous chapters and shown in the latter parts of this chapter, international human rights law ought to apply as a standard in both countries. This is because international human rights law offers more normative clarity and protection for the right to freedom of expression. Such application is not to create a hierarchical feud but it is due to an urgent need for normative and operational clarity given the novelty of platform governance.[14] In making a case for internationalism, this thesis advances five major reasons to consider it over constitutionalism.

One, due to the challenge that most state actors would rather be guided by their local laws than international human rights law, both countries' constitutions have not fully benefited from new developments on the application of the right to freedom of expression online.[15] This challenge in many African countries is due to colonial legal legacies and the misapplications of the permissible limitations under international human rights law through various laws.[16] It is these legacies and misapplications that misdirect most laws made to limit the right to freedom of expression online.[17] Therefore, when it comes to protecting freedom of expression online, actors cannot afford to look only inwardly. They need to draw inspiration from international human rights law to reform laws that prevent online harms and promote online free speech

[13] M Killander & H Adjolohoun 'International law and domestic human rights litigation in Africa: An introduction' in M Killander *International law and domestic human rights litigation in Africa* (2010) 4.
[14] On the novelty of the regulation of platforms see R Gorwa 'The platform governance triangle: Conceptualising the informal regulation of online content' (2019) 8 *Journal on Internet Regulation* 4.
[15] Section 2.4 above.
[16] M Kanna 'Furthering decolonization: Judicial review of colonial criminal laws' (2020) 70 *Duke Law Journal* 412, 452; Media Defence 'Mapping digital rights and online freedom of expression litigation in East, West and Southern Africa' August 2021 https://www.mediadefence.org/resource-hub/wp-content/uploads/sites/3/2021/08/Media-Defence-Mapping-digital-rights.pdf (accessed 30 October 2021); Section 2.5 above.
[17] Section 2.5.1 above.

© University of Pretoria

because it provides clearer directions through its topical normative advancements that states can follow on novel issues such as platform governance.[18]

Two, international human rights law benefits from nuanced contexts for resources that could assist in framing its approach to platform governance while the constitutions of Nigeria and South Africa are likely to focus on the narrow and limited interpretations of these rights.[19] Three, constitutions have to go through rigorous amendment processes or lengthy court cases before constitutional protections or judicial precedent may be set on online harms and platform governance. Four, many technology-related laws in Nigeria and South Africa like those that impact online harms are framed to be extra-territorial while lacking in proper enforcement and, in many instances, they are historically and presently posed to violate human rights and protect the state and not its citizens.[20] Therefore, it would be important to measure these laws against international human rights law which is globally applicable and enforceable.[21] Five, given the gaps that exists in most constitutional interpretations of online expression, applying international human rights law to local contexts has the potential of improving not only policy development on the issue, but also presents an opportunity to inform judicial activism for human rights protection, since both Nigeria and South Africa already have obligations to comply with international human rights law. Given the background above, this following part discusses the various laws and their relationship with online harms in Nigeria and South Africa, and whether they comply with international human rights law.

5.2.1 Legal and regulatory provisions on online harms in Nigeria

Nigeria's 1999 Nigerian Constitution (as amended) is the most primary law of the land and its chapter four provides for fundamental human rights. Section 39 and 45 of the Constitution provides for the right to freedom of expression, opinions and information and their permissible limitations.[22] As at the time of writing this thesis, there is no primary and comprehensive law that limits the right based on online harms. This may be due to two reasons.

---

[18] As above.

[19] G Karekwaivanane *et al* 'Digital rights in closing civic space: lessons from ten African countries' in T Roberts (ed) *Institute of Development Studies* February 2021 https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/15964/Digital_Rights_in_Closing_Civic_Space_Lessons_from_Ten_African_Countries.pdf?sequence=4&isAllowed=y 114, 122, 153, 164 (accessed 23 May 2021).

[20] Section 2.5 above.

[21] See R Wilde 'The extraterritorial application of international human rights law on civil and political rights' (2013) in N Rodley & S Sheeran (eds) *Routledge handbook of international human rights law* 635 – 661; Z Elkins 'Getting to rights: Treaty ratification, constitutional convergence and human rights practice' (2013) 54 *Harvard International Law Journal* 201-234.

[22] Section 39(3) and Section 45 of the 1999 Constitution (as amended).

© University of Pretoria

One, the conversations on online harms and platform governance are fairly new and mostly novel in many countries. Two, provisions on online harms, where found in other laws, are fragmented. Despite these reasons, there is a need to prevent these harms and protect the right to freedom of expression online in order to develop an effective platform governance system. The aim of this part is to consider how various online harms are provided for and whether they are up to international human rights standards. This is because as noted earlier, Nigeria's obligations to protect human rights may be broadly divided into two. First, the Nigerian Constitution requires, like most constitutions, the guarantee and enforcement of fundamental human rights.[23] Second, Nigeria has the obligation to comply with the international human rights it has ratified or acceded to.[24] These situates Nigeria within the constitutionalist and internationalist standards with respect to human rights protection.[25] Therefore, Nigeria has the dual obligation of ensuring that its laws and proposed laws are not only compliant with its Constitution, but that they are also not in violation of international human rights standards.

A Criminal Code and the Penal Code

The Criminal Code Act and the Penal Code are the two laws that provide for substantive criminal offences in Nigeria.[26] The Criminal Code applies to the Southern parts of Nigeria while the Penal Code applies to the Northern parts of Nigeria. Both laws were put in place during the colonial era and most of its provisions, especially those that impact on the right to freedom of expression are still in force till today.[27] Some of these provisions include publication of false news likely to cause fear and alarm; sedition; criminal defamation; and abusive and insulting language.

i. Publication of false news likely to cause fear and alarm

Sections 59(1) of the Criminal Code and section 418 of the Penal Code provide for the offence of 'the publication of false news likely to cause fear and alarm'. Specifically, they criminalise the publication and reproduction of any statement, rumour or report that could alarm public peace. It is punishable by a jail term of three years and categorised as a misdemeanour. Section 59(2) also adds that any person who

[23] Chapter four of the 1999 Constitution (as amended); JA Dada 'Human rights protection in Nigeria: the past, the present and goals for role actors for the future' (2013) 14 *Journal of Law, Policy and Globalisation* 4.

[24] Nigeria is party to major international human rights law instruments and is bound by them. Ratification status for Nigeria https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?CountryID=127&Lang=EN (accessed 12 July 2021).

[25] EO Okebukola 'The application of international law in Nigeria and the façade of dualism' (2020) 11 *Nnamdi Azikiwe University Journal of International Law* 21-27.

[26] Criminal Code Act; Penal Code.

[27] Section 2.5.1 above.

© University of Pretoria

publishes or reproduces such false information cannot claim that they did not know the information was false and therefore would be liable for sharing it. The only way they claim so, according to the provision, is where they show that they took reasonable measures to verify the information.

In the language of online harms, section 59 and 418 of both Codes will qualify as a regulation of information disorder. Criminalisation of reproduction of false information described under the provision would fall under regulation of misinformation where a person shares false information but without the intent to deceive.[28] In terms of regulating disinformation and propaganda, publication of such information may satisfy the requirement of the intent to deceive. As earlier noted in the previous chapter, while information disorder as covered by the provisions of sections 59 and 418 of both Codes may be harmful, under international human rights law, they are not illegal.[29]

In addition to this, the criminalisation of information disorder especially as it relates to these provisions violates online free expression. This is because the legitimate aim of public order being sought to be protected is still not achieved by criminalisation, rather, it further puts online speech in danger.[30] This criminalisation does not factor in that false information tends to spread faster than the true statements especially when it comes to public information due to the elements of information disorder.[31] Additionally, such spread of information disorder is often due to the failure of the government to prioritise proactive access to public information.[32] There are reduced chances of false information when the facts are not only set straight but are accessible in a timely manner. This provision may also be used to restrict unpopular speech online which is a protected form of speech under international human rights law especially as they are not often most palatable public speech.[33] This is because online political speech often stands a chance of stoking emotions and spreads faster when they are not outrightly

[28] Section 3.3.1 A i above.
[29] Section 3.3.1 above.
[30] S Coliver 'Commentary on the Johannesburg Principles on National Security, Freedom of Expression and Access to Information' (1999) https://www.right2info.org/exceptions-to-access/resources/publications/CommentaryontheJohannesburgPrinciples.pdf10-11 (accessed 24 July 2021); AO Salau 'Social media and the prohibition of 'false news': Can the free speech jurisprudence of the African Commission on Human and Peoples' Rights provide a litmus test?' (2020) 4 *African Human Rights Yearbook* 231-254; B Maripe 'Freezing the press: Freedom of expression and statutory limitations in Botswana' (2003) 3 *African Human Rights Law Journal* 66-75.
[31] Section 3.3.1 A & B above.
[32] P Cunliffe-Jones *et al* 'Bad law: Legal and regulatory responses to misinformation in sub-Saharan Africa 2016–2020' (2021) in P Cunliffe-Jones *et al* (eds) *Misinformation policy in sub-Saharan Africa: from laws and regulations to media literacy* 99-218.
[33] Pen International 'Blogger arrested over critical posts, held incommunicado' 30 October 2008 https://ifex.org/blogger-arrested-over-critical-posts-held-incommunicado/ (accessed 14 October 2021); Naijagists 'Kemi Olunloyo & Samuel Welson bail application hearing postponed over judge's absence' 24 March 2017 https://naijagists.com/kemi-olunloyo-samuel-welson-bail-application-hearing-postponed-judges-absence/ (accessed 14 October 2021).

© University of Pretoria

shocking, offensive or disturbing.[34] Criminalisation of such speech does not only attack the right to freedom of expression online, it strikes at the very heart of democratisation as criticisms and dissent are some of the most enriching manures for any aspiring open society.[35]

ii. Abusive and insulting language

Section 399 of the Criminal Code criminalises the use of insulting or abusive language that could provoke another to break public peace. The applicable punishment is a fine or a jail term of two years or both. With respect to the Penal Code, section 204 only criminalises insults to a religion and anyone found guilty is liable to a jail term of two years. This section is problematic because while insults could be used to fuel targeted online violence, they do not fall under the category of speech that are prohibited under international human rights law.[36] Whether such speech is directed to a public, private individual or religion, insults are too vague a term to be used as a limit for online speech especially because its meaning depends on context and that they are part of everyday human interaction.[37]

iii. Sedition

Sedition is the criminalisation of communications that may cause disaffection against a government. Sections 50 - 52 of the Criminal Code and sections 416 - 422 of the Penal Code provide for the offence of sedition. The provisions of both Codes may be used in the case of online speech. The offence of sedition has been linked to Nigeria's colonial past as most colonial administrations do not encourage criticisms against their authoritarian and repressive policies.[38] In addition to these provisions, the offence of sedition under section 50(2) in Nigeria is at direct loggerheads with protected online political speech. This is because as described earlier, the relationship between the government and the governed is hinged on eternal vigilance, especially on the part of the latter. Therefore, disaffection against public policies will continuously take many

---

[34]Section 2.4.3 A iii above.

[35] See WE Adjei 'The protection of freedom of expression in Africa: problems of application and interpretation of Article 9 of the African Charter on Human and Peoples' Rights' unpublished PhD thesis, University of Aberdeen, 2012 69; UA Tar 'The challenges of democracy and democratisation in Africa and Middle East' (2010) 3 *Information, Society and Justice* 88; GAI Nwogu 'Democracy: Its meaning and dissenting opinions of the political class in Nigeria: A philosophical approach' (2015) 6 *Journal of Education and Practice* 131-143.

[36] Section 2.4.3 A iv above.

[37] N Ibrahim 'Kano court sentences singer to death for blasphemy' 10 August 2020 *Premium Times* https://www.premiumtimesng.com/news/headlines/407936-kano-court-sentences-singer-to-death-for-blasphemy.html (accessed 13 October 2021).

[38] CW Ogbondah 'Nigerian press under imperialists and dictators: 1903-1985' *Paper presented at the International Division of the AETMC conference at Portland, Oregon, July 2, 1988* https://files.eric.ed.gov/fulltext/ED296319.pdf (accessed 24 July 2021).

© University of Pretoria

forms including shocking, offensive and disturbing speech.[39] Criminalising such an important need, especially when they do not call directly for violence against others, violates international human rights on one hand, and represses democracy on the other.

iv. Criminal defamation

There are two forms of defamation: libel, which deals with the publication of untrue statements which injures the reputation of another person and slander, which deals with speaking such untrue statements to injure the reputation of another person.[40] Both forms of defamation are provided for and criminalised under the Criminal Code[41] and the Penal Code[42] in Nigeria. Both laws also provide for the punishment of two years while the Penal Code provides for the option of fine.[43] These provisions also relate to online harms specifically in the context of information disorder. In instances where the impacts of a false statement may be ascertained to have violated the reputation of another, international human rights law has provided that proportionate civil remedies that are clear, precise, legitimate and necessary should be adopted instead of criminalisation.[44] There is no way a criminal law can effectively limit falsity or prevent it, first due to its frequency and prevalence and second, due to its impact on media freedoms. This is particularly important in that the right to freedom of expression is not limited to true statements. Criminalisation of defamatory statements also has a huge possibility of chilling all forms of online expression.

B Cybercrime (Prohibition, Prevention etc) Act, 2015

The objectives of the Cybercrime (Prohibition, Prevention etc) Act, 2015 (Cybercrime Act) are majorly two: to provide a legal framework to combat cybercrime and for

[39] H Essien 'Installing Twitter seditious under Penal Code of Northern Nigeria, AGF Malami tells Court' 21 September 2021 *Peoples Gazette* https://gazettengr.com/installing-twitter-seditious-under-penal-code-of-northern-nigeria-agf-malami-tells-court/ (accessed 14 October 2021); Committee to Protect Journalists 'Two journalists charged with sedition over presidential jet story' 27 June 2006 https://cpj.org/2006/06/two-journalists-charged-with-sedition-over-preside/ (accessed 12 October 2021).

[40] IJ Udofa 'Right to freedom of expression and the law of defamation in Nigeria' (2013) 2 *International Journal of Advanced Legal Studies and Governance* 75-84; Committee to Protect Journalists 'Nigerian journalists charged with criminal defamation, breach of peace' 29 October 2019 https://cpj.org/2019/10/nigerian-journalists-charged-with-criminal-defamat/ (accessed 15 July 2021).

[41] Section 373 of the Criminal Code Act.

[42] Section 391 of the Penal Code.

[43] Section 375 of the Criminal Code; Sections 392 and 393 of the Penal Code.

[44] United Nations General Assembly 'Disinformation and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/47/25' 13 April 2021 http://undocs.org/en/A/HRC/47/25 (accessed 26 August 2021).

© University of Pretoria

cybersecurity in Nigeria. While the law does not provide any additional interpretation on what it means by cybercrime, a dictionary meaning refers to cybercrime as 'criminal activities carried out by means of computer or the internet.' This thesis is more concerned about the cybercrime aspect of the Act rather than the cybersecurity provisions. These provisions include sections 23, 24 and 26 as they lean more into the former than the latter. These provisions can be largely described as falling into the finer categories of online harms as described in the third chapter of this thesis.[45] So far, there is a gap in analyses on these harms as they relate to platform governance and the Cybercrime Act. Asides this gap, the provisions of the Act are yet to be considered alongside the responsibilities of both state and non-state actors in preventing online harms and protecting the right to freedom of expression online in Nigeria. The provisions of section 23, 24 and 26 address online harms including disinformation, cyberharassment, Online GBV, violence against children and online hate speech.

Section 23 of the Act provides for the offence and punishment of child pornography online in Nigeria. While child pornography online is a specific aspect of CSAM, CSAM is one of the examples of online violence against children. Therefore, child pornography is a subset of online violence against children. The nature of online violence against children is that it is multifarious and includes sharing of content of torture and cruel treatments of children, which could include abusive and degrading actions. Such part-criminalisation of child pornography online is one step towards many right directions. Therefore, the provision has to comply with Nigeria's obligations under the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography (CRC-OP-SC).[46]

The provisions of section 24(1)(a)(b) of the Act covers the offences of cyberstalking, disinformation and online GBV. The provisions of section 24(1)(a)(b) are divided into six subsections, which are made up of definitions of offences, the applicable punishments and the orders the courts may give. Section 24(1)(a) of the Act criminalises the intentional sending of messages through a computer or network that 'is grossly offensive, pornographic or of an indecent, obscene or menacing character.' Section 24(1)(b) also criminalises intentionally sending a message anyone 'knows to be false, for the purpose of causing annoyance, inconvenience, danger, obstruction, insult, injury, criminal intimidation, enmity, hatred, ill will or needless anxiety to another.' The punishment of these offences is a fine of ₦7,000,000.00 (approx. USD17,000.00) or a jail term of three years or both.

[45] Section 3.3 above.
[46] United Nations General Assembly 'Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography (CRC-OP-SC), A/RES/54/263' 18 January 2002 http://undocs.org/en/A/RES/54/263 (accessed: 26 August 2021).

© University of Pretoria

On the legality of the restrictions, while the law defines the offences that are punishable, the offences do not fall under the categories of speech that may be prohibited under international law.[47] In addition to these, words like 'grossly offensive', 'pornography' (except in cases of child pornography) 'indecent', 'obscene', 'insult' and many others as used in both sections are not clear and sufficiently precise enough to pass the legality test.[48] In addition to these vague words, international law does not prohibit speech that shocks, offends or disturbs. Considering section 24(1)(b) in particular which focuses on disinformation online, the right to freedom of expression whether online or offline is not limited to true statements.[49] It has also been demonstrated in the past that the criminalisation of false statements violates online expression.[50] With respect to legitimate aims which might include cyberstalking, the provision does not include the requirement of repeated messaging in its definition of offence.[51] With respect to section 24(1)(a) for example, there is no clear interest of who it seeks to protect as it does not contain who such a message could be sent to.[52] Comparing the provision to section 24(1)(b) on the other hand, such a message must have been sent to 'another' referring to a person. A closer look at the provisions show criminalisation of online speech through censorious language where civil remedies could be applicable.[53] On necessity, the law has neither demonstrated the social desirability of limiting the right to freedom of expression online based on both provisions. The restriction of right to freedom of expression based on 'annoyance', 'insults' etc. which are all words that could be used colloquially, do not carry any specific meaning and therefore cannot be a basis for the restriction of the right.[54] With respect to the test of proportionality, the offences prescribed for annoying or insulting another person online is a three-year jail term and this is not only over-board, it chills the right to freedom of expression online and encourages self-censorship.

Applying the four-part test of international human rights law to these provisions, it appears that they are in violation of the right to freedom of expression online. In addition to this violation, it shows that if cybercrime are criminal activities carried out

[47] Section 2.4.1 A-C above.
[48] United Nations General Assembly 'Contemporary challenges on freedom of expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/71/373' 6 September 2016 http://undocs.org/en/A/71/373 (accessed 26 August 2021) paras 12-16.
[49] D Kaye *Speech police: The global struggle to govern the Internet* (2019) 70.
[50] United Nations General Assembly (n 44 above).
[51] Section 3.3.2 A above.
[52] The provision only states that a message that 'grossly offensive, pornographic or of an indecent, obscene or menacing character or causes any such message or matter to be so sent' is an offence. It does not identify who such could be sent to.
[53] Section 2.4.1 above. Media Foundation for West Africa 'Nigeria's cybercrime law being selectively applied' 9 October 2020 *IFEX* https://ifex.org/nigerias-cybercrime-law-being-selectively-applied/ 24 July 2021.
[54] United Nations General Assembly (n 44 above).

© University of Pretoria

through the use of a computer or network and the offences fabricated by the Act under section 24(1)(a)(b) do not fall under international human rights law directions on the speech that can be criminalised, the provisions pose threats to online expression.

With respect to section 24(2)(a-c) of the Act which prohibits speech or communication that threatens death, kidnapping or harm to property, there seems to be some compliance with international human rights law standards. However, a closer reading of section 24(2)(c) of the Act which includes threats to reputation of an addressee or the reputation of a deceased person is the transplantation of criminal defamation into online space and which is not congruent with international law standards.[55] Applying the four-part test to these provisions show that while the provisions define the offences with some clarity and precision, provides a legitimate basis and it could be necessary in order to protect the rights of others or that of the public, the applicable punishments seem disproportionate in that section 24(2)(c)(i) provides for a jail term of ten years and a minimum fine of ₦25 000 000.00 (approx. USD70 000.00). It is also noteworthy that the provision of 24(2)(c)(ii) does not only criminalise defamation which ought to be settled as a civil matter in terms of damage to reputation with a jail term of five years or ₦15 000 000.00 (approx. USD36 000.00) or both, it also criminalises a non-existent subsubsection (d) not provided for in the Act.

In a general analysis of the provisions, the offence of cyberstalking sought to be criminalised by the provision failed to mention a major ingredient of cyberstalking which includes repeated messaging or sending of intrusive communications.[56] None of the provisions under section 24 provide for such repetition; rather, they provide that any message or communication satisfies the ingredient of cyberstalking. Using this point, it shows that even section 24(2)(a-c) is quite problematic and cannot be said to comply with international human rights law.

The Court of Appeal has had the opportunity to consider the provisions of section 24 of the Cybercrime Act in the case of *Paradigm Initiative & Others v Attorney General of the Federation & Others*.[57] In the *Paradigm Initiative* case, the appellant had prayed that the provisions of sections 24 and 38 of the Cybercrime (Prohibition, Prevention etc) Act of 2015 be declared unconstitutional, null and void. These sections provide for the offence of cyberstalking and interception of communications by law enforcement. The Court of Appeal found against the appellants and ruled that the provisions are made pursuant to the provisions of the Constitution. After considering some Nigerian case law and foreign authorities, the Court of Appeal concluded that

[55] Article 19(3) of the ICCPR on the four-part test of legality, legitimacy, necessity and proportionality; Principle 22 of the Declaration of Principles on Freedom of Expression and Access to Information in Africa 2019 https://www.achpr.org/presspublic/publication?id=80 (accessed 1 December 2021). See section 2.4 above.
[56] Section 3.3.1 A above.
[57] CA/L/556/2017 (Court of Appeal) delivered on 1 June 2018.

© University of Pretoria

'there is therefore, no compelling reason to resort to foreign jurisprudence in view of the rich indigenous case law and materials (supra) to resolve the controversy.'[58] If the need to protect the right to freedom of expression online is compelling enough, the Court would have benefited from foreign jurisprudence in at least four major ways.

One, according to the Court, foreign jurisprudence would be applicable only if there was no indigenous case law and legal material to be applied that is the same as the one to be decided. A closer look at the provisions of section 24 shows that it deals with online speech or content, that is the kind of speech shared through networks which include the Internet. If the provisions of section 24 had been compared strictly with the limitations provided for under section 39(3)(a) of the 1999 Constitution of Nigeria, it would show that instances where laws may be used to limit speech do not include online networks which the provisions of section 24 deals with. As a result, not only are there no indigenous case law and materials that could have helped to interpret section 39(3)(a) on online speech or content, the Constitution, which is the most 'indigenous legal material' does not include it in the most proximate limitative provisions of the right to freedom of expression – section 39(3)(a). An internationalist approach would have benefitted the protection of the right by reading international human rights law into domestic legislative gaps such as this.[59]

Two, the Court based its decision on the constitutional validity of section 24 which deals with online free speech on the provisions of section 36(12) of the Constitution which was about the constitutionality of criminalisation.[60] This basis is only one of the many legs the Court ought to have based its judgment on. This is because in determining the issue in question exhaustively, it was necessary for the Court to also consider the provisions of section 39 of the Constitution on freedom of expression which is one of the most substantive issues before the court. The other legs are the catalogue of international authorities cited by the appellants. An internationalist approach could have improved the judgment by contributing to the local jurisprudence of online speech in Nigeria in that it will present the opportunity to have the provisions of section 24 reviewed against the provisions of section 39 of the Constitution.

Three, in determining whether an offence is criminal, the crime of armed robbery is clearer to the members of the public than what might constitute insult or annoyance. The Court's position that section 24 is constitutional on the basis of section 36(12) suggests that the lawmaker can criminalise any activity that appears offensive while

[58] Court of Appeal (n 57 above) 29.
[59] Judicial borrowing for interpretation of rights is not only desirable but necessary when building jurisprudence for regulation of hate speech. See S Fredman *et al* 'Comparative hate speech law: Memorandum' (2012) *Oxford Pro Bono Publico* https://www.law.ox.ac.uk/sites/files/oxlaw/1._comparative_hate_speech_-_lrc.pdf (accessed 15 June 2021).
[60] Court of Appeal (n 57 above) 24.

© University of Pretoria

the Court will always rise to the occasion to interpret such laws justly.[61] The problem with such a position is that allowing patently problematic laws with the hope that they will be properly interpreted by the courts costs in human freedom. Before the Court turns around to deliver justice, avoidable rights violations with far-fetched negative impacts would have occurred.

Four, the second issue on the protection of the right to privacy and section 38 of the Cybercrime Act was found to be constitutional based on the misapplication of section 39(3)(a) which dealt with the right to freedom of expression. In determining whether the provisions of section 38 of the Act violates the right to privacy provided for under section 37, the court used the provisions of 39(3)(a) which was about the instances law could be used to limit the right to freedom of expression. The most proximate section that applies to the limitation of the right to privacy is section 45(1)(a)(b) which allows such limitation based on public interests and the rights of others.

In contrast, in *Laws and Awareness Initiatives v Federal Republic of Nigeria*[62] which was filed before the Court of Justice of the Economic Community of West African States (ECOWAS Court), the provisions of section 24 of the Act were issues for determination before the Court. Some of the issues for determination were whether the provisions of section 24 of the Cybercrime Act were in violation of international human rights law and whether Nigeria can be ordered to repeal the provision. One of the reliefs sought by the applicant was that the ECOWAS Court should declare the provisions of section 24 as a violation of international human rights law. Assuming jurisdiction on the application, the ECOWAS Court applied international human rights law principles in determining the merits of the case. It relied extensively on international human rights standards highlighted under chapter two that the provisions of section 24 do violate international human rights law the law should be either amended or repealed in line with the law.

In providing justification for its decision, the ECOWAS Court considered the section as a whole. The ECOWAS Court found that section 24 complied with the principles of legality because states have the powers to criminalise any conduct in their national legislation.[63] It also found that the section complied with the test of legitimacy under international law because it aims to safeguard the rights of others.[64] It however found that the provisions of section 24 are not necessary in a democratic setting and they are disproportionate because of the criminal penalties attached.[65] While the overall reasoning of the ECOWAS Court finds the provisions in violation of international human rights standards, some of its rationale are at variance with international human

[61] Court of Appeal (n 57 above) 27.
[62] ECW/CCJ/JUD/16/20 (ECOWAS Court) delivered on 10 July 2020.
[63] ECOWAS Court (n 62 above) 32.
[64] ECOWAS Court (n 62 above) 33.
[65] ECOWAS Court (n 62 above) 38.

© University of Pretoria

rights law on the principle of legality. This variance is due to the ECOWAS Court's position with respect to the words used in the Act. As earlier noted, vague words are capable of being misinterpreted both by an ordinary person and state authorities.

Furthermore, section 26 of the Act provides for racist and xenophobic offences; however, on a closer look, it is an attempt to criminalise online hate speech. Structurally, it can be divided into two broad parts: the offence and punishment section and the definition section. The first part is further subdivided into four parts. The first two parts, section 26(1)(a-b) criminalises the intentional distribution of any racist or xenophobic materials and threats through a computer system or network against a person or a group of persons based on a set of characteristics including race, colour, descent, national or ethnic origin and religion. The third part, section 26(1)(c) criminalises insults against a person or a group of persons through a computer system or network based on the characteristics above. The first section of the last part, section 26(1)(d) criminalises distribution of materials through a computer system or network justifying genocide or crime against humanity. The second section of the last part, section 26(2) defines the meanings of crime against humanity, genocide and racist and xenophobic material as used in section 26(1).

Under international human rights law, section 26 of the Act fails on two main grounds. The first ground is that in limiting speech through hate speech, it must comply with the four-part test of legality, legitimacy, necessity and proportionality.[66] Second, such limitations must also consider the six major factors that must be considered in determining whether a speech is hateful or not.[67] A careful analysis shows that on the principle of legality, while the provision defines the offence and may be sufficiently precise, it becomes vague when it considers 'insult' under section 26(1)(c) as hateful or prohibited and uses it to limit speech. Determining insults in formal and colloquial speech in many instances is as relative as it is dynamic.[68] With respect to the principle of legitimacy and necessity, the need to limit speech based on online hate speech in order to forestall the precipitation of both online and offline violence is urgent in a society like Nigeria. However, while such limitation is allowed under international human rights law, the law also requires that such criminalisation should be reserved for only serious offences and as a last resort in order to consider alternative means of regulation.[69]

Just as there is no evidence of criminalisation of hate speech under section 26 as a last resort, there is no evidence that it considers other alternative measures to

[66] Section 2.4.1 A iii above.
[67] As above.
[68] A Clooney & P Webb 'The right to insult under international law' (2017) 48 *Columbia Human Rights Review* 25.
[69] Section 2.4.1 A iii above.

© University of Pretoria

criminalisation.[70] As a result, while the principle of legitimacy may be complied with in terms of protecting the rights of others and public interests, it misses the opportunity to legitimise and pass the necessity test through alternative methods to hate speech interventions like education, sensitisation and awareness, media training and literacy and many others. With respect to proportionality, the offence of online racist and xenophobic attacks do not have any punishment attached to them under the section and in the Act in general. This is because the punishment section with a jail term of five years or a N10 000 000.00 fine or both in section 26(1)(d) is with respect to distribution of online materials that justify genocide and crime against humanity, not online racist and xenophobic attacks.

These provisions present at least two shortcomings. First, it misses an opportunity to adequately provide for online hate speech under international human rights law. This is because international human rights law does not forbid criminalisation of hate speech, which includes a clear definition of the offence and a punishment section that both comply with article 19(3) of the ICCPR. In addition to this, the law requires that such criminalisation must be reserved for serious offences, online racist and xenophobic attacks are some, and must be a last resort. While the serious offences aspect is however complied with under section 26(1) and the definition of the offence in 26(2), criminalisation as last resort which must clearly state the punishment involved and be in line with international human rights law is absent and therefore not complied with. The second issue is that in limiting hate speech, six major factors must be considered.[71] This is because, given the cultural and historical differences in many heterogeneous societies, what may amount to hate speech is informed by context. These factors, especially during judicial review, helps to ensure that the regulation of hate speech is stripped of its negative impacts on free speech and the surrounding circumstances of each instance of hate speech is considered to arrive at a decision. Section 26 clearly does not provide for any of these factors. All of these shows that section 26 of the Act does not comply with the four-part test under international human rights law.

C Violence Against Persons (Prohibition) Act, 2015

The Violence Against Persons (Prohibition) (VAPP) Act, 2015 has the objective of prohibiting all forms of violence against persons in private and public life, providing maximum protection and ensuring effective remedies for victims and punishments for offenders. It is the most primary and comprehensive law with respect to violence against persons in Nigeria. This re-statement of the objectives is important in showing

[70] A Scheffler 'The inherent danger of hate speech legislation: A case study from Rwanda and Kenya on the failure of a preventative measure' (2015) *fesmedia Africa series* https://library.fes.de/pdf-files/bueros/africa-media/12462.pdf (accessed 25 June 2021).
[71] Section 2.4.1 A iii above.

© University of Pretoria

that even in the objectives of the law, there are a number of shortcomings with respect to effectively combating online harms in Nigeria.

In assessing the law, while the provisions for psychological, emotional and verbal abuse are included as a form of violence under the Act, the general reading of the law shows that such abuse is limited to the physical space. Sections 3, 5, 14, 17 and 18 all provide for the offences of psychological coercion, force and threat that is detrimental to psychological well-being, emotional, verbal and psychological abuse, stalking and intimidation respectively. Two proofs point to the Act being limited to physical occurrences only. One, section 32 which provides for the powers of the Police in the Act refers to the 'scene of an incident of violence' which connotes that such offences contained in the Act are limited to physical spaces. Two, the meanings of various words like 'emotional, verbal and physical abuse', 'stalking', 'victim' and 'violence' under section 46 do not make any specific mention of these definitions to apply to the online space in Nigeria to protect persons from violence.

However, it is important to note that the meaning of 'harassment' under the section may include 'repeatedly sending, delivering, or causing delivery of information such as … electronic mail, text messages or other objects to any person.' The challenge posed by this law however presents two opportunities. First, the law could be amended to include the various dimensions of online violence against persons to help to combat online gender-based violence in Nigeria. Second, it provides an opportunity to have the problematic provisions of section 24 of the Cybercrime Act moved to a more theme-specific law like the VAPP Act with respect to online gender-based violence in Nigeria.

D Child Rights Act, 2003

The Child Rights Act (CRA), 2003 is the most primary and comprehensive legislation on the protection and promotion of the rights of the child in Nigeria. Even though the law does not provide specifically for the right to freedom of expression for children, section 3 of the Act provides that the law would apply the provisions of Chapter IV of the 1999 Constitution of Nigeria on fundamental human rights which includes the rights under the Chapter for children in Nigeria. In addition to this provision, section 1 of the Act provides that every action taken must be in the 'best interest of the child.' Sections 35 and 36 prohibits and punishes importation of harmful publication while section 277 defines harmful publication as any information that targets children and portrays harmful information such as crime, violence, repulsive incidents, immoral words or character and obscene and indecent publication. It also defines harm as 'ill-treatment or the impairment of physical, mental, intellectual, emotional, or behavioural health or development.' These references to harms or violence under the Act however do not include specific reference to the dimensions of violence children face online in Nigeria which shows a gap in legislative framework that could guide proximate actors on the protection of children online in Nigeria.

© University of Pretoria

E The Nigeria Communications Act, 2003

The Nigeria Communications Act (NCC Act), 2003, is the most primary legislation with respect to telecommunications regulation in Nigeria. The Act establishes the Nigerian Communications Commission (NCC) which enforces and implements the provisions of the Act. While the Act does not have a direct relationship with online harms, it has a direct impact on the right to freedom of expression online and how online harms may be regulated by state actors. The Nigerian government has used a combination of the laws analysed above to limit not only the right of expression online but also other forms of digital rights.[72] For example, in limiting access to a number of websites allegedly disseminating separatists' views and spreading online harms in Nigeria, the government ordered their block from ISPs using the NCC Act.

In the Act, the provisions of section 146(1) provides that a licensee[73] (an ISP) should not allow its facilities to be used for criminal offences to the best of their ability. This provision does not consider the impacts of problematic provisions of the Criminal and Penal Codes, the Cybercrime Act and other laws that are problematic for online expression. Under this provision, so far there is an extant law that prescribes an offence, whether rights-respecting or not, a licensee must obey such law. With respect to section 146(2), the Commission or any other authority may give a written request to the licensee to assist in the prevention of any crime in any written law in Nigeria. This is problematic in that what it would take for such a request is a written request from the Commission or any government authority for a licensee to comply with the request. Here, there is no evidence of any checks against the powers of the Commission or any other authority in the Act or any other. The absence of such checks, given how most state actors are complicit in abuse of the rule of law, may give rise to many human rights violations. There should be at least one institutional check either by the judiciary or the legislature or where possible, both, before the granting of such a request by a licensee.

The provisions of section 146(3) also shield ISPs from any criminal or civil sanctions with respect to the acts carried out under sections 146(1) & (2) in good faith. This provision gives rise to two major challenges. First, it does not encourage private sector

[72] Paradigm Initiative & OONI 'Tightening the noose on freedom of expression: 2018 Status of Internet Freedom in Nigeria' 11 June 2019 https://ooni.org/documents/nigeria-report.pdf (accessed 15 August 2021). Luminate 'Data and digital rights in Nigeria: Assessing the activities, issues and opportunities' (2021) https://luminategroup.com/storage/1361/Data-%26-Digital-Rights-in-Nigeria-Report-%5BFINAL%5D.pdf (accessed 15 August 2021).

[73] The Act defines a network service provider as a person who provides network services under section 157 of the Act. Section 32 of the Act described the class of licences that may be issued to prospective licences as those to be used for operation and provision of telecommunications services. In this context, network service provider may be referred to as an Internet Service Provider.

© University of Pretoria

actors to perform their special duties and responsibilities[74] to protect human rights and second, it forecloses the possibility of seeking effective redress in the case of human rights violations as may be ordered by the state and carried out by the ISPs. The requirement of good faith should only avail the ISPs where they have required and received human rights-based assessments of the proposed act from the NCC or such government authority making such request, a judicial warrant by a judicial officer in a superior court of record with respect to the request and publication of such requests by the NCC to the public annually or at such time as may be determined by stakeholders during a legal review of the provisions.[75]

Considering section 146 as a whole and its compliance or otherwise with respect to international human rights law, it does not prescribe specifically the categories of offences where it would be applied and as a result does not comply with the requirement of legality where the law must be sufficiently precise. While the provisions might comply with the principle of legitimacy, it does not show that requests to licensee, oftentimes for blocking of online content, would consider the least intrusive means where the action would reasonably infringe on a right. In addition to this challenge, the Commission or government authority cannot justify their requests as it is less transparent in what it must include and even far less accountable as such requests cannot be checked by any other oversight system, therefore, the proportionality cannot be determined.

In addition to the provisions of section 146, the powers of the NCC to suspend, withdraw or make certain orders with respect to communication by an ISP or the general public under section 148(1)(a-d) of the NCA is premised on the occurrence of either public emergency or public safety. An ISP here could be a person or company that provides telecommunications services including Internet Service Providers (ISPs). A closer look at these powers granted to the NCC under the provisions is not only overbroad and not in compliance with international human rights law, it could be counter-productive to the public safety and public emergency aim sought to be achieved.

On its non-compliance with international human rights law particularly on the right to freedom of expression online, article 19(2) of the ICCPR guarantees the 'freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers… through any other media of …choice.' This right is however qualified by the provisions of article 19(3) which provides a four-part test that any restrictions on the right must comply with – legality, legitimacy, necessity and proportionality. It is also a notorious

[74] Section 3.4.2 A above.

[75] ARTICLE 19 & Electronic Frontier Foundation (EFF) 'Necessary and proportionate: International principles on the application of human rights law communication surveillance: Background and supporting international legal analysis' May 2014 https://www.article19.org/data/files/medialibrary/37564/N&P-analysis-2-final.pdf (accessed 24 August 2021).

© University of Pretoria

fact that these tests must be jointly achieved before a restriction based on article 19(3) can be said to be compliant under international law. While public emergencies or public safety can be a basis for restricting the right to freedom of expression online under article 19(3), it is not clear what examples would qualify as 'public safety' under the law. On the necessity and proportionality, the unfettered powers of the NCC without any form of external oversight from another arm of government (legislature or judiciary) on a role as important as facilitating telecommunications services cannot be said to be necessary in a democratic society. At least, a judicial oversight must be made mandatory before the exercise of such powers. In carrying out such oversight, a judge of a superior court of record would be assigned to review the request in order to grant it or refuse it. In instances where it could be difficult to get such orders due to imminent threat to life and property, NCC must show cause before a sitting judge for its actions after exercising such powers by submitting a detailed report on its activities and the possible impacts on human rights.

The basis for assessing such powers on whether it followed due process is to consider the period within which such powers may be exercised, for what purpose and its impacts on human rights. In addition to this, ISPs are not allowed to demand compliance with due process for such a request by the NCC. Given the nature of communications infrastructure, governments need private sector investments just as private investments need governments. Therefore, in carrying out requests from the NCC with respect to section 148(1)(a-d), given this power dynamics and the economic loss arguments for ISPs and telecom providers, the latter should be able to demand a judicial warrant and a human rights impacts assessment of a proposed order for cutting communication access for Nigerians.[76]

On how counterproductive the powers granted the NCC under section 148(1)(a-d) are, first, the responsibilities of public safety or issues that arise during public emergencies do not require cutting off communication access. Rather, it requires that communication of official and public information must be ramped up so that the public is adequately informed for its safety and other emergencies.[77] This is because not only is the right to freedom of expression important, access to information is a right in and of itself during emergencies. Leaving the powers to cut off modern public means of communication like Internet access in an age where survival is nearly driven by Internet access to just the NCC or only the Executive could encourage gross abuse of state powers. These powers could also become problematic as the NCC could be the victim of political whims rather than the dictates of rule of law. Second, a truly democratic government in Nigeria needs to rethink its concept of power in terms of

[76] As above.
[77] United Nations, Educational, Scientific and Cultural Organisation (UNESCO) 'The right to information in the time of crisis' 2020 https://en.unesco.org/sites/default/files/unesco_ati_iduai2020_english_sep_24.pdf (accessed 23 August 2021).

© University of Pretoria

digital communications such as Internet services and how it impacts human rights. This is because in order to ensure effective regulatory governance in the telecoms sector especially as it relates to preventing online harms and protecting online speech, state actors such as the NCC have to reassess their powers and provide means for holding themselves accountable.

F The Protection from Internet Falsehoods and Manipulation and other Related Matters Bill, 2019

The Protection from Internet Falsehoods and Manipulation and Other Related Matters (PIFM) Bill, 2019 is a proposed legislation awaiting committee report before the Nigerian Senate with a twin-objective to suppress falsehoods and manipulation.[78] The Bill is divided into six major parts which focus on aims and objectives; prohibition of transmission of false declaration of fact; regulations dealing with transmission of false declaration of fact; regulations for Internet intermediaries and providers of mass media services and declaration of online locations respectively.

In addition to various analyses of the Bill which point to the violation of the right to freedom of expression online in Nigeria, section 3 of the Bill is similar to the provisions of sections 50 - 52 of the Criminal Code and sections 416 - 422 of the Penal Code which provide for the offence of sedition online as it criminalises speech against the state. In addition to these provisions, the content of the Bill is similar to that of Singapore on the same subject which has been highlighted as problematic in the report by United Nations Special Rapporteur on the Right to Freedom of Expression and Opinion on disinformation.[79] Some of the criticism by the Special Rapporteur against the Singaporean version, and in effect, the Nigerian version, is that the provisions are vague, the harm sought to be prevented is not clear and there is no concrete connection between the legitimate aim and the harm sought to be neutralised. The criticism also includes the lack of accountability, transparency and disproportionality of punishment in the provisions as they concentrate powers in the executive which is prone to abuse without judicial or legislative oversight.[80]

G National Commission for the Prohibition of Hate Speeches (Est. etc.) Bill, 2019

The objectives of the National Commission for the Prohibition of Hate Speeches (Est. etc.) (NCPHS) Bill include the promotion of national cohesion and integration and

[78] Explanatory memorandum *Protection from Internet Falsehoods and Manipulation and Other Related Matters Bill 2019* https://placbillstrack.org/upload/SB132.pdf (accessed 15 October 2021).
[79] United Nations General Assembly 'Disinformation and Freedom of Opinion and Expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/47/25' 13 April 2021 http://undocs.org/en/A/HRC/47/25 (accessed 26 August 2021).
[80] United Nations General Assembly (n 44 above) paras 63-80.

© University of Pretoria

outlawing unfair discrimination, hate speech and to provide for the establishment of an independent commission for the prohibition of hate speeches. The Bill is divided into four main parts including the preliminary section; discrimination to which the bill applies; establishment, powers and functions of the independent commission for the prohibition of hate speeches and enforcement respectively. In an analysis of the Bill with respect to its compliance with international human rights law, it was found that not only were the provisions overbroad and unclear, its punishments are disproportionate.[81]

H The need for legal reform in preventing online harms in Nigeria

In an analysis of the sources of free speech provisions in Nigeria and most other African countries, Jeffrey noted that these provisions are largely shaped by colonial laws and are therefore in need of reform 'to meet the changed conditions of the modern world.'[82] Laws of colonial origin are described as those laws which were in operation before the official independence of colonised states. Laws having colonial legacies on the other hand, are such laws which even though they are made after colonialism officially ended, they still show the imprints of colonial laws.[83] Some of these laws have impacts on general issues such as media licensing and regulatory laws and specific problematic provisions such as publication of false news likely to cause fear and alarm, insult, insult to religion, sedition and criminal defamation. Provisions of these are still provided for in the Criminal and Penal Codes of Nigeria as examined above despite being of colonial origin. These laws have outlived their usefulness not because they are colonial relics and served the end of oppression and discrimination, they were fashioned to violate the right to freedom of expression.

This is not the first time a relationship has been established between colonial laws and human rights protection in Nigeria. In their work, Ogbondah and Onyedike drew a straight line that connected the relationship between colonial press laws and post-independence press laws in Nigeria.[84] In another study, Ogbondah noted that the most defining feature of both systems, even though marked by different justification, was the oppressive and authoritarian characteristics of both the colonial systems and the military governments that afflicted Nigeria post-independence.[85] According to Dada,

---

[81] T Ilori 'A socio-legal analysis of Nigeria's Protection from Internet Falsehoods, Manipulations and other Related Matters Bill' 5 December 2019 *AfricLaw* https://africlaw.com/2019/12/05/a-socio-legal-analysis-of-nigerias-protection-from-internet-falsehoods-manipulations-and-other-related-matters-bill/ (accessed 24 August 2021).

[82] AJ Jeffrey 'Media freedom in an African state: Nigerian law in its historical and constitutional context' unpublished PhD thesis, University of London, 1983 58; Adjei (n 35 above) 439.

[83] Jeffery (n 82 above).

[84] CW Ogbondah & EU Onyedike 'Origins and interpretations of Nigerian press laws' (1991) 5 *Africa Media Review* 59-70.

[85] Ogbondah (n 38 above).

© University of Pretoria

'the advent of the colonialists inevitably made the Nigerian societies become subject to the political, economic and social domination and subjugation of the colonial power.'[86] To Eze, the protection of fundamental rights cannot thrive on the colonial legal system in Nigeria particularly as it was predominantly racist and authoritarian.[87]

Today, online communication in Nigeria has become powerful in the sense that they have not only challenged the traditional notions of its dissemination,[88] they have also caused the Nigerian government to change its policies from time to time.[89] Most Nigerian laws that involve the use of digital technologies often seek to limit the rights provided for under chapter four of the 1999 Constitution (as amended). While some of them do not make reference to the need to protect the substantive right as provided for under the Constitution, the ones that do make such reference proceed directly to violate the said right. This obvious dissonance suggests a disconnection between the understanding of state responsibilities in protecting human rights under the Nigerian Constitution and the laws that are enacted to further operationalise them. These responsibilities are those performed by the legislature in making the laws, the executive in implementing them and the judiciary in interpreting them.

Therefore, as a starting point, there is a need to repeal the provisions of publication of false news likely to cause fear and alarm; sedition; criminal defamation; and abusive and insulting language. In addition to this, the provisions of sections 24(1)(a-b), (2)(a-c) and 26 must also be amended. With respect to the VAPP Act, sections 3, 5, 14, 17 and 18 should be amended to accommodate the online dimensions of these forms of violence. The Child Rights Act should also be completely reviewed in its various provisions to accommodate the various harms that could impact children online in Nigeria. The provisions of 146 and 148 should also be amended. These repeals and amendments must be carried out in line with international human rights law. Both bills before the National Assembly on disinformation and hate speech should also be stopped and reviewed based on international human rights law. Asides these, there is also a gap of a rights-respecting framework with respect to online harms like information disorder, cyberbullying, cyberagression, online gender-based violence, online violence against children and online hate speech in Nigeria.

[86] JA Dada 'Human rights protection in Nigeria: the past, the present and goals for role actors for the future' (2013) 14 *Journal of Law, Policy and Globalisation* 1-13.
[87] OC Eze *Human rights in Africa: Some selected problems* (1984) 1-314.
[88] Section 3.3.1 B above.
[89] F Kperogi *Nigeria's digital diaspora: Citizen media, democracy, and participation* (2020) 113-134.

© University of Pretoria

5.2.2 Legal and regulatory provisions on online harms in South Africa

The South African Constitution provides for a bill of rights, the right to freedom of expression, its scope and its limitations.[90] It also provides for a unique application of international law by the Courts. This provision requires the judiciary to 'prefer any reasonable interpretation of the legislation that is consistent with international law over any alternative interpretation that is inconsistent with international law.'[91] South Africa does not have a comprehensive or primary law with respect to online harms or platform governance. In dealing with online harms in South Africa, a close attention must be paid to its turbulent historical and racial history. This is particularly important to ensure that laws and regulatory designs do not perpetuate this history but at the same time prevent online harms while also protecting the right to freedom of expression online. One of the ways of achieving this, is by paying close attention to how current laws or proposed laws on online harms comply with South Africa's human rights obligations. Like Nigeria, various provisions on online harms are found in other laws some of which are analysed below in line with international human rights law.

A Criminal Procedure Act of 1977

The Criminal Procedure Act of 1977 deals mainly with proceedings with respect to criminal law in South Africa and does not define offences. Section 104 deals with proceedings with respect to 'printing, publishing, manufacturing, making or producing blasphemous, seditious, obscene or defamatory matter…' while section 242 of the Act deals with proceedings with respect to defamation. These provisions suggest that defamation, sedition, blasphemy and obscenity are still criminal offences in South Africa.

In dealing with online defamation especially on social media platforms however, South African courts have largely adopted civil and administrative remedies than criminal sanctions which seems more attune to international human rights standards.[92] For example, in *Heroldt v Wills*,[93] an applicant alleged that his reputation was injured by the respondent's Facebook post which portrayed him as someone suffering from addiction. In arriving at its decision in the case, the South Gauteng High Court per

[90] Section 7, 16, & 36 of the Constitution of South Africa, 1996. Section 36(1)(a-e) applies the four-part test for limitation of rights including the right to freedom of expression namely: nature of the right (legality); importance of the purpose of the limitation (necessity); relation between the limitation and the purpose (legitimate aim); and less restrictive means to achieve purpose (proportionality).
[91] Section 233 of the Constitution of South Africa, 1996.
[92] United Nations General Assembly 'Expert workshops on the prohibition of incitement to national, racial or religious hatred, A/HRC/22/17/Add.4' 11 January 2011 http://undocs.org/en/A/HRC/22/17/Add.4 (accessed 1 December 2021) para 12.
[93] (2013 (2) SA 530 GSJ)). See also *RM v RB* (2015) (1) SA 270 (KZP), para 28 where the Court noted the importance of social media platforms but noted that defamatory postings could pose risks against the reputational integrity of individuals.

© University of Pretoria

Willis held that while there is a *lacuna* in South African law on social media expressions, the provisions of the South African Constitution on the rights to freedom of expression and privacy were enough basis. As a result, the Court found in favour of the applicant and granted his request for removal of the post. However, the Court refused the applicant's requests that the respondent should be restrained from making injurious comments and be jailed for 30 days.

In *Mwanele Manyi v Mcebo Freedom Dhlamini*,[94] the court found that a Whatsapp post infringed on the plaintiff's right to dignity as the plaintiff was referred to as a 'lame horny donkey.' The court also held that the words were defamatory and the 'seriousness of the defamation, the nature and extent of the publication, the reputation, character and conduct of the plaintiff, the motives and conduct of the defendant' must be taken into consideration when arriving at a judicial decision on defamatory statements. A total amount of R 55 000 (USD3 500) was awarded as damages against the defendant.

In a recent case, *EFF & Others v Manuel*,[95] decided by the Supreme Court of Appeal (SCA), the Court dismissed the appeal of the applicant and upheld the decision of the Guateng Division of the High Court. The High Court had found that the applicant published a defamatory statement on its Twitter account that was later retweeted by Julius Malema, EFF's leader. Both the EFF and Malema had a combined followership of about 2.8 million on Twitter. The statement stated that the respondent had a special relationship with an appointee of the President for the South African Revenue Service. The respondent had chaired the committee that recommended the appointee to the President. The applicant later went on to describe the appointment as corrupt and nepotistic. The SCA held that the applicant did not establish any evidence that the claim was true and as a result, the message was calculated to injure the respondent. The SCA ordered the applicant to remove the statement within twenty-four hours but granted a leave to appeal the award of damages granted by the High Court.

According to South Africa's Constitutional Court, the right to freedom of expression, its limitation and social media use is that the right is not unfettered and that in expressing oneself, there should be due regard to the right of others.[96]

B Cybercrime Act of 2020

The Cybercrime Act of 2020 provides for various aspects of regulating cybercrime in South Africa and some of them include the definition and prosecution of offences, mutual assistance with respect to international cybercrimes and others. The Act

[94] Case number 36077/18 (un reported) heard in the High Court of RSA Gauteng Division, Pretoria.
[95] 2021 (3) SA 425 (SCA) (17 December 2020)
[96] *The Citizen 1978 (Pty) Ltd and others v. McBride* (Johnstone and others as amicus curiae) 2011 (8) BCLR 816 (CC), para 152.

© University of Pretoria

provides for four major aspects that touches on prevention of online harms and online speech governance in South Africa. The first aspect is provided for under section 14 which provides for the offence of inciting damage or violence against a group of persons online while the second one is provided for in section 15 on the offence of threatening damage or violence to a person or group of persons online. The third one is in section 16 which provides for the offence of non-consensual sharing of intimate images online, which includes real or simulated images. Section 19(7) of the Act provides for the punishment of sections 14, 15 and 16 to include a fine or imprisonment of up to three years or both. The fourth one are the provisions of sections 21 and 22 of the Act which provides for the power of the court to make an order with respect to the provisions of sections 14, 15 and 16 of the Act and the duty of electronic communications service providers to furnish particulars to court respectively.

Applying international human rights standards to each of these provisions, especially sections 14 and 15, show that while there are legitimate and necessary bases for these provisions to combat harmful online speech, there are challenges as to the requirements of legality and proportionality. For example, section 14 which provides for data messages with the intention to incite damage to property or violence against a person or groups of persons is largely deficient. This deficiency is seen in how such provision, even while it might be well-intentioned, does not take into account the requirements under international human rights law on the prohibition of incitement to hatred, hostility or discrimination.[97] These requirements are such that laws should not only criminalise these acts but provide for additional six factors which includes prevailing social and political context; status of the speaker to audience; clear intent to incite; content and form of speech; extent of speech, public nature; size of audience and means of dissemination; and real likelihood and imminence of harm. The implication of not adding these requirements is that the provisions of section 14 may be used to chill 'offensive, shocking and disturbing speech' which are permissible forms of speech under international human rights law.[98] Additionally, the terms of punishment under section 19(7) for this offence is disproportionate without considering these additional factors and background.

Equally, the provisions of section 15(a)(b) on electronic communications that threatens to damage property or cause violence against a person or group requires a delicate balancing between the person or group threatened and the person who issues the threat. While this provision may be said to be a legitimate basis for limiting online expression, as worded, with respect to the legality principle, the provisions of section 15(a)(b) is vague and could be used to chill legitimate expression. In addition to this,

[97] United Nations (n 92 above) para 29.

[98] United Nations General Assembly 'Online hate speech and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/74/486' 7 September 2012 http://undocs.org/en/A/74/486 (accessed 6 October 2021) para 13.

© University of Pretoria

it also focuses only on the recipient of the threats without adequate application of contextual factors. This is because based on the general principles of criminal law of *Mens Rea* and *Actus Reus* which refer to the intention and actual commission of crime as basis for arriving at an offender's guilt, treating threats through electronic communication as a basis for arriving at an offender's guilt is challenging because while threats could be precursors to the commission of a crime, they could also be empty.[99] For example, some forms of threats are as a result of legitimate human outburst which could be rendered in the heat of the moment. However, the provisions of section 15(b)(ii) attempt to remedy this vagueness by stating that the person being threatened should reasonably perceive the communication as threatening. The attempt still does not cure the vagueness under section 15(a)(b). Failure to ensure that and criminalising threats could chill and unduly rein in legitimate expression. Until there is a direct link between a threat through speech, actually carrying out such threat and such threat impacting others, criminalising just the threat could adversely affect legitimate speech. However, this does not necessarily foreclose the rights of the person being threatened or the responsibilities of law enforcement to investigate the credibility of such threat. What is most important is that a delicate balance is struck between the right to express and the right to safety of life and property. What could have provided the much needed balance is for the provision to ensure that such person or group of persons must show that they have suffered some type of harm or violence as a result of such threat. An additional practical way of striking such balance is by applying the six factors highlighted above under international human rights law on limiting prohibited speech to determine the possibility of such threat taking place. As a result, the terms of punishment under section 19(7) are disproportionate without considering these additional factors and background.

The provisions of section 16 on the disclosure of intimate images through electronic communication may be detailed to an extent in its compliance with the criteria of legality, legitimacy, necessity and proportionality under international human rights standards. This is because unlike many other laws that criminalise non-consensual sharing of intimate images, the provision centres the consent of the person whose image is shared – it becomes a crime only when the sharing is not based on their consent. This presents two interesting perspectives. First, it limits the decision to share such images between two persons but places the autonomy of the person whose image is being shared at the centre. Second, it whittles down the power of the state in the regulation of human bodies by emphasising on the importance of consent of the individual and not the approval of the state. It is also noteworthy that the provision is gender neutral and shows that anyone, regardless of their gender can be victims of non-consensual sharing of intimate images by using neutral terms such as 'A' and 'B' to refer to persons. However, it would have been far more useful for the provisions to

[99] A Guichard 'Hate crime in the cyberspace: the challenges of substantive criminal law' 18 *Information and Communications Technology Law* 211.

© University of Pretoria

expand on the duration of consent which means that the consent of the person whose image is being shared can be withdrawn at any time and not respecting such withdrawal would still amount to a criminal offence.[100] The provision can also do more to consider the conditions upon which the consent is secured because consent secured under duress, or consent of a minor should not be construed as giving consent with respect to sharing of intimate images. It will therefore be important to consider this additional background to ensure that punishments provided for under section 19(7) are proportionate.

The provisions of sections 20 and 21 of the Act is central to the implementation of the sections highlighted above. This is because both provisions vests powers in the court to address the harms caused by the offences under sections 14, 15 and 16 discussed above while also giving clarity on the role of the electronic communications service provider to furnish necessary particulars to the court. An electronic communications service provider is described under section 1 as a person (natural or juristic) who provides an electronic communication service to the public under the provisions of the Electronic Communications Act, 2005 or a person who has lawful authority on the use of private electronic communication and is exempted in terms of the Electronic Communications Act, 2005. Such definition of a provider suggests that social media platforms may be included as electronic communications service providers. These provisions are important in that they establish the necessary relationship between judicial oversight on these offences and the responsibilities of electronic communications service providers. For example, section 20 of the Act requires the magistrate court to make two kinds of orders expeditiously and *ex parte* with respect to the provisions of sections 14, 15 and 16 of the Act: one, stop a person from further disclosing data messages based on the offences in sections 14, 15 and 16 or two, order an electronic service provider whose services is used to host the data message to remove or disable such person's access on its platform. This is particularly important in that it conforms with international human rights requirement that in limiting content online, the specific problematic content should be restricted and not the entire platform hosting such content.[101] Section 21 of the Act generally provides for the responsibilities of the electronic service provider in furnishing the court with the details that would assist the court in adjudicating on the offences provided for in sections 14, 15 and 16.

C Domestic Violence Act of 1998

The Domestic Violence Act (DVA) of 1998 seeks to eliminate domestic violence against women, children and other vulnerable groups in South Africa. It makes references to the protection of rights like equality and security of the person as

[100] DK Citron & MA Franks 'Criminalizing revenge porn' (2014) 49 *Wake Forest Law Review* 355.
[101] United Nations General Assembly 'General Comment No 34, CCPR/C/GC/34' 12 September 2011 http://undocs.org/en/CCPR/C/GC/34 (accessed 26 June 2021) para 22.

© University of Pretoria

provided for under the South African Constitution and international human rights treaties. Its definition of domestic violence has evolved over a number of amendments to the law to include 'emotional, verbal and psychological abuse', 'intimidation', 'harassment', 'stalking' and any other abusive behaviour that causes or may cause imminent harm to the safety and well-being of the complainant in online spaces.[102] It further defined the meaning of emotional, verbal and psychological abuse to include repeated insults, ridicule or name calling, threats to cause emotional pain, and obsessive possessiveness that invades a complainant's privacy, integrity or security. Harassment is also defined to include repeated stalking, calling or delivery of electronic information.[103]

It is however noteworthy that despite these definitions of domestic violence and its examples, domestic violence is not defined as an offence. What this lack of definition means is that the act of domestic violence and its various forms are neither specifically criminalised nor is there a punishment prescribed under the law for domestic violence. What exists, is the criminalisation of domestic violence through offences such as 'assault (either common or with intent to cause grievous bodily harm), pointing a firearm, intimidation, rape, or attempted murder (among other charges).'[104]

In a submission by civil society organisations that work on women's rights online in and outside South Africa, the proposed amendment bill, which provides for new definitions such as 'coercive behaviour', 'emotional, verbal and psychological abuse', 'controlling behaviour', 'economic abuse', 'harassment', 'intimidation', 'sexual harassment', 'electronic communications', need to be expanded, with the inclusion of words that convey the true impacts of online-based gender violence.[105] In their submission, the meaning of 'electronic communication' in the Act should carry additional instances where such communication is 'real, simulated or manipulated' which accommodates the reality of women being the most victims of deepfakes and other manipulated media. For 'emotional, verbal and psychological abuse', it is suggested that the Act includes threats to disclose the complainant's gender, gender identity, sexual identity, sexual orientation or perceived sexual orientation without the complainant's consent. All of these seem to accommodate the new realities of online gender-based violence as a form of online harm.

The DVA therefore is positioned to comply with international human rights law as its purpose is clear, precise, legitimate, necessary and proportionate. However, there are

[102] Section 1 of the Domestic Violence Act, 1998.
[103] As above.
[104] L Vetten 'Domestic violence in South Africa', November 2014 https://issafrica.s3.amazonaws.com/site/uploads/PolBrief71.pdf (accessed 24 June 2021).
[105] Alt Advisory & Research ICT Africa 'Domestic Violence Amendment Bill: Joint submission by Research ICT Africa and Alt Advisory' 9 July 2021 https://altadvisory.africa/wp-content/uploads/2021/07/ALT-Advisory-Research-ICT-Africa-Joint-Submissions-Domestic-Violence-Amendment-Bill.pdf (accessed 15 August 2021).

© University of Pretoria

no clear means of enforcing the online harms sought to be prevented under the Act. This is in addition to the need for both the definition of domestic violence as an offence and the punishments that are applicable to be provided for in the amendment especially as it affects the online space. This is because not only will domestic violence become a substantive crime, its ramifications in the online space will also be captured. Criminalising domestic violence indirectly through assault does not provide the effective avenue for tackling it as a substantive violation of rights.

D Children's Act of 2005

South Africa's Children's Act was enacted in 2005 with similar provisions like that of Nigeria. Perhaps due to the time it was passed, there were no meaningful engagements with respect to children's protection online because interactions with digital technologies were not as pronounced then like they are today. Section 7(i)(ii) of the Act provides against violence against children to include the protection of children from physical and psychological harm. This points to a policy and legislative gap that requires filling, given that children now face various challenges online with respect to their safety.[106]

E Electronic Communications and Transaction Act (ECTA) of 2002

Sections 73 and 77 provides for limited liability of ISPs and requirement of take-down requests for illegal content including child pornography, defamatory materials and copyright issues respectively. Section 71 of the ECTA provides for the recognition of an industry representative body, which is currently the ISPA.[107] It provides regulatory representation, support and advisory with respect to key regulations like the ECTA, ECA and others. The ISPA is primarily responsible for content regulation decisions through self-regulation of its members. The ISPA has provided clarity on a number of occasions with respect to its role in online content regulation.[108] In one of its

[106] Section 3.3.2 C above.

[107] Department of Communications 'Electronic Communications and Transactions Act (25/2002): Guidelines for Recognition of Industry Representative Bodies of Information System Service Providers' December 2006, https://www.ellipsis.co.za/wp-content/uploads/2011/02/IRB-Regulations-Gazette-29474.pdf (accessed 13 August 2021); ISPA 'Press release: ISPA recognised as an Industry Representative Body' 20 May 2009, https://ispa.org.za/press_releases/ispa-recognised-as-an-industry-representative-body/ (accessed 13 August 2021).

[108] ISPA 'ISPA submissions on the draft Films and Publications Regulations, 2020' 17 August 2020 https://ispa.org.za/wp-content/uploads/2020/08/ISPA-Submission-Draft-Films-and-Publications-Regulations-20200817.pdf accessed: 14 August 2021; ISPA, 'Submissions on the draft online content regulation policy' 15 July 2015, https://ispa.org.za/wp-content/uploads/2012/06/Internet-Service-Providers-Association-ISPA.pdf (accessed 14 August 2021).

© University of Pretoria

submissions on the regulation, it states that it does not have the power to regulate content hosted by specific social media platforms like Facebook or Twitter.[109]

F Promotion of Equality and Prevention of Unfair Discrimination of 2000 (PEPUDA)

Dealing with the regulation of hate speech in South Africa requires a contextually sensitive application of the law. This requirement is necessary as a result of South Africa's deeply problematic racial history.[110] Botha and Naidoo have argued that a free speech model that will be most effective in South Africa should be dominantly communitarian.[111] What this suggests is that laws that seek to limit free speech based on hate speech must have a clear and precise definition of hate speech with appropriate balances of government powers; the meaning of hate speech should be interpreted to include only 'deeply felt emotions of detestations, calumny or vilification' and not mere offense or ridicule; there must be a clear link between the hate speech and the harm intended; and the consideration of contextual and pluralistic factors.

Hate speech jurisprudence, whether online or offline, is still limited in South Africa.[112] This limitation is not as a result of lack of laws or academic discussions,[113] but as a result of limited judicial engagement by the courts, especially with the online dimensions of it. However, there have been a few cases, decided by the apex court in South Africa on hate speech that could set the judicial tone. It is important to consider one of the most comprehensive legislation on the regulation of hate speech in South Africa, the Promotion of Equality and Prevention of Unfair Discrimination Act 4 of 2000 (the PEPUDA). The first point to note with respect to the PEPUDA asides its objectives, which is to promote the rights provided for by the Constitution, is the

---

[109] ISPA 'ISPA submissions on the Hate Crimes and Hate Speech bill' 31 January 2019, https://ispa.org.za/wp-content/uploads/2019/02/ISPA-Hate-Crimes-and-Hate-Speech-Bill-31-January-2019.pdf (accessed 14 August 2021).

[110] K Naidoo 'Factors which influenced the enactment of hate-crime legislation in the United States of America: Quo Vadis South Africa' (2016) *Journal of South African Law* 705; J Botha & A Govindjee 'Regulating 'extreme cases of hate speech' inn South Africa: A suggested framework for a legislated criminal sanction' (2014) 27 *South African Journal of Criminal Justice* 154.

[111] J Botha 'Towards a South African free-speech model' (2017) 134 *South African Law Journal* 815-818.

[112] The SAHRC admitted to the constant evolution of hate speech regulation in South Africa. See SAHRC 'Findings of the South African Human Rights Commission regarding certain statements made by Mr Julius Malema and another member of the Economic Freedom Fighters' March 2019, para 13.3 https://www.sahrc.org.za/home/21/files/SAHRC%20Finding%20Julius%20Malema%20&%20Other%20March%202019.pdf (accessed 15 June 2021).

[113] J Botha '"Swartman": Racial descriptor or racial slur? *Rustenburg Platinum Mine v SAEWA obo Bester* [2018] ZACC 13; 2018 (5) SA 78 (CC)' 2020 10 *Constitutional Court Review* 353-377; F Cassim 'Regulating hate speech and freedom of expression on the Internet: Promoting tolerance and diversity' 28 (2015) *South African Journal of Criminal Justice* 303-336; K Naidoo 'The origins of hate-crime laws' 22 *Fundamina* 53-66; J Botha & A Govindjee (n 106 above)117-155.

© University of Pretoria

provisions of its section 3. This provision makes the PEPUDA an open-ended legislation with the capacity to receive international best practices into South Africa's jurisprudence on combating hate speech.

Section 10 of the PEPUDA provides:

> no person may publish, propagate, advocate or communicate words based on one or more of the prohibited grounds, against any person, that could reasonably be construed to demonstrate a clear intention to –
> (a) be hurtful;
> (b) be harmful or to incite harm;
> (c) promote or propagate hatred

The test for this provision came to the fore in the case of *Qwelane v South African Human Rights Commission and Another* (*Qwelane* case) where the Constitutional Court found that the provision was unconstitutional even though the applicant was guilty of hate speech.[114] One of the major findings of the Court in determining the unconstitutionality of the provision is section 10(1)(a) in that it is vague and cannot be used as an objective means to assess hate speech.[115] This was done by the Court leveraging the provisions of section 233 of the Constitution on the application of international law and specifically referring to the provisions of the ICCPR, ICERD and the African Charter in the process.[116]

G Films and Publications Amendment Act of 2019

The most comprehensive and primary law with respect to the classification of films, games and publications in South Africa is the Films and Publications Amendment Act of 2019. In addition to these objectives, the Act establishes the Films and Publications Board (FPB) which has responsibilities to implement and enforce the provisions of the Act. While it is not primarily focused on online content, aspects of the law, especially its various amendments have potential impacts on online content governance. One of the major objectives of the amendments is to reflect the increasing demands for online content and technological advances in the law.[117]

The most recent in the series of amendments of the law is the Films and Publications Amendment Act of 2019. Under the new amendment, 'harm' is defined to include causing 'emotional, psychological and moral distress… through on or offline medium including through the Internet.'[118] In addition to these definitions, it defines publication to include 'any content made available using the Internet.' This meaning of 'harm' suggests that it is one of the aims of the amendment to regulate online harms such as

[114] [2021] ZACC 22 (Constitutional Court) para 198.
[115] Constitutional Court (n 114 above) paras 144, 159.
[116] Constitutional Court (n 114 above) para 87.
[117] Memorandum on the objects of the Films and Publications Amendment Bill of 2015, para 1.2.
[118] Section 1(j) of the Films and Publications Amendment Act of 2019.

© University of Pretoria

online violence against children in form of child pornography online, online abuse, online hate speech and prohibited content as provided for in the Act. Each of these online harms are also punished with the use of the words like 'harm' and 'publication' as defined in the interpretation section. Other regulatory provisions on online harms include the responsibilities of private sector actors including social media platforms and ISPs in preventing such harms.

With respect to child pornography online, the Act criminalises the possession, production and distribution of child pornography.[119] The requirement of the FPB that ISPs must register with it in order to fight child pornography points to the fact that such offence also applies to online environment. With respect to online hate speech, the latest amendment provides for the prohibition of hate speech and advocacy of such as violence or harm through section 16(2)(b). It also provided that incitement of violence, advocacy of hatred and incitement of violence against persons with identifiable group characteristics is imminent as an offence under section 16(2)(b).[120] It also criminalises the distribution of such publication online including the Internet or social media platforms.[121] In addition to this, ISPs are required to prevent, report to the Police and preserve the evidence of failure of which could amount to various punishments.[122] In addition to this requirement by ISPs, section 16(4) provides for the classification of certain publications as 'XX' including explicit sexual conduct that violates or disrespect human dignity; bestiality, incest, rape or other act degrading human beings, incitement, encouragement and promotion of harmful behaviour; explicit infliction of sexual or domestic violence; and explicit depiction of extreme violence.

The amendment also provides for the criminalisation of non-consensual sharing of intimate images, which may be described as a form of online gender-based violence.[123] The elements of the offence is such that it must have been shared without consent of the victim(s) and must be intended to cause harm. It also provided for instances where such content publication shared may be referred to as 'sexual' to include all or part of an 'individual female's breasts, anus, genitals to pubic areas;'[124] shows something a reasonable person may consider as sexual because of its nature;[125] or the content as a whole will be considered as sexual by a reasonable person.[126] It also provides that the ISP should furnish the Board or Police with information on the identity of the person who publishes such content.[127]

[119] Section 2(d), 18G, 24B(5) & 24F of the Films and Publications Amendment Act of 2019.
[120] These offences are prohibited under section 24G of the Act.
[121] 18H of the Films and Publications Amendment Act of 2019.
[122] Section 27A of the Films and Publications Amendment Act of 2019.
[123] Section 18F & 24E of the Films and Publications Amendment Act of 2019.
[124] 18G(5)(a) of the Films and Publications Amendment Act of 2019.
[125] 18G(5)(b) of the Films and Publications Amendment Act of 2019.
[126] 18G(5)(c) of the Films and Publications Amendment Act of 2019.
[127] 18(F)(6) of the Films and Publications Amendment Act of 2019.

© University of Pretoria

Under the Act, on what constitutes online harms such as child pornography, online hate speech and non-consensual sharing of intimate images, is clear, precise, legitimate, necessary and proportionate. However, the FPB's mandate in regulating these online harms could be better realised by engaging with more actors. For example, the scope of online harms covered by the FPB and the regulation could be expanded but with a rights-respecting regulatory plan. This means that the FPB should be open to rethinking its regulatory powers in order to effectively implement its mandate under the Act.

H Prevention and Combating of Hate Crimes and Hate Speech Bill

The objectives of the Bill include giving effect to South Africa's obligations regarding combating prejudice and intolerance under international instruments.[128] The bill provides for an elaborate definition of hate speech which includes publication or communication of intention to be harmful or incites harm.[129] Section 4(1)(b) makes provision for online hate speech as an offence under the bill. Section 6(3) of the bill punishes the offence of hate speech by a fine or an imprisonment term of not more than three years for a first offender or conviction and up to five years with or without a fine for a subsequent offender or conviction. Section 9 of the Act vests the awareness of hate crimes and hate speech in the SAHRC and the Commission for Gender Equality.[130]

I The need for legal reform in preventing online harms in South Africa

Given the background analyses of laws in South Africa, there is need to carry out legal reforms in order to prevent online harms. For example, the provisions of sections 104 and 242 of the Criminal Procedure Act on criminal defamation, sedition, blasphemy and obscenity need to be repealed in line with international human rights standards. In addition to this, the provisions of the Cybercrime Act of 2020 on incitement to

[128] Preambular text of the Hate Crimes and Hate Speech bill.
[129] Section 4 of the Hate Crimes and Hate Speech bill.
[130] In a submission on the provisions of the bill by *Rule of Law,* a civil society organisation recommended that it should be stopped as there are already laws that adequately provide for hate speech in South Africa. They also noted that if it will be passed, the provisions of section 4 of the bill on the grounds for hate speech should be reviewed to remove subsections (b) and (c) which provides for grounds of albinism and birth. Rule of Law 'Submission to the Portfolio Committee on Justice and Correctional Services on the Prevention and Combating of Hate Crimes and Hate Speech bill, 2018' September 2021 https://www.freemarketfoundation.com/dynamicdata/documents/20210927-submission-on-hate-speech-bill.pdf (accessed12 October 2021). Helen Suzman Foundation, a civil society organisation also submitted that the bill is not necessary and also opposes the criminalisation of hate speech. Helen Suzman Foundation 'Submission in response to the Prevention and Combating of Hate Crimes and Hate Speech Bill (Gazette No 41543 of 29 March 2018)' 14 February 2019 https://hsf.org.za/publications/submissions/hsf-submission-hate-crimes-and-hate-speech-bill.pdf (accessed 14 October 2021).

© University of Pretoria

violence, threats of violence and non-consensual sharing of intimate images fall short of international human rights standards requiring that laws that limit online content must legal (precisely worded); legitimate (protect the right of others); necessary (the least intrusive means is considered) and proportionate (means of addressing the harm is not overbroad). However, the provisions of the Act on the role of the court and electronic communication service providers seem quite clear and in line with international human rights standards. As pointed out above, there are various gaps in the provisions of the Domestic Violence Act, Children's Act and the Films and Publications Amendment Act that could make the regulation of online harms in South Africa tedious and difficult. In addition to these, it will be far more useful to understand the responsibilities of ISPs under the ECTA to the public and both state and non-state actors that seeks to balance the right of the individual to freely express themselves and the rights of others to be protected from harmful content. A close reading of sections 20 and 21 of the Cybercrimes Act, applicable provisions of the ECTA and international human rights requirement on limitation of online content may be useful in this regard. There is also a need address the duplicated provisions under section 16(1) of the Cybercrimes Act of 2020 on non-consensual sharing of intimate images which reads as same under section 18(F) of the Films and Publications Amendment Act. It is also required that the relevant provisions of the PEPUDA and the Prevention and Combating of Hate Crimes and Hate Speech bill provide for a clear and precise definition of hate speech and in addition ensure that contextual and pluralistic factors are carefully considered both in the legal texts and in the adjudication before the courts.

Considering the legal and regulatory landscape for online harms in South Africa, there are scattered provisions across various legislation and the regulatory approach is mainly self-governance by industry actors. However, in order to effectively regulate online harms, not only must the laws be clear and precise, it would benefit policy development to have them all in one place – a multistakeholder sourced body of rules. This body of rules will be an exercise in bricolage – built by diverse and broader sets of stakeholders like end-users.[131] In addition to this, while self-regulation is a useful approach for platform governance, given the complex and diverse stakeholders in online speech governance, governance would have to be as diverse as possible. Therefore, the repeals and amendments necessary for legal reform on online harms in South Africa should be best focused on working towards a soft legislation that focuses on using a rights-based approach to prevent them. Such soft legislation is however not possible without proximate actors in the platform governance ecosystem.

[131] Section 4.4.1 above.

© University of Pretoria

5.3 Framing the roles of state and non-state actors in ensuring a rights-respecting platform governance in Nigeria and South Africa

So far, considering the various laws in Nigeria and South Africa and that both countries share histories with colonialism, this thesis makes four observations. One, there is hardly any evidence that colonial legacies on right to freedom expression influence the language of modern laws on online speech in South Africa as they do in Nigeria. Two, South Africa is influenced by international human rights law in preventing online harms compared to Nigeria. Three, there is a strong working relationship between policy makers, regulatory bodies and civil society in the development of laws that impact on the right to freedom of expression online in South Africa when compared to Nigeria. For example, there is hardly any proof of major submissions made by stakeholders with respect to laws that impact on the right in Nigeria. Oftentimes, bills that could impact online speech spring up before the legislature without any knowledge of its drafting by major stakeholders that could assist in improving them. Whereas, in South Africa various actors contribute to the development of laws or amendments that could prevent online harms and protect online speech.[132] Internal state actors in South Africa like the policymakers, South African Human Rights Commission (SAHRC), ISPA, FPB all have clear mandates on online speech governance and are most likely to work together compared to Nigeria's actors who have barely shown any of such collaboration.

Given these issues, and notwithstanding that South Africa may fare better in terms of applying international human rights law to online speech governance, there is need for state and non-state actors, whether internal or external in both countries, to work together in repealing, amending and enacting frameworks where necessary. The chronicling of issues from the first chapter of this thesis to this point, is to show actors in Nigeria and South Africa that online harms pose threats to the right to freedom of expression especially on social media platforms. That these harms have various negative impacts not only on online speech but on democratic development altogether. In ensuring that online harms are prevented and the right to freedom of expression is protected online in Nigeria and South Africa, internal state actors need to commit to the repeal, amendments and where necessary, enactment of laws.

[132] Alt Advisory & Research ICT Africa (n 99 above); ISPA (n 102 above); ISPA (n 103 above); AfriForum 'Submission on the Prevention and Combating of Hate Crimes and Hate Speech Bill' (2018) https://www.afriforum.co.za/wp-content/uploads/2019/09/Afriforum-Submission-B9-2018-MO-003.pdf (accessed 18 August 2021); AfriForum 'Comments on the Draft Regulations of the Internet Censorship Amendment Act' (2020) https://afriforum.co.za/wp-content/uploads/2020/08/AfriForum-commentary-Internet-Censorship-Amendment-Bill.pdf (accessed 18 August 2021); J Duncan 'Monitoring and defending freedom of expression and privacy on the internet in South Africa' (2011) https://www.apc.org/sites/default/files/SouthAfrica_GISW11_UP_web.pdf (accessed 18 August 2021).

© University of Pretoria

Therefore, without a doubt, there is a need to re-imagine platform governance in both contexts.[133] This is because the main challenge of online speech governance today still persists, which is that governments abuse their traditional regulatory legitimacy and social media companies wield unchecked powers.[134] In Nigeria, more than in South Africa, the state's traditional regulatory powers are made worse due to influence of colonial laws. For social media companies, there is hardly any transparency or accountability and where such is possible, warped foundations created by problematic legal and regulatory frameworks make them worse. What flows from the analyses above is that there is a need to do away with bad laws that primarily violate online expression, improve on the potentially good ones that currently have gaps and begin to lay the foundation for a rights-respecting platform governance system. This is not possible without engagement with actors, both internal and external who have both vertical and horizontal responsibilities in protecting human rights in Nigeria and South Africa.[135] These actors can be broadly divided into four major categories:

a. Internal state actors;
b. Internal non-state actors;
c. External state actors; and
d. External non-state actors

In this context, internal state actors[136] mainly refer to governments and their institutions while internal non-state actors refer to civil society, ISPs and other corporate actors and academia in Nigeria and South Africa.[137] External state actors also known as state-created bodies or quasi-state bodies refer to treaty-making bodies in the UN and the AU human rights systems[138] and external non-state actors refer to social media platforms, international NGOs and philanthropy organisations.[139]

---

[133] See C Papaevangelou 'The existential stakes of platform governance: A critical literature review' (2021) *Open Research Europe* 1-25.

[134] See R MacKinnon *Consent of the networked: The worldwide struggle for internet freedom* (2013) 321-325; F Akpan 'Bridging the gap between non-state actors and the state in governance: Evidence from Nigeria' (2011) 6 *International Journal of Development and Management Review* 62-71.

[135] P Willets 'Transnational actors and international organisation of global politics' in JB Baylis & S Smith (eds) *The globalisation of world politics* (2001) 35 -383; N Nasiritousi *et al* 'Normative arguments for non-state actor participation in international policymaking processes: Functionalism, neocorporatism or democratic pluralism?' (2016) 22 *European Journal of International Relations* 920-943.

[136] N Santarelli 'Non-state actors' human rights obligations and responsibility under international law' (2008) 15 *Revista Electronica De Estudios Internacionales* 1-10.

[137] Y Ronel 'Human rights obligations of territorial non-state actors' (2013) 46 *Cornell International Law Journal* 21-47.

[138] B Kabumba 'Soft law and legitimacy in the African Union: the case of the Pretoria Principles of Ending Mass Atrocities Pursuant to Article 4(H) of the AU Constitutive Act' in O Shyllon (ed) *The Model Law on Access to Information for Africa and other regional instruments: Soft law and human rights in Africa* (2018) 167.

[139] N Tusikov 'Transnational non-state regulatory regimes' in P Drahos (ed) *Regulatory theory: Foundations and applications* (2017) 339-354.

© University of Pretoria

Getting these actors to work collaboratively on the prevention of online harms and ensuring a rights-respecting platform governance would be an onerous task. This is because online harms are global in reach even though their impacts are local and further exacerbated by various factors.[140] For example, while the governments are complicit in the violation of the right to freedom of expression online, they are still needed in the implementation of solutions that would help prevent online harms. This is because they still retain the formal legitimacy of administering policies within their various contexts.[141] Social media platforms as non-state actors are also important in the value chain of an effective platform governance due to their roles as new speech governors.[142] The civil society also plays an important role in advocacy, objective analysis and monitoring of how best to govern the future of online speech. However, these actors cannot work in silos and neither can they continue to work within their traditional influences or self-interests.[143]

This regulatory challenge is the major reason why even though the human rights-based approach to platform governance is desirable, it must be creatively designed, normatively sound and generative in its processes. Such creativity would include rethinking traditional roles in governing online expression; it would be normatively sound by anchoring these roles on international human rights standards and it would be generative in its processes by using soft laws first to regulate online expression before the use of hard laws. By generative, it would also involve as many stakeholders as possible that seek to learn from their errors and successes to build stronger governance systems. Simply put, the generative approach is the multistakeholder approach in motion, that is rights-based, stakeholder-driven, open, transparent and

---

[140] Internet & Jurisdiction Policy Network 'Toolkit Cross-border Content Moderation' March 2021, https://www.internetjurisdiction.net/uploads/pdfs/Internet-Jurisdiction-Policy-Network-21-104-Toolkit-Cross-border-Content-Moderation-2021.pdf (accessed 15 August 2021).

[141] J York 'The global impact of content moderation' 7 April 2020 *ARTICLE 19* https://www.article19.org/resources/the-global-impact-of-content-moderation/ (accessed 20 August 2021); M Karanicolas 'Moderate globally, impact locally: A series on content moderation in the Global South' 5 August 2020 *Yale Law School* https://law.yale.edu/isp/initiatives/wikimedia-initiative-intermediaries-and-information/wiii-blog/moderate-globally-impact-locally-series-content-moderation-global-south (accessed 20 August 2021); OHCHR 'Moderating online content: fighting harm or silencing dissent' 23 July 2021, https://www.ohchr.org/EN/NewsEvents/Pages/Online-content-regulation.aspx (accessed 20 August 2021); T Gillespie *et al* 'Expanding the debate about content moderation: scholarly research agendas for the coming policy debates' (2020) 9 *Internet Policy Review* 24.

[142] K Klonick 'The new governors: the people, rules and processes governing online speech' (2018) 131 *Harvard Law Review* 1599-1670.

[143] Kaye referred to just the governments shedding their traditional areas of powers but in order to ensure a generative approach, all proximate actors including the government must make compromises on their interests and influence in order to prevent online harms and protect online free speech. See Kaye (n 49 above) 93.

© University of Pretoria

consensus-based.[144] The most important responsibilities for both state and non-state actors using the generative approach is to optimise the use of international human rights law to achieve legitimacy, representation and consensus for online speech governance.[145]

In terms of making the generative approach work, this role might create challenges for power relations. This is because most governments often seek to maintain an upper hand in policy deliberations and such understanding of their powers might prove difficult in light of governing platforms.[146] However, online harms will not regulate themselves and neither will online speech be protected unless governments collaborate with other actors.[147] In terms of representation, there are resources necessary for the development of a multistakeholder system and the inclusion of under-represented groups such as vulnerable persons and netizens.[148] However, creating a system for determining how these groups, persons and netizens contribute to the platform governance debate might be a monumental challenge. This challenge may be resolved by ensuring that deliberations on inclusion are had by more marginalised and vulnerable groups. Perhaps the most painstaking responsibility for stakeholders is achieving consensus. Often, it is the ability of the majority of the stakeholders reaching an agreement to commit to the use of international human rights law, critical logic and reasoning in their processes.[149] One of the practical ways of seeking consensus is ensuring clear rules and modes of discussion which must come from mutual respect and clear objectives. Consensus might therefore create challenges for the legitimacy of the approach if not managed properly.

Moving towards designing for such a system that accommodates this approach, Jørgensen notes that platform governance systems across the world are fractal and complex but it would be necessary to have an 'authoritative human rights guidance for the major online platforms.'[150] Gillespie proposes a law, which could also be understood as Jørgensen's authoritative human rights guidance by stating that:

---

[144] LE Strickling & JF Hill 'Multi-stakeholder governance innovations to protect free expression, diversity and civility online' in E Donahoe & FO Hampson (eds) *Governance innovation for a connected world: Protecting free expression, diversity and civic engagement in the global digital ecosystem* (2018) 45 - 50 https://apo.org.au/sites/default/files/resource-files/2018-11/apo-nid203391.pdf (accessed 20 August 2021).

[145] As above.

[146] Principle 1(2) of the revised Declaration.

[147] Principle 16(3); DM Chirwa 'The doctrine of state responsibility as a potential means of holding private actors accountable for human rights' (2004) 5 *Melbourne Journal of International Law* 250.

[148] Strickling & Hill (n 144 above) 49.

[149] As above.

[150] R Jørgensen 'Human rights and private actors in the online domain' in MK Land & JD Aronson (eds) *New technologies for human rights law* (2018) 243-269.

© University of Pretoria

> … it is high time to reconsider the responsibilities of platforms. This should include crafting a new principle of law tailored for social media platforms, not borrowed whole cloth from a law designed for ISPs and search engines. It should include articulating normative expectations for what platforms are – legally, culturally, and ethically – not just passes for what they don't have to be.[151]

In agreement with both Jørgensen and Gillespie, and noting Jørgensen's strong objection to soft law, so far, only international human rights law as customary law offer strength to the legitimacy, representation and consensus of state actors and non-state actors in the design of online speech rules. So far, only a set of internationally set rules would cater for both their concerns. State actors, whether internal or external, must consider the trans-border and complex nature of the Internet and social media platforms, ensure compromises of traditional powers where necessary and collaborate to address basic norms on online speech governance through international organisations like the UN and the AU's treaty-making bodies. Due to their design, the UN and the AU's treaty-making bodies are not only made up of states who have the most traditional legitimacy but also have the resources and opportunity for inclusion. The UN and the AU treaty-making bodies also have the opportunity of achieving consensus among stakeholders and this may achieve two major things.

First, such an approach would recognise the nature of the Internet as a globally free, open and interoperable network which also includes social media platforms.[152] It will provide specific provisions, typologies, mandates and responsibilities with respect to protecting online speech using the multistakeholder approach. For example, both institutions could develop a 'soft law' on online speech governance which will offer a more applicable and descriptive system of rules for countries.[153] A meaningful point to start from is through ongoing efforts by the civil society like the BSR report and the SMCs. Second, such an approach continues to maximise both institutions as the most formal and diverse system of global governance on international human rights law and its application. This is because, through the inter-operable relationship between various stakeholders like state actors, NHRIs, private actors and the civil society that both institutions provide, it could begin to design more timely and creative solutions to thorny global challenges such as online speech governance.

Practically, the approach can be achieved by first drawing up Social Media Charters (SMChs) internationally and nationally which would be the precursor for a hard national law like an Online Harms Act as a primary, regulatory and substantive law in local contexts. The SMChs could be used to accommodate diverse consultations to shape a rights-respecting platform governance approach which can be best achieved through national 'soft laws.' These SMChs will lean in on the various inputs on the SMC's models in terms of structure while using the BSR's report and other

[151] T Gillespie 'Regulation of and by platforms' in J Burgess, T Poell & A Marwick (eds) *SAGE handbook of social media* (2017) 254.
[152] Section 4.2 above.
[153] Gillespie (n 151 above).

© University of Pretoria

multistakeholder driven normative processes as substantive rules. It would then be the failures, lessons and opportunities that these Charters provide that will give considerations for the enactment of an Online Harms Act – a rights-centred law with the force of law to design responsibilities for duty-bearers and right-holders in a national law.

5.3.1 A generative rights-based approach to platform governance in Nigeria and South Africa

Sourced from the United Nations' Guiding Principles on Business and Human Rights (UNGPs), the human rights-based approach may be generally referred to as the need to 'Protect, Respect and Remedy' that is divided alongside responsibilities for actors.[154] For example, the State has the duty to protect against human rights abuses in their territory, companies must respect human rights by not violating rights and remedying wrongs in which they are involved and both actors have responsibilities of ensuring access to remedies.[155] In achieving the generative approach with this background for both the SMChs and the law, various actors will have specific roles which must include three pillars of the rights-respecting approach to platform governance: substance dimension, process dimensions and procedural dimension.[156]

A Substance dimension

An important feature of the substantive dimension to an approach based on human rights is basing rules that govern online speech on international human rights law.[157] It applies to both national legal reform and self-governance by platforms. The first most important step towards applying the substantive dimension to preventing online harms by state actors is through theme-specific national legal reforms. This kind of legal reform is necessary in that there are wider theme-based concerns with respect to the impacts of online harms. Such include the relationship between gender and online harms, the role of international organisations and online harms, political advertisement and online harms and many other concerns. However, the legal reform being referred to in this research is that which is specific to safeguarding the right to freedom of expression online while preventing online harms in Nigeria and South Africa. State actors will be required to amend existing problematic laws that have been identified above, provide for criminal offences for harms that are illegal and harmful and ensure that all these laws are in line with international human rights law while social media

[154] See OHCHR 'Guiding principles on business and human rights' https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf (accessed 24 August 2021).
[155] As above.
[156] B Sander 'Freedom of expression in the age of online platforms: The promise and pitfalls of a human rights-based approach to content moderation' (2020) 43 *Fordham International Law Journal* 939 - 1006.
[157] Sander (n 156 above) 970.

© University of Pretoria

will be also be required to amend their policy documents on online speech governance and remove harmful content that are not illegal in line with international human rights law.[158] As a result, both categories of actors are on the same page with respect to the centrality of international human rights law in their governance approaches and as a result will be able to effectively combat online harms.

The application of the generative approach here is that they are non-traditional and theme-based. It is non-traditional because the traditional expectations of state actors to control governance will be required to be flexible for the application and implementation of international human rights law.

Such legal reform is also theme-based because it will be geared specifically towards solving the challenge posed to the right to freedom of expression through online harms. It is not the role of a rights-based approach to preventing online harms and promoting online privacy, association, assembly or even guidelines on online political advertising; it is a legal proposal targeted towards combating the harms posed to free speech as a result of online harms. Such legal reform is therefore generative in that while the rules are clear, context will continue to pose a challenge and as a result of such dynamism, online speech governance cannot afford to be a hard and unyielding system of rules. This system of rules is what will require the development of a soft law mechanism that would guide the path towards a substantive hard law on online harms in both countries. It is this kind of legal reform, generative in its goal and self-learning in its processes, that would bring both the state and non-state actors together on the need to have a human rights-based approach to online content governance.

B Process dimension

The process-based dimension is broadly divided into three including carrying out human rights impact assessments, taking effective measures to mitigate harms and 'radical transparency.'[159] These are further divided into rule-making, decision-making, content and advertising and regulatory compliance.[160] The first part requires that social media platforms must open up on how they make their rules to ensure more inclusion. The second part focuses on more clarity on how these platforms apply these rules while the third aspect focuses on sharing more information on the content and advertisement that online users view. For example, that could include who paid for what advert and whether it has any pre-assessed human rights implications. The last part on regulation requires that when platforms are faced with problematic national

[158] J Woodhouse 'Regulating online harms' House of Lords 21 August 2021, https://researchbriefings.files.parliament.uk/documents/CBP-8743/CBP-8743.pdf 31- 32 (accessed 23 August 2021).
[159] Sander (n 156 above) 988.
[160] Sander (n 156 above) 990.

© University of Pretoria

laws, higher consideration must be given to international human rights law as much as possible.

Using an example of how to prevent online harms and promote freedom of expression for example, according to Azelmat, for social media platforms to prevent online gender-based violence, they must focus less on digital colonialism and tool-focused solutions and more on root causes, prevention, improved transparency and accountability.[161] Azelmat's argument demonstrates that a human rights based approach to combating online gender-based violence is possible especially on the part of social media platforms. However, as highlighted earlier, online gender-based violence is both harmful and illegal and as a result constitutes a crime which is primarily within the powers of state actors to prevent.

Therefore, a generative approach to combating online gender-based violence for example would be to coordinate actors on various forms of online gender-based violence and the applicable methods of sanctions online and offline. While social media platforms have seemed most feasible as the implementing actor for a human rights-based approach, state actors should also be able to develop laws and policies that could help prevent it offline. Therefore, while social media platforms continue to directly apply their reworked policies to the online gender-based violence with more transparency and accountability, states are also able to apply rights-respecting laws and not violate the right to freedom of expression as they have been doing.[162] In this context, it would be necessary to draw up a typology of transparency and oversight responsibilities in laws and social media community guidelines.

C Procedure-remedial dimension

The procedural-remedial aspect of this approach focuses on how wrongs can be righted by establishing or participating in 'an effective operational-level grievance mechanisms for individuals and communities who may be adversely affected.'[163] Here, actors like the NHRIs and social media platforms are required to ensure the development of a system whereby wrongs are clear and responsibilities for righting them are accessible and effective. In this context, it would be necessary to draw up a typology of procedural guarantees and remediation applicable in laws and social media community guidelines.[164]

---

[161] M Azelmat 'Can social media platforms tackle online violence without structural change?' 17 August 2021 *Association for Progressive Communications* https://genderit.org/node/5511 (accessed 23 August 2021).
[162] Karekwaivanane & Msonza (n 19 above).
[163] Principle 29 of the UNGP.
[164] See E Goldman 'Content moderation remedies' (2021) *Michigan Technology Law Review* 1-76.

© University of Pretoria

Therefore, the goal of the generative approach as an effective approach should be such that it allows two major aspects of governance: the normative standard and the operational communication for enforcement. With respect to these aspects, a set of rules or norms that would allow for preventing online harms and promoting online speech must consider: [165]

a. clear definitions of what constitutes online harms, why[166] and the nature of sanctions applicable to each;[167]
b. the categories of persons affected;[168]
c. the legitimate aim sought to be achieved with such sanctions which must be 'concrete threat to life, limb and liberty of person';[169]
d. adequacy of such aim in neutralising the harm;
e. least intrusive measures;[170]
f. user notification;[171]
g. right to remedy;[172]
h. transparency;[173]
i. oversight systems;[174]
j. regulators working on a dynamic model;[175]
k. opening up mode of communication between subsystems;[176]
l. using the feedback loop to improve the dynamic model.

In order to ensure that this set of rules are developed generatively and rights-complaint, state and non-state actors, whether internal or external have specific roles below.

[165] S Benesch 'Proposals for improved regulation of harmful online content' https://dangerousspeech.org/wp-content/uploads/2020/06/Proposals-for-Improved-Regulation-of-Harmful-Online-Content-Formatted-v5.2.1.pdf (accessed 24 August 2021).
[166] Sander (n 156 above) 977.
[167] Sander (n 156 above) 971.
[168] Sander (n 156 above) 979.
[169] Sander (n 156 above) 973, 1003.
[170] Sander (n 156 above) 984, 988.
[171] Sander (n 156 above) 1001.
[172] Goldman (n 164 above).
[173] Kaye (n 49 above) 89, 91.
[174] Kaye (n 49 above) 90.
[175] Role for public institutions and other proximate actors. See Kaye (n 49 above) 92; Sander (n 156 above) 998.
[176] Kaye (n 49 above) 90.

© University of Pretoria

5.3.2 The roles of internal state actors

A Government

In the context of this thesis, the government refers to the executive, legislature and the judiciary. In terms of policy implementation roles carried out by the executive, its ministries, institutions or agencies established by law to perform specific roles that have both remote and immediate impact on the relationship between technologies and human rights. This is because these ministries, institutions or agencies have specific mandates either towards the use of technologies and the protection of human rights or both and therefore are in a better position to not only advise other state actors but will be able to contribute meaningfully to the generative process of platform governance.

For example, in Nigeria, the NCC should be in talks with the NHRC to consider the various impacts its laws and their implementation have on human rights. In addition to this relationship, other institutions whose mandates fall explicitly or implicitly on the regulation of technologies and human rights development should be at a table to rethink reform and chart a path forward. Here, the traditional responsibilities of the government will not be to only make, implement, or interpret laws but to collaborate with other proximate actors in the value chain of ensuring effective online speech governance in Nigeria. For South Africa, the FPB, ICASA, ISPA and other government actors must have a sit-down with the SAHRC on the various opportunities under international human rights law to mainstream rights-respecting approaches to online content regulation. Some of these approaches include training and workshops that update the knowledge of each of these actors on the tasks at hand on preventing online harms and protecting online free speech.

B National Human Rights Institutions

National Human Rights Institutions (NHRIs) are state-regulated public bodies provided for in the Constitution or legislation that ensures the promotion, protection, advisory, monitoring, coordination and cooperation on human rights development.[177] They are not under the direct control of the government even though they are largely funded by them and neither are they non-governmental institutions as they are created by an extant law. Therefore, it is trite to refer to NHRIs as both state and non-state actors in order to ensure that their functions are performed and their powers are exercised independently. In many instances, they have been the most frontal state institution with respect to human rights development, the bridge between state and non-state actors and also become the link between the international human rights system and

[177] OHCHR 'National human rights institutions: History, principles, roles and responsibilities' (2010) https://www.ohchr.org/Documents/Publications/PTS-4Rev1-NHRI_en.pdf (accessed 24 August 2021).

© University of Pretoria

its application in local contexts.[178] Given the unique placement of NHRIs in human rights development and to ensure an effective platform governance that is generative, each NHRI in Nigeria and South Africa have roles to play.

In both countries, it will be necessary to funnel the provisions of the content governance by the BSR for its substantive and process-based elements and SMC for its proposed structuring of institutions in governing online speech. By doing so, the NHRIs perform the roles of a 'regulator of online speech' working on a dynamic model of interaction between internal and external non-state actors on platform governance.[179] This model also allows for inclusive deliberations by opening up modes of communication between these actors before, during and after the development of such a model. Due to the dynamism of digital technologies and the harms they pose, the NHRIs could also use the feedback loop to improve the dynamic model.

To begin with, for Nigeria, the National Human Rights Commission Act establishes the National Human Rights Commission (NHRC). Among other responsibilities, the functions of the NHRC includes the monitoring of implementation of major treaties Nigeria has obligations to comply with[180] and to:

> Undertake studies on all matters pertaining to human rights and assist the Federal, State and Local Governments, where it considers it appropriate to do so, in the formulation of appropriate policies on the guarantee of human rights… Examine any existing legislation, administrative provisions and propose bills or bye-laws for the purpose of ascertaining whether such enactment or proposed bill or bye-laws are consistent with human rights norms… Carry out all such other function as are necessary or expedient for the performance of these functions under the Act.[181]

This shows that the responsibility of carrying out the generative platform governance lies with the NHRC because it is more positioned as one of the most subject-matter relevant agencies with respect to human rights protection in Nigeria. In addition to this role, its dual role both as a state and non-state actor in the implementation of human rights makes it the most viable formal institution to begin the process of realising context-based and effective platform governance in Nigeria. However, before performing this responsibility, the NHRC must realign its goals, resources and

---

[178] S Lagoutte 'The role of state actors within the national human rights system' (2019) 37 *Nordic Journal of Human Rights*, 177-194; SLB Jensen *et al* 'The domestic institutionalisation of human rights: An introduction' (2019) 37 *Nordic Journal of Human Rights* 165-176.

[179] Section 5(1)(f)(g) of the NHRC Act and Section 13 of the SAHRC Act. Given the impacts of various digital rights violations like the right to freedom of expression online as described under section [Refer to chapter 4], drawing up a Social Media Charter might be regarded as incidental, necessary, conducive, or expedient for the performance of the functions of both Acts. This also compares with the provisions on Principle 43(1) of the revised Declaration on its implementation.

[180] Preambular text of the NHRC Act.

[181] Section 5 of the NHRC Act.

© University of Pretoria

thematic areas to accommodate the protection of digital rights.[182] This re-alignment can begin by first including 'digital rights' as the NHRC's twentieth thematic area of focus in order for its 'effective performance and result oriented approach' to be relevant in the digital age.[183]

In realising such a plan, the NHRC is empowered to carry out an extensive and elaborate review of existing and proposed policies that have been identified in this research in order to bring them in line with international human rights law.[184] The NHRC should also sponsor amendment bills at the legislature in order to effect the legal reforms necessary as empowered by the Act.[185] After this process has been completed, the NHRC will undertake studies and report on action that should be taken in preventing online harms and promoting online free speech.[186] This should culminate into the process of drafting a Social Media Charter.

In carrying out these major functions above, it would take the lead through internal actors and consult widely with various proximate stakeholders including the executive and its agencies, the legislature and its staff, and the judiciary and its supporting agencies, ISPs, local civil society, academic institutions etc.[187] This consultation must include civil society actors, ISPs, representatives of social media platforms, academia and researchers. This Charter will determine the powers, functions and roles of various stakeholders in activating the various aspects above.

In South Africa, the NHRI, the South African Human Rights Commission (SAHRC) is established as one of the six chapter nine institutions under the Constitution.[188] The Constitution provides for the responsibilities and powers of the SAHRC to include promotion of respect, protection, development and attainment of human rights. This role also includes monitoring and assessing the observance of human rights in South Africa. It also provides for the powers of the SAHRC, including investigation, reporting, redress, research and education on human rights should be regulated by legislation. Section 13(1)(a)(b) of the South African Human Rights Commission (SAHRC), 2013

[182] A Fund can be co-created by state and non-state actors to fund the digital rights program at the Commission under section 13 & 15 of the Act.
[183] NHRC, 'What are human rights' https://www.nigeriarights.gov.ng/about/nhrc-mandate.html (accessed 15 August 2021).
[184] One of the objectives of the NHRC in its draft National Action Plan from 2021 to 2025 is the 'Protection of citizens against misinformation, disinformation and fake news. This objective needs to be expanded in scope and objectives to accommodate other forms of online harms. See 'Draft National Action Plan: 2021-2025' 13 August 2021 https://www.nigeriarights.gov.ng/files/nap/NAP%20for%20final%20Review%20July%209%20(3)-converted.pdf (accessed 15 August 2021) 60.
[185] Section 5(1)(d)(g)(h)(k)(n)(o) of the NHRC Act.
[186] Section 5 (j) of the NHRC Act.
[187] Section 5(f)(g)(h) of the NHRC Act.
[188] Section 184 of the Constitution of South Africa, 1996.

© University of Pretoria

Act provides for the powers of the SAHRC. These provisions empower the SAHRC to make recommendations to state actors, undertake studies, collaborate with other bodies with similar objectives, review government policies on human rights issues etc. These powers similarly position the SAHRC like its Nigerian counterpart to take on the fight for the prevention of online harms and protection of online speech in South Africa.

The SAHRC must therefore take a decisive lead on the development of an effective platform governance for South Africa. This is because online harms primarily pose threats to the enjoyment of online freedoms which also fall within the purview of the SAHRC.[189] While the SAHRC has already taken steps in terms of drafting a Social Media Charter, this Charter will have to be more inclusive in order to ensure an effective platform governance approach in South Africa.[190] Through its core operational programmes, the Commission should carry out more collaborative research with other stakeholders and increase its activities with respect to drawing up a multistakeholder Social Media Charter and in the end, an Online Harms Act. In addition to this, in its Strategic Plan for 2015-2020, the Commission includes its role in ensuring rights-respecting policies on information technologies.[191] In its next Plan, the Commission should ensure that not only online harms are catered for, the larger scope of digital rights should be included.[192]

5.3.3 The roles of internal non-state actors

A Civil society

Local civil society actors like NGOs who work on technologies, human rights and their various intersections have a major role to play in ensuring a rights-respecting platform governance in both countries. In Nigeria, for example, the most significant legislative effort so far on digital rights of which includes online content regulation has been the Digital Rights and Freedom Bill. The bill seeks to provide both positive and negative duties of state actors in the promotion and protection of digital rights in Nigeria. The bill, which was majorly championed by civil society, also identified the NHRC as the

[189] Section 13 of the SAHRC Act.
[190] SAHRC recently made a call for developing a Social Media Charter which has now advanced to a draft. See here for the call: SAHRC 'Terms of Reference: Develop a draft Social Media Charter for the South African Human Rights Commission https://www.sahrc.org.za/home/21/files/Terms%20of%20reference%20-%20Social%20Media%20Charter%20-%20Final.doc (accessed 15 August 2021).
[191] SAHRC 'Revised strategic plan for the fiscal years 2015 to 2020' https://www.sahrc.org.za/home/21/files/SAHRC%20Revised%20Strategic%20Plan%202015%20-%202020.pdf 18 (accessed 15 August 2021).
[192] SAHRC (n 191 above) 22.

© University of Pretoria

implementing government agency. However, the bill, which has been rejected by the President, falls short of major needs for digital rights in two ways.[193]

First, the bill needs to be unbundled into more specific aspects of digital rights. Currently as the bill stands, various aspects of digital rights capable of being broken down into more specific laws are lumped together without any clear and specific enforcement and implementation steps. Second, there is no clear evidence of thorough engagement with the NHRC especially in terms of using the bill to key into its roles as the most frontal human rights institution in Nigeria. One clear area for synergy between the bill and the NHRC is to create a digital rights fund to be managed by the NHRC, part of which will be used to take care of NHRC's resources needed for its digital rights mandate, including those on online content governance. In this instance, the civil society plays a major role first by ensuring legislative reform and implementation. The civil society has to work with the NHRIs and other actors in ensuring legal reform of laws and policies on online speech governance by providing advisory and contextually relevant policy research. In terms of implementation, civil society will co-monitor the development of policies with the NHRIs on online speech governance.

B Internet service providers

ISPs are also involved in ensuring an effective platform governance. One of the two ways they can be more involved asides collaborating with other actors is to commit to accountability and transparent processes.[194] In terms of accountability, ISPs should require a judicial review of orders to block social media platforms. In addition to this, ISPs must also require human rights-based impact assessments (HRIAs) from state actors with respect to blocking social media platforms.[195] It might be difficult to fulfil such a request where there are 'concrete threats to life, limb or liberty of a person.'[196] In such an instance, an ISP may go ahead to grant such a request by the state actor but must demand a timeous and retroactive accountability – demand that the HRIA and judicial review is submitted to them. With respect to transparency, ISPs must publish government requests for blocking of online content bi-annually and report to

[193] V Ekwealor 'Nigeria's president refused to sign its digital rights bill, what happens now?' 27 March 2019 *Techpoint* https://techpoint.africa/2019/03/27/nigerian-president-declines-digital-rights-bill-assent/ (accessed 20 August 2021).
[194] ARTICLE 19 & EFF (n 75 above).
[195] According to Jørgensen *et al*, HRIAs can be a useful tool for infrastructure providers like ISPs. However, such Assessment must include 'determining responsibility for human rights harms, proposing rights-respecting solutions to the current governance gap and other challenges' R Jørgensen *et al* 'Exploring the role of HRIA in the information and communication technologies (ICT) sector' in N Götzmann (ed) *Handbook on Human Rights Impact Assessment* (2019) 13.
[196] Geneva Academy 'Defending the boundary: constraints and requirements on the use of autonomous weapon systems under international humanitarian and human rights law' May 2016, page 63 https://www.geneva-academy.ch/joomlatools-files/docman-files/Briefing9_interactif.pdf (accessed 20 August 2021).

© University of Pretoria

the legislative committees on human rights annually on the same issue in both countries. These roles should be debated and the product of such debate included in the SMChs and the Online Harm Acts.

C Academia

Academia also has a role to play in ensuring an effective platform governance in Nigeria and South Africa. This is because the impacts of online harms are multi-faceted and cannot be considered only from legal lenses. There is constant need to advance academic literature on how these harms can be prevented and policy developed in both settings.

5.3.4 The roles of external state actors

External or quasi state actors also known as state-created bodies are involved in the promotion and protection of human rights. These actors are mainly the UN human rights system and AU human rights system. For example, within the UN, the standard-setting role of the Special Rapporteur will be to clarify the various aspects of the twelve principles highlighted above with respect to online harms. Within the AU human rights system, the African Commission through the Special Rapporteurs (SPs) on Freedom of Expression and Access to Information will also do the same to contextualise the work of the UN SP regionally.

The role of external state actors is to take the lead among external actors. This standard-setting responsibility on online harms best falls on the shoulders of these two external actors for two main reasons. First, the complementary and collaborative working nature of the UN SPs and regional SPs provide a great avenue to ensure an inclusive development of norms with respect to online harms because of their roles in standard-setting for the right to freedom of expression.[197] Second, both systems have a unique relationship with the main internal state actor – NHRIs in the application of the promotion and protection of human rights not only in Nigeria and South Africa but across the region at large. It is this relationship that should be leveraged first by the external state actors and the NHRIs.[198]

5.3.5 The roles of external non-state actors

A Social media platforms

Social media platforms have at least six important roles to play in ensuring effective platform governance in Nigeria and South Africa. Major social media platforms such

[197] Section 2.4.3 above.
[198] OHCHR (n 154 above).

© University of Pretoria

as Facebook, Twitter, Google and others all have the responsibility to collaborate with other actors.[199] One, platforms must ensure that they bring their policy guidelines on online harms up to international human rights standards.[200] This is similar to having a Social Media Charter but in this case, such 'Charter' would apply internally as the basis for making content moderation decisions.

Two, the role of social media platforms should include working directly with the NHRI and civil society in each country directly in order to apply human rights to their decisions on online speech governance from each country. Three, they must also work with the NHRIs, civil society and other actors to publish annual reports on the requests for take-downs, those approved, those denied and the reasons for each of these decisions. Four, collaborate with both internal and external non-state actors with respect to online speech governance and with internal state actors in order to have diverse perspectives. Five, social media platforms should provide resources to support digital rights projects including effective online speech governance to internal state and non-state actors.[201] Six, for social media platforms to build more trust in African contexts, they must hire locally. Such hires must have cultural knowledge of the context and rights-respecting content moderation skills to make the necessary decision to balance protecting online expression while preventing harms.

B International NGOs and philanthropy organisations

International NGOs have the role of teaming up with NHRI's, UN and AU treaty-making bodies, and local civil society actors. This relationship is necessary in order to funnel global trends and dynamics into the development of an effective online speech governance in African countries. For example, the role of these NGOs has been seen in the development of research such as the BSR report, the Global Network Initiative's (GNI) Content Regulation and ARTICLE 19's proposal on Social Media Councils. These research outputs have the potential to guide both state and non-state actors in the application of online content governance not only in Nigeria and South Africa, but across the globe.

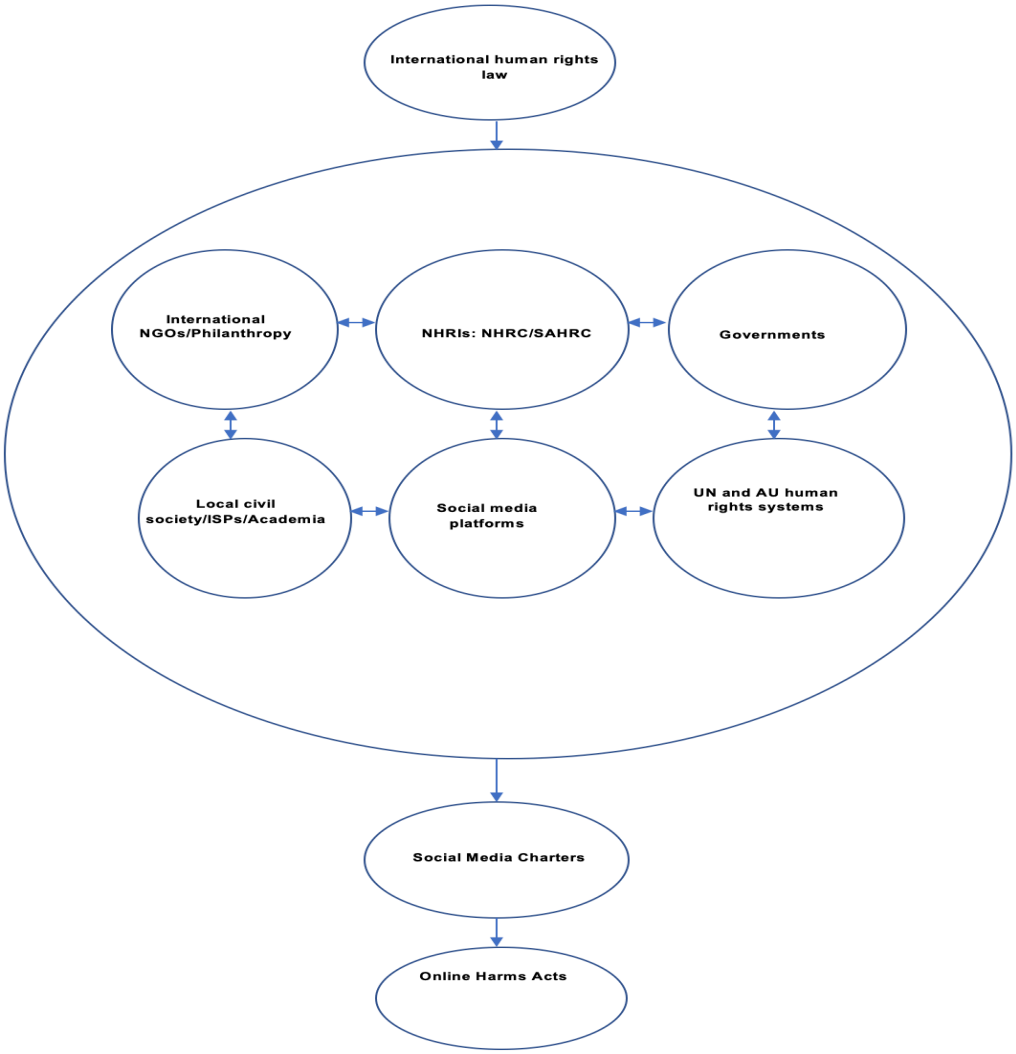
Philanthropy organisations in the social justice sector should therefore fund more projects on digital rights and online speech governance.[202] In terms of prioritising

[199] A Callamard 'The human rights obligations of non-state actors' in RF Jørgensen (ed) *Human rights in the age of platforms* (2019) 199.
[200] Kaye suggests doing this by social media platforms committing to 'industry-wide oversight and accountability.' Kaye (49 above).
[201] D Kaye 'Ethiopia, the scourge of 'hate speech' & American social media' 9 December 2019 https://dkisaway.medium.com/ethiopia-the-scourge-of-hate-speech-american-social-media-952c9228e21c (accessed 2 August 2021).
[202] A program or strategy on digital rights may be developed by both the NHRC and the SAHRC called the 'Digital Rights Project.' This project will focus broadly on the impacts of technologies on human

© University of Pretoria

funding for digital rights and online speech governance, funding from philanthropy organisations with track records of integrity, flexible donor conditions, and non-partisanship should be prioritised. This prioritisation is in order to ensure more resource sustainability and less dependence on private sector funding which may have ethical strings attached. However, in order to resolve these ethical strings private sector funding may have on such resources, the extent to which the private entity has the track record similar to that of philanthropy organisations would determine prioritising them for such support.

Illustration 2: A generative approach to platform governance in Nigeria and South Africa



rights with specific focus areas. Platform governance would be designed as one of the sub-projects and one of the objectives of the sub-project would be to ensure a rights-respecting regulation of online harms. This program or strategy can also be created conversely, that is, the platform governance sub-project kicks off first, before the broader Digital Rights Project. The major function of the sub-project would be to monitor the legal reforms and other measures necessary with respect to regulating online harms and promoting online speech in both countries.

© University of Pretoria

5.4 Conclusion

This chapter set out to identify the roles of state and non-state actors in ensuring a rights-respecting approach to platform governance in Nigeria and South Africa. In doing this, it identified the legal and regulatory provisions on online harms, their compliance with international human rights law and the need to embark on a legal reform that ensures the repeal, amendments and enactment of rights-respecting laws on online harms in Nigeria and South Africa.

It further identified that in order to carry out these reforms, internal and external state and non-state actors will have to play specific roles anchored on international human rights law that is generatively achieved. For internal state actors, NHRIs are encouraged to take leadership due to their unique role in the promotion and protection of human rights nationally and internationally. For external state actors, the UN and the AU are both encouraged to lead through their norm-setting mandates.

Civil societies, both local and international together with other stakeholders like social media platforms are also encouraged to team up across various sectors to draw up rights-based rules and an operational system that protects online speech. In summary, this chapter shows that it is possible to govern social media platforms, but such governance must be generative. This generativity must however be anchored on international human rights law and operationalised by various actors performing their specific responsibilities in ensuring that online harms are prevented and the right to freedom of expression online is protected in Nigeria and South Africa. It proposes that a practical output of such compliance would be building a dynamic regulatory matrix (multistakeholder governance) that first adopts a soft law (SMChs) before a hard law (Online Harms Act or Law) in both countries. The next chapter summarises this thesis, its major findings, further areas of research and concludes.

© University of Pretoria

CHAPTER SIX: SUMMARY OF FINDINGS, FURTHER AREAS FOR RESEARCH AND CONCLUSION

6.1 Introduction

Social media platforms may be said to be as recent as the attempts to govern them. These attempts at governance have however been difficult and complex. In African countries, this difficulty and complexity are exacerbated in part by problematic legal and regulatory provisions on the right to freedom of expression and the unchecked powers of social media platforms. The need for governance has therefore not been more urgent especially with the rise of online harms. These harms which have been majorly identified as information disorder and targeted online violence now need actors to re-imagine their roles in order to effectively regulate these harms. In illustrating these roles in Nigeria and South Africa, this thesis proposed a major research question on how a rights-respecting approach can be applied to ensure the prevention of online harms and protection of online expression on social media platforms in Nigeria and South Africa. In answering this question, this thesis further considered the following four research sub-questions:

a. To what extent have the theoretical perspectives and recent normative developments on the protection of the right to freedom to expression online been implemented in Africa?
b. What are the impacts of online harms on the right to freedom of expression online in Africa?
c. How can rights-respecting platform governance be used to prevent online harms and protect freedom of expression online on social media platforms?
d. What are the roles of state and non-state actors, internal and external, in ensuring an effective rights-based approach to social media platform governance in Nigeria and South Africa?

The first chapter set the stage for this thesis by providing a snapshot. The second chapter focused on the first sub-question on the extent theoretical perspectives and recent normative developments on the protection of the right to freedom to expression online have been implemented in Africa. The third chapter focused on the second sub-question on the impacts of online harms on the right to freedom of expression online in Africa. The fourth chapter examined the third research sub-question on how a rights-respecting platform governance could prevent them and promote online expression. The fifth chapter then considered the fourth research sub-question on the roles of internal and external state and non-state in ensuring an effective rights-based approach to social media platform governance in Nigeria and South Africa. This chapter provides summaries of these chapters and their findings, identifies further areas for research and concludes that the international human rights law system must be complied with in order to prevent online harms and protect the right to freedom of

© University of Pretoria

expression online. However, such compliance must be generatively applied – a dynamic regulatory matrix (multistakeholder governance) that first adopts a soft law (SMChs) before a hard law (Online Harms Act or Law) in both countries.

6.2 Summary of key findings

The impacts of digital technologies on human life, whether good or challenging, will continue to be relevant conversation for a long time. Specifically, the impacts of social media platforms as examples of these technologies will also continue to be a legal and regulatory debate for a while. This thesis considered an example of these impacts – the impacts of online harms on the right to freedom of expression online and the roles of internal and external state and non-state actors in preventing these harms and promoting freedom of expression online through a rights-respecting platform governance in Nigeria and South Africa. In identifying these impacts and roles of actors in ensuring a rights-respecting platform governance, it further broke down the thesis into four sub-theses that led to the various findings that are summarised below.

6.2.1 The implementation of theoretical perspectives and recent normative developments on the protection of the right to freedom to expression online in Africa

The second chapter considered the first sub-question on the extent of implementation of theoretical perspectives and recent normative developments on the protection of the right to freedom of expression online in African countries. In answering this sub-question, three questions were posed to examine how major theoretical perspectives have contributed to the protection of the right to freedom of expression in Africa; what the recent normative developments on the protection of the right to freedom of expression online in Africa are; and what the gaps between these recent updates and their implementation in African national contexts are.

In answering the first question, it noted that the importance of the right to freedom of expression is understood in African indigenous societies just as it is understood in Western societies that have claimed dominant influence on the right as protected under international human rights law today.[1] It then answered the second question by examining recent normative developments on the right to freedom of expression online under the international human rights standards. It noted that these recent normative developments which include information disorder, gender-based violence, online hate speech and the regulation of online content are generally not reflected in African national contexts.[2] In answering the third question, it noted that one of the reasons for such lack of implementation is that colonial laws lay the foundation for most cyber laws in Africa and these laws are the basis upon which most social media companies

[1] Section 2.3.3 above.
[2] Section 2.4 above.

© University of Pretoria

govern their platforms.[3] These problematic colonial laws that violate the right to freedom of expression continue to reach into provisions of current online content-related laws, thereby creating the difficulty of protecting the right to freedom of expression under international human rights law today. This is described as digital colonialism, the use or continued impact of colonial laws that shape the future of the right to freedom of expression in Africa. It also noted that digital colonialism exacerbates online harms. Online harms are engendered by various violations that occur both from state and non-state actors as a result of problematic cyberpolicies. It established that in order to implement these new developments under international human rights law in preventing online harms, problematic colonial laws and cyber laws cannot form the basis of such prevention.

Therefore, the common-law maxim *ex turpi causa non oritur actio* therefore applies but with a slight modification. The phrase which means you cannot place something on nothing, and expect it to stand, fits better into this context in that you cannot place something (platform governance) on a faulty thing (problematic colonial and cyber laws) and expect it to stand (effectiveness). The foundation is faulty, whatever is built on it will be faulty as well. In answering this sub-question, while there are theoretical and normative developments on the protection of the right to freedom of expression online under international human rights standards, most African national contexts have not benefited from these developments which exacerbates online harms.

6.2.2 The impacts of online harms on the right to freedom of expression online in Africa

In order to understand these online harms and their impacts on the right to freedom of expression in Africa, the third chapter focused on the second sub-question on the impacts of online harms on the right to freedom of expression online in Africa. In illustrating this question, it considered three additional questions on what online harms are; what their forms, methods and classifications are; and how these harms impact the right to freedom of expression in Africa.

In answering the first question, this chapter defined online harms as technology-mediated harms that are capable of violence.[4] In answering the second question, it identified its major forms as information disorder and targeted online violence in African countries. It further identified the finer examples of information disorder to include misinformation, disinformation and malinformation.[5] For targeted online

[3] Section 2.5 above.
[4] Section 3.2 above.
[5] Section 3.3.1 A-C above.

© University of Pretoria

violence, it identified cyberstalking, cyberharassment, cyberagression, online gender-based violence, online violence against children and online hate speech.[6]

It then considered the methods of online harms and how they are engendered to include primary methods as emotive narrative and constructs; fabricated multimedia; and artificial online entities while it identified secondary methods in terms of actors and dissemination.[7] It also analysed the differences, similarities and features of these disorders especially as they manifest online. It used the analysis to classify online harms based on harmfulness and legality under international human rights law. It found that while information disorder may be harmful, they are legal – that the right to freedom of expression is not limited to only true statements. It found that all forms of targeted online violence are both harmful – based on the violence they inflict on others and illegal – as a basis for limitation of the right to freedom of expression.[8]

The chapter answered the third question by identifying the impacts of these harms on online expression in African countries to include colonial legacies, distortion of the right to political participation, narrowing the space for political and unpopular speech, exclusion of women, conflation of forms of online harms that lead to problematic legal and regulatory frameworks, less protection for children's rights online and the precipitation of offline violence through online hate speech.[9] In combating these challenges, it identified a rights-respecting approach to platform governance.

6.2.3 A rights-respecting approach to platform governance in preventing online harms and protecting freedom of expression online

Having identified the impacts of online harms, the fourth chapter examined a rights-respecting approach to combating them. It examined the third sub-question of this thesis on the way a rights-respecting platform governance may be used to prevent online harms and protect freedom of expression online. In doing this, it considered three sub-questions on how various stakeholders, especially state and non-state actors understand platform governance; what a rights-respecting approach to preventing online harms is; and the way these stakeholders perform these responsibilities in order to prevent online harms and protect online expression through platform governance.

In answering the first question, the chapter identified platform governance as social media platform governance. It noted that one of the similarities between Internet and platform governance is the multistakeholder approach but that the application of this

[6] Section 3.3.2 A-D above.
[7] Section 3.4 above.
[8] Section 3.5 above.
[9] Section 3.6 above.

© University of Pretoria

approach to the latter must be more nuanced.[10] It identified the governance of social media platforms and their approaches which include content moderation as carried by platforms themselves and as a body of rules sought to be enforced outside social media platforms. It identified the various forms of platform governance, their strengths and weaknesses.[11]

These forms include traditional governance (as carried out by governments through laws); sectoral governance (as carried out by the industry players); self-governance (as carried out by platforms themselves) and multistakeholder governance (as carried out by more stakeholders with the aim of democratising governance). It identified the multistakeholder form as the most feasible for platform governance due to the novelty and complexity involved in governing a right as dynamic as online expression. It then drew up a typology of platform governance by form, actor type, mode of governance, enforcement processes and penalties that often apply.

It further identified the various limitations of platform governance to include surveillance capitalism and economic might as carried out by social media platforms; the lack of context by social media platforms in drawing up and applying their rules; clashes between international human rights laws, national laws and various actors; and patently problematic legal and regulatory frameworks.

In answering the second question, it then proposed that a human rights framework would address these limitations. It noted that such framework must be based on international human rights law through substance, process and procedural dimensions.[12] It however noted, answering the third question that such framework must be based on generative platform governance model.[13] It noted that this model would be such that is achieved incrementally with dynamic matrix of stakeholders. It identified two major proposals on content governance that could form the basis for discussing such matrix: the report by the Business and Social Responsibility (BSR) on content governance and the ongoing global consultations by ARTICLE 19 with respect to Social Media Councils (SMCs).[14] The chapter concluded that in order to be able to achieve a rights-respecting approach to platform governance, it must be generative such that all internal and external state and non-state actors must re-imagine their roles to ensure a dynamic model of governance, ensure interactive communication between other actors and learn from their failures.[15]

[10] Section 4.2.1 above.
[11] Section 4.2.4 & Section 4.3 above.
[12] Section 4.4 above.
[13] Section 4.6 above.
[14] Section 4.6.1 & Section 4.6.2 above.
[15] Section 4.7 above.

© University of Pretoria

6.2.4 The roles of state and non-state actors in platform governance, internal and external, in ensuring an effective rights-based approach to platform governance in Nigeria and South Africa

The fifth chapter examined the fourth sub-question of this thesis on the roles of state and non-state actors in platform governance, internal and external, in ensuring an effective rights-based approach to platform governance in Nigeria and South Africa. In doing this, it considered three more questions on what the legal and regulatory provisions of online harms in Nigeria and South Africa are; the ways these legal and regulatory provisions can be brought in line with international human rights standards and the roles of state and non-state actors, internally and externally in ensuring an effective rights-respecting approach to platform governance in order to prevent online harms and protect the right to freedom of expression online.

In answering the first and second question, for Nigeria, it examined several laws like the Criminal Code Act and the Penal Code, Cybercrime Act, Children's Act, Violence Against Persons Act, Hate Speech Bill, PIFM bill and the NCC Act. It noted that the legal and regulatory provisions in Nigeria that seek to regulate online harms have colonial imprints that violate international human rights standards.[16] These laws include the applicable provisions of the Criminal Code Act, the Penal Code, Cybercrime Act, NCC Act, PIFM bill and the NCHPS. Where there are no imprints, it identified legislative gaps. This includes filling gaps to regulate harms like online GBV and online violence against children. It noted that law reform should be carried out to repeal the Criminal Code Act and the Penal Code, amend the Cybercrime Act, NCC Act, PIFM bill and the Hate Speech, and enact laws to fill gaps on online GBV and violence against children online.

For South Africa, while there is no strong evidence of colonial influence, it considered the Criminal Procedure Act, Domestic Violence Act, PEPUDA, Children's Act, Electronic Communication and Transactions Act, Prevention and Combating of Hate Crimes and Hate Speech Bill. It identified the need for legal reform just as it did for Nigeria.[17]This legal reform would entail the repeal of sections 104 and 242 of the Criminal Procedure Act, and amendments of the applicable provisions of the Domestic Violence Act, Children's Act and the Films and Publications Amendment Act in line with international human rights law in order to fill policy and legislative gaps that would prevent online harms and protect online expression in South Africa.

In doing this and answering the third question, it identified that various actors in Nigeria and South Africa need to rethink their powers and interests in order to use a rights-based platform governance effectively.[18] This involves developing a Social Media

[16] Section 5.2.1 above.
[17] Section 5.2.2 above.
[18] Section 5.3 above.

© University of Pretoria

Charter as a soft law that would lead to a comprehensive, primary and substantive law on platform governance – an Online Harms Act. The four categories of actors are internal state actors, internal non-state actors, external state actors and external non-state actors.[19] It identified internal state actors as those actors that are directly funded or established by law of the state like ministries/departments, National Human Rights Institution (NHRIs) and so on while internal non-state actors are those without any ties to the state, which include civil society, private sector, academia and other proximate actors. For the external state actors, it identified the UN and AU's human rights systems and external non-state actors as social media platforms, international NGOs and philanthropy organisations.

In applying a generative model of governance, the substantive, process and procedural aspects to platform governance must be designed by these actors.[20] For internal actors, NHRC and SAHRC are to take the lead in Nigeria and South Africa respectively due to their unique roles in the promotion and protection of human rights in both countries.[21] This will be done collaboratively with other stakeholders, both internal and external. For external state actors, the treaty-monitoring bodies will take the lead by deploying their norm-setting mandates through the Special Rapporteurs on the Right to Freedom of Expression while also working with NHRIs, and other actors.[22] This dynamic interactions between various actors, both internal and external, is geared towards applying international human rights law to prevent online harms and promote online expression that is generative.[23]

6.3 Further areas for research

Given the novelty of platform governance and the new questions it raises for legal and regulatory solutions on online expression, there is a need to identify further areas for research beyond this thesis. This is because this thesis is not a silver bullet for solving the challenges of platform governance, it is primarily geared towards nudging various actors to re-think their approaches towards a creatively designed, normatively sound and a generative approach to platform governance. Therefore, there are still further aspects of research that could benefit from academic inquiry.

6.3.1 Online harms and other human rights

In understanding the wider impacts of online harms beyond the right to freedom of expression online, it will be useful for policy-setting to understand how it affects other human rights. This is because due to the inter-relatedness and interdependence of

[19] As above.
[20] Section 5.3.1 above.
[21] Section 5.3.2 above.
[22] Section 5.3.4; Section 5.3.3; Section 5.3.5 above.
[23] Section 5.4 above.

© University of Pretoria

human rights, their violations also have various negative ripple effects on each other. Given the specific focus of this thesis, it would be useful to understand the impacts of these online harms on the rights to freedom of association and assembly online alongside the impacts they may have on the prospects of a vibrant civil society in the digital age, online protests and democratic development. It will also be useful, to consider the various impacts of online harms on socio-economic rights and the role of the law using both quantitative and qualitative methods.

Noting that the area of research on online harms is still new but fast growing globally, it requires a more urgent attention in most African countries in that there are various systemic challenges it faces in the region. This can be done using comparative regional and national case studies. Therefore, its various impacts on other human rights will be useful in designing more analytical and objective systems that could assist in preventing online harms.

6.3.2 Contextual policy design for platform governance in Africa

An ambitious question that this research aims to kick-start is the re-consideration of the possibilities of protecting online speech in all African countries. This can be done, first, by critically engaging the roles of national and international state and non-state actors in social media platform governance and online speech and later, by engaging the larger Internet ecosystem on the same to ensure rights-respecting regulation approach across board. This research focuses on narrower case studies in Nigeria and South Africa and one major takeaway is the need for legal reforms. The major question that then follows after taking care of this need is how these reforms can be mainstreamed into effective policy paradigms. This research did not meaningfully engage such paradigms, rather, it focuses more on the need to get the legal reform right first, before thinking through policies that will be built on them. Therefore, there is a need for more contextual policy design which must flow from foundational and effective legal reform.

Such policy design may consider major questions of content moderation such as the role of contextual nuance that applies to an online harm such as hate speech online. It may seek to design policies that are not only sound in effective governance and administrative techniques, but that will also ensure creative application of human rights law to contemporary challenges such as new technologies and protection of free speech. For example, in what ways can socio-historical, socio-political or socio-legal perspectives influence policy mainstreaming for the protection of online speech in Africa? What is the role of the law in policy-making for the protection of online speech in Africa? For example, it may be important to understand what policy design Rwanda must consider with respect to hate speech and how that may be different from South Africa's after legal reforms. It will also be important to ensure that such designs are in compliance with constitutional and international law requirements on protecting free speech while combating its prohibited form.

© University of Pretoria

6.3.3 National Human Rights Institutions and human rights online

Human rights that can be protected offline must be protected online. Therefore, the scope of functions and capacities of NHRIs must begin to accommodate the protection of human rights as they apply online. Known as 'digital rights', human rights online should be mainstreamed into the promotional and protective mandates of NHRIs. Additional research which will consider the strong independence and effective oversight responsibilities of the NHRIs should be considered. This is because given the dynamism of digital technologies, the roles of NHRIs are bound to change in ensuring that online harms are prevented and human rights are protected. There could also be further research in determining the approach the NHRIs must take in promoting and protecting human rights in the digital age. For example, what would be the role of the NHRI in ensuring a rights-respecting legal framework for artificial intelligence or any other frontier technologies?

6.3.4 The revised Declaration and the role of the African Commission in ensuring effective online content governance in Africa

Some parts of the revised Declaration of Principles on Freedom of Expression and Access to Information in Africa have been used to illustrate the roles of states in protecting online speech in the digital age in Africa. This is best seen in the role of the Declaration as the most proximate 'soft law' on the role of states in protecting online speech. However, beyond this illustration, there exists the need to flesh out finer aspects of the Declaration that could require more descriptive details for states to follow on protecting online speech in Africa. This is necessary because most states' criticisms against international human rights law is that not only are there no specific descriptive directions on how to apply these rights-based frameworks with respect to new technologies, even where they exist, they lack context. For example, a set of guidelines that states can adopt with respect to online content governance, which could flow from the provisions of Principles 37 - 39 of the Declaration could help with such specific instructions and fill contextual gaps that states claim to contend with.

Therefore, there is need for more engagements with the revised Declaration to continuously and gradually move international human rights law from being prescriptive, abstract and less culturally relevant towards being more descriptive, applicable and contextually useful, especially with respect to the protection of online speech. These examples have existed in the past as the African Commission has drafted several model laws, declarations, guidelines and even recommendations to guide states with respect to a human rights policy challenge. The research that may look into this need could focus on the history of implementation of the African Commission's 'soft laws' especially with respect to digital technologies and human rights. It could also consider how such implementation could be improved and mainstreamed into national contexts.

© University of Pretoria

6.3.5 A feminist legal theory approach to combating online harms

This research applied post-colonial legal theory to rethink the impacts of online harms as caused by digital colonialism. Postcolonial legal theory is one out of the many lenses of critical legal theories. Another lens which could be useful in combating online harms in African countries could be the feminist legal theory. In a rather robust meaning of what such could mean in explaining online harms, it would mean investigating these harms through the prisms of race, gender, capitalism, sexuality and imperialism. As a prominent feature of online harms, not only does online gender-based violence feature prominently, vulnerable groups like persons living with disability, sexual minorities, refugees and others, are adversely affected. In her closing remarks on the need to centre women in conversations on digital coloniality in Africa, Tamale argued that:

> While African women's digital footprint online may be viewed as a good thing as they are less likely to be affected by digital coloniality, the downside is their exclusion from the positive aspects of ICT. Women need to critically challenge the disempowering elements of technology.[24]

It would therefore be useful to study how state and non-state actors must pay attention to the peculiar online harms faced by vulnerable persons. For state actors, such academic pursuit could interrogate the role of soft laws as opposed to hard laws in engaging the feminist dimensions of online harms in African countries. Such inquiry could also engage the relationship between colonialism, the feminist legal theory and online harms in African countries. All of these assist in grounding policies on online harms in thorough analyses that could deliver thoughtful and effective solutions.

6.4 Conclusion

This thesis examined how to prevent online harms and protect online expression through a rights-respecting approach to platform governance. It found that due to the complexity of platform governance, the roles of actors in platform governance in Nigeria and South Africa must be generative while anchored on international human rights law. Calls for anchoring online speech rules on international human rights law is not new, what is new however, and to which this thesis makes a contribution, is how to apply this law generatively in order to ensure prevention of online harms and the protection of online free speech. This generative approach is particularly necessary given the need to apply context to platform governance through multistakeholder perspectives. For example, an application of just international human rights law, without contextual factors like problematic colonial and new laws in Nigeria and South Africa will be building the future of online speech on patently problematic laws. Therefore, the generative approach allows for a thorough ground-up development of

[24] S Tamale *Decolonisation and afro-feminism* (2020) 395.

© University of Pretoria

online speech rules and this can be seen in the roles that have been recommended for various actors in Nigeria and South Africa.

State actors and non-state actors, whether external or internal, have been identified in this thesis as having an important role to play in ensuring a rights-respecting platform governance. At the centre of these responsibilities lies the need for all actors to commit to a generative approach that is dynamic, open and communicative. For internal state actors, this role is to ensure legal reform through enactment and implementation of rights-respecting laws and policies.[25] These actors also have the responsibility to relax their expectations of traditional rule-making given the complexity of Internet and platform governance and allow for logical, creative and dynamic application of international human rights law. It is recommended that the NHRIs, given their unique mandate on promotion and protection of human rights in Nigeria and South Africa should take up this role for the internal state actors. With respect to internal non-state actors, it is recommended that civil society should play a key role in skills and capacity development with respect to how international human rights law applies to digital technologies and the policy needs that arise.

For external state actors, it is recommended that the standard-setting roles of Special Rapporteurs in the UN and AU treaty-making bodies should be used to set up a process for ensuring a body of rules anchored on international human rights law for state actors on online speech governance. For non-external state actors like social media platforms, their role is to ensure necessary review of their policies in line with international human rights law and collaborate more with local stakeholders with respect to norm and standard setting. For international NGOs, they have the responsibility of working with diverse actors in the online speech ecosystem to understand the nuance and context necessary in drawing up online speech rules. Finally, it recommends that philanthropy organisations commit more strategically to support internal state and non-state and external non-state actors' mandates on ensuring rights-respecting online speech governance not only in Nigeria and South Africa, but also across African countries and other jurisdictions outside Africa.

[25] Section 5.3.1 above.

© University of Pretoria

BIBLIOGRAPHY AND REFERENCES

Books

Amoah, GY *Groundwork of government for West Africa* (Ilorin: Gbenle Press 1988)

Arewa, OB *Disrupting Africa: Technology, law, and development* (Cambridge: Cambridge University Press 2021)

Ayittey, G *Indigenous African institutions* (Ardsley: Transnational Publishers 2006)

Benkler, Y, Faris, R & Roberts, H *Network propaganda: Manipulation, disinformation, and radicalization in American politics* (New York: Oxford University Press 2018)

Berlin, I *Essays on liberty* (Oxford: Oxford University Press 1969)

Bohannan, P & Bohannan, L *Tiv economy* (Evanston: Northwestern University Press 1968)

Bottomley, S & Bronitt, S *Law in context* (Annandale: Federation Press 2006)

Busia, KA *Africa in search of democracy* (London: Routledge and Kegan Paul 1967)

Busia, KA *The position of the chief in the modern political system of Ashanti* (Oxford: Oxford University Press 1951)

Chanock, M *Law, custom and social order* (Cambridge: Cambridge University Press 1985)

Chanock, M *The making of South African legal culture* (Cambridge: Cambridge University Press 2004)

Citron, D *Hate crimes in cyberspace* (Cambridge: Harvard University Press 2014)

DeNardis, L *The global war for internet governance* (New Haven: Yale University Press 2014)

Donnelly, J *Universal human rights in theory and practice* (Ithaca: Cornell University Press 2013)

Dutton, W *The Oxford handbook of internet studies* (Oxford: Oxford University Press

© University of Pretoria

2013)

Elias, T *British colonial law: A comparative study of the interaction between English and local laws in British dependencies.* (London: Stevens 1962)

Elias, T *Nigerian land law* (London: Sweet and Maxwell 1971)

Ellul, J, Kellen, K & Lerner, J *Propaganda: The formation of men's attitudes* (New York: Knopf 1965)

Emerson, T *The system of freedom of expression* (New York: Random House 1970)

Eze, O *Human rights in Africa: Some selected problems* (Lagos: Macmillan Nigeria 1984)

Fitzpatrick, P *Modernism and the grounds of law* (Cambridge: Cambridge University Press 2001)

Fitzpatrick, P *The mythology of modern law* (London: Routledge 1992)

Fuchs, C *Social media: A critical introduction* (London: SAGE Publications Ltd 2014)

Gardiner, GG & Lansdown, LW *South African Criminal Law and Procedure* (Cape Town: Juta & Co Limited 1924)

Gillespie, T *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media* (New Haven: Yale University Press 2018)

Goldsmith, J & Wu, T *Who controls the internet: Illusions of a borderless world* (Oxford: Oxford University Press 2006)

Harris E *Africans and their history* (New York: New American Library 1987)

Hart, HLA *Law, liberty, and morality* (California: Stanford University Press 1963)

Hawkins, M *Social Darwinism in European and American thought, 1860–1945: Nature as model and nature as threat* (Cambridge: Cambridge University Press 1997)

© University of Pretoria

Howard, R *Human rights in commonwealth Africa* (Totowa: Rowman & Littlefield 1986)

Humphrey, J *No distant millennium: The international law of human rights* (Paris: UNESCO 1989)

Ibhawoh, B *Human rights in Africa* (Cambridge: Cambridge University Press 2018)

Ibhawoh, B *Imperialism and human rights: Colonial discourses of rights and liberties in African history* (Albany: State University of New York Press 2007)

Jeong, S *The internet of garbage* (Vox Media, Inc 2015)

Jowett, G & O'Donnell, V *Propaganda & persuasion* (Thousand Oaks: SAGE Publications 2005)

Kaye, D *Speech police: The global struggle to govern the internet* (New York: Columbia Global Reports 2019)

Kosseff, J *The twenty-six words that created the internet* (Ithaca: Cornell University Press 2019)

Kperogi, F *Nigeria's digital diaspora: Citizen media, democracy, and participation* (Rochester: University of Rochester Press 2020)

Lessig, L *Code: And other laws of cyberspace* (New York: Basic Books 2006)

Locke, J *Two treatises of government* (London: Awnsham Churchill 1689)

Lodge, M & Taber, C *The rationalizing voter* (Cambridge: Cambridge University Press 2013)

MacKinnon, R *Consent of the networked: The worldwide struggle for internet freedom* (New York: Basic Books 2013)

Mamdani, M *Citizen and subject: Contemporary Africa and the legacy of late colonialism* (New Jersey: Princeton University Press 1996)

Mandela, W *Part of my soul went with him* (New York: Norton 1984)

Marshall, J *Personal freedom through human rights law* (Leiden: Martinus Nijhoff

© University of Pretoria

Publishers 2009)

Mill, J *On liberty* (London: Longmans, Green, and Company 1859)

Milo, D *Defamation and freedom of speech* (Oxford: Oxford University Press 2008)

Moore, M *Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age* (London: Oneworld 2018)

Mueller, M *Networks and states: The global politics of internet governance* (Cambridge: MIT Press 2010)

Mueller, M *Ruling the root: Internet governance and the taming of cyberspace* (Cambridge: MIT Press 2004)

Murray, A *The regulation of cyberspace: Control in the online environment* (Abingdon: Routledge-Cavendish 2007)

Mutua, M (ed) *Human rights NGOs in East Africa: Political and normative tensions* (Philadelphia: University of Pennsylvania Press 2009)

Nelson, R *A Chronology and glossary of propaganda in the United States* (Connecticut: Greenwood Press 1996)

Nowak, M *UN Covenant on civil and political rights: CCPR commentary* (Kehl: Engel 2005)

Nyabola, N *Digital democracy, analogue politics: How the internet era is transforming politics in Kenya* (London: Zed 2018)

Olowu, D *An integrative rights-based approach to human development in Africa* (Pretoria: Pretoria University Law Press 2009)

Papacharissi, Z *Affective publics: Sentiment, technology, and politics* (New York: Oxford University Press 2015)

Popper, K *Conjectures and refutations: The growth of scientific knowledge* (London: Routledge 1989)

© University of Pretoria

Rawls, J *A theory of justice* (Cambridge: Belknap Press 1971)

Rid, T *Active measures: The secret history of disinformation and political warfare* (New York: Farrar, Straus and Giroux 2020)

Roberts, S *Behind the screen: Content moderation in the shadows of social media* (New Haven: Yale University Press 2019)

Spijkers, O, *The United Nations: The evolution of global values and international law* (Cambridge: Intersentia 2011)

Suzor, N *Lawless: The secret rules that govern our digital lives* (Cambridge: Cambridge University Press 2019)

Tamale, S *Decolonisation and afro-feminism* (Québec: Daraja Press 2020).

Ten, CL *Mill on liberty* (Oxford: Clarendon Press 1980)

Unger, R *Law in modern society: Toward a criticism of social theory* (New York: Free Press 1986)

van Notten, M *The law of the Somalis* (Trenton: Red Sea Press 2006)

Waldron, J *The harm in hate speech* (Cambridge: Harvard University Press 2012)

Ward, I *An introduction to critical legal theory* (London: Cavendish Publishing 1998)

Welch, C *Protecting human rights in Africa: Roles and strategies of non-governmental organizations* (Philadelphia: University of Pennsylvania Press 1995)

Williams, C *The destruction of black civilization* (Chicago: Third World Press 1987)

Woodman, G & Obilade, A *African Law and Legal Theory* (New York: New York University Press 1995)

Viljoen, F *International human rights law in Africa* (Oxford: Oxford University Press 2012)

Chapters in books

Andrew, J 'Introduction' in Andrew, J & Bernard, F (eds), *Human rights responsibilities*

© University of Pretoria

in the digital age: States, companies, and individuals (Gordonsville: Hart Publishing 2021)

Bell, E 'The unintentional press: How technology companies fail as publishers' in Bollinger, LC and Stone, GR (eds) *The free speech century* (New York: Oxford University Press 2018)

Bickert, M 'Defining the boundaries of free speech on social media' in Bollinger LC and Stone GR (eds) *The free speech century* (New York: Oxford University Press 2019)

Callamard, A 'The human rights obligations of non-state actors' in Jørgensen, RF (ed) *Human rights in the age of platforms* (Cambridge: MIT Press 2019)

Clarke, J 'Law and race: The position of indigenous people' in Bottomley S and Parker S (eds) *Law in context* (Annandale: Federation Press 1997)

Cunliffe-Jones, P, Diagne, A, Finlay, A & Schiffrin, A 'Bad law – legal and regulatory responses to misinformation in sub-Saharan Africa 2016–2020' in Cunliffe-Jones, P, Diagne, A, Finlay, A, Gaye, W, Gichunge, C, Onumah, C, Pretorius, A & Schiffrin, A (eds) *Misinformation policy in sub-Saharan Africa* (London: University of Westminster Press 2021)

Darian-Smith, E and Fitzpatrick, P 'Laws of the postcolonial: An insistent introduction' in Darian-Smith, E & Fitzpatrick, P (eds) *Laws of the postcolonial* (Ann Arbor: University of Michigan Press 1999)

Davies, M 'Race and colonialism: Legal theory as White mythology' in Davies, M (ed) *Asking the law question: The dissolution of legal theory* (Sydney: Law Book Company 2002)

Dwyer, M & Molony, T 'Mapping the study of politics and social media use in Africa' in Dwyer, M & Molony, T (eds) *Social media and politics in Africa: Democracy, censorship and security* (London: Zed Books 2019)

Edwards, L 'Pornography, censorship and the internet' in Edwards, L and Waelde, C (eds) *Law and the internet* (Oxford: Hart Publishing 2009)

© University of Pretoria

Gillespie, T 'Regulation of and by platforms' in Burgess, J, Poell, T & Marwick, A (eds) *SAGE handbook of social media* (London: SAGE Publications Ltd 2017)

Grimm, D 'Freedom of speech in a globalized world' in I Hare and J Weinstein (eds) *Extreme speech and democracy* (Oxford: Oxford University Press 2009)

Harman, E 'Harming as causing harm' in Roberts, MA and Wasserman, DT (eds) *Harming future persons* (Dordrecht: Springer 2009)

Harmer, E & Lumsden, K 'Conclusion: Researching "online othering"—future agendas and lines of inquiry' in Harmer, E & Lumsden, K (eds) *Online othering: Exploring digital violence and discrimination on the web* (London: Palgrave Macmillan 2019)

Jasanoff, S 'The idiom of co-production' in Jasanoff, S (ed) *States of knowledge: The co-production of science and social order* (London: Routledge 2004)

Jørgensen, R 'Human rights and private actors in the online domain' in Land, MK & Aronson, JD (eds) *New technologies for human rights law and practice* (Cambridge: Cambridge University Press 2018)

Jørgensen, R, Veiberg, C and & Oever, N 'Exploring the Role of HRIA in the Information and Communication Technologies (ICT) Sector' in Götzmann, N (ed) *Handbook on human rights impact assessment* (Cheltenham: Edward Elgar Publishing 2018)

Kabumba, B 'Soft law and legitimacy in the African union system: The case of the pretoria principles on ending mass atrocities pursuant to article 4(h) of the Constitutive Act of the African Union' in Shyllon, O (ed) *The model law on access of information for Africa and other regional instruments: Soft law and human rights in Africa* (Pretoria: Pretoria University Law Press 2018)

Karp, I 'African systems of thought' in Karp, I and Bird, CS (eds) *Explorations in African systems of thought* (Bloomington: Indiana University Press 1986)

Killander, M & Adjolohoun, H 'International law and domestic human rights litigation in Africa: An introduction' in M Killander *International law and domestic human rights litigation in Africa* (Pretoria: Pretoria University Law Press 2010)

© University of Pretoria

Kuwali, D 'Decoding Afrocentrism: Decolonising legal theory' in Onazi, O (ed) *African legal theory and contemporary problems: Critical essays* (Dordrecht: Springer 2014)

Land, M 'Regulating private harms online: Content regulation under human rights law' in Jørgensen, RF (ed), *Human rights in the age of platforms* (Cambridge: MIT Press 2019)

Land, M & Hamilton, R 'Beyond takedown: Expanding the toolkit for responding to online hate' in Dojčinović, P (ed) *Propaganda and international criminal law: From cognition to criminality* (Abingdon: Routledge 2020)

Lauren, PG 'The foundations of justice and human rights in early legal texts and thought' in Shelton, D (ed) *The Oxford handbook of international human rights law* (Oxford: Oxford University Press 2013)

McGonagle, T 'Freedom of expression and information in the UN' in McGonagle, T & Donders, Y (eds) *The United Nations and freedom of expression and information: Critical perspectives* (Cambridge: Cambridge University Press 2015)

McGonagle, T 'The development of freedom of expression and information within the UN: leaps and bounds or fits and starts' in T McGonagle & Y Donders (eds) *The United Nations and freedom of expression and information* (2015) 21

Mendel, T 'The UN Special Rapporteur on Freedom of Opinion and Expression: Progressive development of international standards relating to freedom of expression' in McGonagle, T and Donders, Y (eds) *The United Nations and freedom of expression and information: Critical perspectives* (Cambridge: Cambridge University Press 2015)

Parekh, B 'Is there a case for banning hate speech' in Herz, M & P Molnár, P (eds) *The content and context of hate speech: Rethinking regulation and responses* (Cambridge: Cambridge University Press 2012)

Persily, N & Tucker, JA 'Introduction' in N Persily & JA Tucker *Social media and democracy: The state of the field and prospects for reform* (Cambridge: Cambridge University Press 2020)

© University of Pretoria

Price, M & Verhulst, S 'The concept of self-regulation and the internet' in Waltermann, J & Machill, M (eds), *Protecting our children on the internet: Towards a new culture of responsibility* (Gütersloh: Bertelsmann Foundation Publishers 2000)

Ranger, T 'The invention of tradition in colonial Africa: Anthropological contribution' in Hobsbawm, EJ & Ranger, T (eds) *The invention of tradition* (Cambridge: Cambridge University Press 1983)

Siebert, F 'The libertarian theory of the press' in Siebert, F, Peterson, T & Schramm, W (eds) *Four theories of the press: The authoritarian, libertarian, social responsibility, and Soviet communist concepts of what the press should be and do* (Urbana: University of Illinois Press 1963)

Silungwe, C 'On African legal theory: A possibility, an impossibility or mere conundrum?' in Onazi, O (ed) *African legal theory and contemporary problems: Critical essays* (Dordrecht: Springer 2014)

Smith, P, Steffgen, G & Sittichai, R 'The nature of cyberbullying, and an international network' in Smith, PK & Steffgen, G (eds), *Cyberbullying through the new media: Findings from an international network* (London: Psychology Press 2013)

Snyder, F & Summer, C 'Colonialism and legal form: The creation of "customary law" in Senegal' *Crime, justice and underdevelopment* (London: Routledge 1981)

Spivak, G 'Post-structuralism, marginality, postcoloniality and value' in Collier, P & Geyer-Ryan, H (eds) *Literary theory today* (Cambridge: Polity Press 1990)

Strauss, DA 'Freedom of expression and the common-law constitution' in Stone, GR & Bollinger, LC (eds) *Eternally vigilant: Free speech in the modern era* (Chicago: University of Chicago Press 2002)

Sunstein, C 'The future of free speech' in Stone, GR and Bollinger, LC (eds) *Eternally vigilant: Free speech in the modern era* (Chicago: University of Chicago Press 2002)

Swire, B and Ecker, U 'Misinformation and its correction: Cognitive mechanisms and recommendations for mass communication' in Southwell, B, Thorson, EA & Sheble, L (eds) *Misinformation and mass audiences* (Austin: University of Texas Press 2018)

© University of Pretoria

Timan, T, Galič, M & Koops, B 'Surveillance theory and its implications for law' in Brownsword, R, Scotford, E & Yeung, K (eds) *The Oxford handbook of law, regulation and technology* (Oxford: Oxford University Press 2017)

Tusikov, N 'Transnational non-state regulatory regimes' in Drahos, P (ed) *Regulatory theory: Foundations and applications* (Acton: Australian National University Press 2017)

Ward, SJA 'Classical liberal theory in a digital world' in Fortner, RS & Fackler, PM (eds) *The handbook of media and mass communication theory* (Chichester: John Wiley & Sons 2014)

Wilde, R 'The extraterritorial application of international human rights law on civil and political rights' in Sheeran, S and Rodley, N (eds) *Routledge handbook of international human rights law* (Abingdon: Routledge 2013)

Willetts, P 'Transnational actors and international organizations in global politics' in Baylis, JB & Smith, S (eds), *The globalisation of world politics* (Oxford: Oxford University Press 2001)

Wittenberg, C and Berinsky, J 'Misinformation and its correction' in Persily, N & Tucker, JA (eds) *Social media and democracy: The state of the field, prospects for reform* (Cambridge: Cambridge University Press 2020)

Zuboff, S '"We make them dance": Surveillance capitalism, the rise of instrumentarian power, and the threat to human rights' in Jørgensen, RF (ed) *Human rights in the age of platforms* (The Cambridge: MIT Press 2019)

Journal articles

Adibe, R, Ike, CC & Udeogu, CU 'Press freedom and Nigeria's Cybercrime Act of 2015: An assessment (2017) 52 *Africa Spectrum* 117

Agrafiotis, I, Nurse, JRC, Goldsmith, M, Creese, S & Upton, U 'A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate' (2018) *Journal of Cybersecurity* 1

© University of Pretoria

Akpan, F 'Bridging the gap between non-state actors and the state in governance: Evidence from Nigeria' (2011) 6 *International Journal of Development and Management Review* 62

Arun, C 'Facebook's faces' 135 *Harvard Law Review Forum* 2

Asogwa, N & Ezeibe, C 'The state, hate speech regulation and sustainable democracy in Africa: A study of Nigeria and Kenya' (2020) *African Identities* 1

Aswad, EM 'Losing the freedom to be human' (2020) 52 *Columbia Human Rights Law Review* 2

Aswad, EM 'The future of freedom of expression online' (2018) 17 *Duke Law & Technology Review* 26

Balkin, J 'The future of free expression in a digital age' 36 *Pepperdine Law Review* 2009 434

Bauman, S & Bellmore, A 'New Directions in Cyberbullying Research' (2015) *Journal of School Violence* 2

Benda-Beckman, F 'Law out of context: A comment on the creation of tradition law discussion' (1984) 28 *Journal of African Law* 28

Benvenisti, E & Harel, A 'Embracing the tension between national and international human rights law: The case for discordant parity' (2017) 15 *International Journal of Constitutional Law* 36

Berghel, H 'Malice domestic: The Cambridge Analytica dystopia' (2018) *IEEE Computer Society* 85

Berinsky, AJ 'Rumours and health care reform: Experiments in political misinformation' (2017) 47 *British Journal of Political Science* 241

Bloch-Wehba, H 'Global platform governance: Private power in the shadow of the state' 72 *Southern Methodist University Law Review* 55

Bogolyubova, O, Panicheva, P, Tikhonov, R, Ivanov, V & Ledovaya, Y 'Dark personalities on Facebook: Harmful online behaviors and language' (2018) 78 *Computers in Human Behavior* 151

© University of Pretoria

Bontcheva, K & Posetti, K 'Balancing act: Countering digital disinformation while respecting freedom of expression' (2020) *International Telecommunications Union* 8 22

Bosch, T 'Twitter activism and youth in South Africa: The case of #RhodesMustFall' (2016) *Information, Communication & Society* 10

Botha, J 'Towards a South African free-speech model' (2017) 134 *South African Law Journal* 815-818

Botha, J '"Swartman": Racial descriptor or racial slur?' *Rustenburg Platinum Mine v SAEWA obo Bester* [2018] ZACC 13; 2018 (5) SA 78 (CC)' 2020 10 *Constitutional Court Review* 353-377

Botha, J & Govindjee, A 'Regulating "extreme cases of hate speech" in South Africa: A suggested framework for a legislated criminal sanction' (2014) 27 *South African Journal of Criminal Justice* 154

Broberg, M & Sano, H 'Strengths and weaknesses in a human rights-based approach to international development – an analysis of a rights-based approach to development assistance based on practical experiences' (2018) 22 *The International Journal of Human Rights* 664

Budree, A, Fietkiewicz, K & Lins, E 'Investigating usage of social media platforms in South Africa' 11 *The African Journal of Information Systems* 315

Bushman, BJ & Anderson, CA 'Is it time to pull the plug on hostile versus instrumental aggression dichotomy?' (2001) 108 *Psychological Review* 274

Callamard, A 'Accountability, transparency and freedom of expression in Africa' (2010) 77 *Social Research* 1211

Carmi, GE '"Dignity," the enemy from within: A theoretical and comparative analysis of human dignity as a free speech justification' (2006-2007) 9 *University of Pennsylvania Journal of Constitutional Law* 957

Cassim, F 'Regulating hate speech and freedom of expression on the Internet: Promoting tolerance and diversity' 28 (2015) *South African Journal of Criminal Justice* 303-336

Chari, T 'Future prospects of the print newspaper in Zimbabwe' (2011) 3 *Journal of African Media Studies* 367

© University of Pretoria

Cheeseman, N, Fisher, J, Hassan I & Hitchen, J 'Social media disruption: Nigeria's WhatsApp politics' 31 *Journal of Democracy* 156

Chesney, B & Citron, DK 'Deep fakes: A looming challenge for privacy, democracy, and national security' (2019) 10 *California Law Review* 1744

Chirwa, DM 'The doctrine of state responsibility as a potential means of holding private actors accountable for human rights' (2004) 5 *Melbourne Journal of International Law* 250

Citron, DK & Franks, MA 'Criminalizing revenge porn' (2014) 49 *Wake Forest Law Review* 355

Clooney, A & Webb, P 'The right to insult under international law' (2017) 48 *Columbia Human Rights Review* 25

Cohn, C 'Bad facts make bad law: How platform censorship has failed so far and how to ensure that the response to Neo-Nazis doesn't make it worse' (2018) 2 *George Law Technology Review* 442

Coleman, D 'Digital colonialism: The 21st Century scramble for Africa through the extraction and control of user data and the limitations of data protection laws' (2019) 24 *Michigan Journal of Race & Law* 439

Corcoran, L, McGuckin, C & Prentice, G 'Cyberbullying or cyber aggression?: A review of existing definitions of cyber-based peer-to-peer aggression (2015) 5 *Societies* 247

Couldry, N & Mejias, UA 'Data colonialism: Rethinking big data's relation to the contemporary subject' (2019) 20(4) *Television & New Media* 339

Dada, JA 'Human rights protection in Nigeria: The past, the present and goals for role actors for the future' (2013) 14 *Journal of Law, Policy and Globalisation* 1

Dahl, RA 'What political institutions does large-scale democracies require?' (2005) 120 *Political Science Quarterly* 187

De Baets, A 'The impact of the Universal Declaration of Human Rights on the study of history' (2009) 48 *History and Theory* 20

Deem, A Gagliardone, I, Csuka, L & Udupa, S 'Hate speech, information disorder and conflict' (2020) *Social Science Research Council* 3

© University of Pretoria

De Gregorio, G & Stremlau, N 'Internet shutdowns and the limits of the law' (2020) *International Journal of Communications* 4224

Demobour, MB 'What are human rights: Four schools of thoughts' (2010) 32 *Human Rights Quarterly* 1

DeNardis, L 'Internet governance by social media platforms' (2015) 39 *Telecommunications Policy* 761

Diamond, L 'The democratic rollback: The resurgence of the predatory state' (2008) 87 *Foreign Affairs* 37

Donnelly, J 'The relative universality of human rights' (2007) 29 *Human Rights Quarterly* 281

Douek, E 'The limits of international law in content moderation' (2021) 6 *UC Irvine Journal of International, Transnational, and Comparative Law* 72

Douek, E 'Governing online speech: From 'posts-as-Trumps' to proportionality and probability' 121 *Columbia Law Review* 803

Dripps, DA 'The liberal critique of the harm principle' (1998) 17 *Criminal Justice Ethics* 3

Ecker, UKH, Lewandowsky, S, Swire, B & Chang, D 'Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction' (2011) 18 *Psychonomic Bulletin & Review* 570

Ecker, UKH, Lewandowsky, S, Cheung, CSC & Maybery, MT 'He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation' (2015) 85 *Journal of Memory and Language* 101

Ekdale, B & Tully, M 'African elections as a testing ground: Comparing Coverage of Cambridge Analytica in Nigerian and Kenyan Newspapers (2019) 40 *African Journalism Studies* 13

Elkins, Z 'Getting to rights: Treaty ratification, constitutional convergence and human rights practice' (2013) 54 *Harvard International Law Journal* 201

Essien, VLK 'The Northern Nigeria Penal Code: A reflection of diverse values in penal legislation' (1985) 5 *New York Law School Journal of International and Comparative Law* 89

© University of Pretoria

Ezeibe, CC & Ikeanyibe, OM 'Ethnic politics, hate speech, and access to political power in Nigeria' (2017) 63 *Africa Today* 65

Farhangpour, P, Maluleke, C & Mutshaeni, KN 'Emotional and academic effects of cyberbullying on students in a rural high school in the Limpopo province, South Africa' (2019) 21 *South African Journal of Information Management* 8

Fetzer, JH 'Disinformation: The use of false information' (2004) 14 *Minds and Machines* 231

Fitzpatrick, P 'Traditionalism and tradition law' (1984) 28 *Journal of African Law* 20

Flew, T, Martin, F & Suzor, N 'Internet regulation as media policy: Rethinking the question of digital communication platform governance' (2019) 10 *Journal of Digital Media and Policy* 33

Franck, TM 'Is personal freedom a Western value' (1997) 91 *American Journal of International Law* 595

Galtung, J 'Kulturelle Gewalt' (1993) 43 *Der Burger im Staat* 106

Gill, K 'Regulating platforms' invisible hand: Content moderation policies and processes' (2020) 21 *Wake Forest Journal of Business and Intellectual Property Law* 173-212

Gillespie, T 'Introduction: Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates' (2020) 9 *Internet Policy Review: Journal of Internet Regulation* 2020 2

Goldman, E 'Content moderation remedies' (2021) *Michigan Technology Law Review* 1

Gorwa, R 'The platform governance triangle: Conceptualising the informal regulation of online content' (2019) 8 *Internet Policy Review* 1

Gorwa, R 'What is Platform Governance?' (2019) 22 *Information, Communication & Society* 854

Grigg, DW 'Cyber-aggression: Definition and concept of cyberbullying' (2010) 20(2) *Australian Journal of Guidance and Counselling* 143

© University of Pretoria

Grimmelmann, J 'The virtues of moderation' (2015) 17 *Yale Journal of Law & Technology* 63

Guichard, A 'Hate crime in the cyberspace: the challenges of substantive criminal law' 18 *Information and Communications Technology Law* 211

Haggart, B and Keller, CI 'Democratic legitimacy in global platform governance' (2021) 45 *Telecommunications Policy* 3

Hamilton, RJ 'Governing the global public square' (2021) 62 Harvard International Law Journal 117-174

Hannum, H 'Reinvigorating human rights for the twenty-first century' (2016) 16 *Human Rights Law Review* 439

Hartmann, IA 'A new framework for online content moderation' (2020) 36 *Computer Law & Security Review* 105376

Heath, FD 'Tribal society and democracy' (2001) 5 *The Laissez Faire City Times* 22

Helberger, N, Pierson, J and Poell, T 'Governing online platforms: From contested to cooperative responsibility' (2018) 34 *The Information Society* 3

Helmond, A 'The Platformization of the web: Making web data platform ready' (2015) 1 *Social Media + Society*

Herring, SC 'Cyber violence: Recognizing and resisting abuse in online environments' (2002) 14 *Asian Women* 187

Hillis, S, Mercy, J & Amobi, A 'Global prevalence of past-year violence against children: A systematic review and minimum estimates' (2016) 137 *Pediatrics* 6

Hinduja, S & Patchin, JW 'Offline consequences of online victimization' (2007) 6 *Journal of School Violence* 107

Ho, K 'Structural violence as a human rights violation' 4 *Essex Human Rights Review* 1

Hogan, B & Quan-Haase, A 'Persistence and change in social media' (2010) 30 *Bulletin of Science, Technology & Society* 309

Humphrey, J 'The international bill of rights: Scope and implementation' (1976) 17 *William & Mary Law Review* 527

© University of Pretoria

Jensen, SLB, Lagoutte, S & Lorion, S 'The domestic institutionalisation of human rights: An introduction' (2019) 37 *Nordic Journal of Human Rights* 165

Johnson, HM & Seifert, CM 'Sources of the continued influence effect: When misinformation in memory affects later inferences (1994) 20 *Journal of Experimental Psychology: Learning, Memory and Cognition* 1420

Kanna, M 'Furthering decolonization: Judicial review of colonial criminal laws' (2020) 70 *Duke Law Journal* 411, 424

Kaplan, AM & Haenlein, M 'Users of the world, unite! The challenges and opportunities of social media' (2009) 53 *Business Horizons* 59

Kleinwächter, W 'Internet co-governance: Towards a multilayer multiplayer mechanism of consultation, coordination and cooperation (M3C3)' (2006) 3 *E-Learning and Digital Media* 473

Klonick, K 'The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression' (2020) 129 *Yale Law Journal* 2437

Klonick, K 'The new governors: The people, rules and processes governing online speech' (2018) 131 *Harvard Law Review* 1603

Kragh, M & Åsberg, S 'Russia's strategy for influence through public diplomacy and active measures: The Swedish case' (2017) 40 *Journal of Strategic Studies* 25

Krayanak, RP 'John Locke from absolutism to toleration' (1980) 74 *The American Political Science Review* 53

Kshetri, N 'Cybercrime and cybersecurity in Africa' (2019) 22 *Journal of Global Information* 77

Lagoutte, S 'The role of state actors within the national human rights system' (2019) 37 *Nordic Journal of Human Rights* 177

Land, MK 'Against privatized censorship: Proposals for responsible delegation' (2019) 60 *Virginia Journal of International Law* 363

Land, MK 'Towards an international law of the Internet' 54 *Harvard International Law Journal* 393

© University of Pretoria

Langos, C 'Cyberbullying: The challenge to define' (2012) 15 *Cyberpsychology, Behavior, and Social Networking* 286

Lechler, M & MacNamee, L 'Indirect colonial rule undermines support for democracy: Evidence from a natural experiment from Namibia' (2018) 51 *Comparative Political Studies* 1858

Lewandowsky, S, Ecker, UKH, Seifert, CM, Schwarz, N & Cook, J 'Misinformation and its correction: Continued influence and successful debiasing' (2012) 13 *Psychological Science in the Public Interest* 106

MacAllister, JM 'The doxing dilemma: Seeking a remedy for the malicious publication of personal information' (2017) *85 Fordham Law Review* 2455

Maréchal, N 'Networked authoritarianism and the geopolitics of information: Understanding russian internet policy' (2017) 5 *Media and Communication* 29

Marchant, E & Stremlau, N 'The changing landscape of internet shutdowns in Africa' 14 (2020) *International Journal of Communications* 4216 4220

Maripe, B 'Freezing the press: Freedom of expression and statutory limitations in Botswana' (2003) 3 *African Human Rights Law Journal* 66

Martin, DA, Shapiro, JN & Nedashkovskaya, M 'Recent trends in online foreign influence efforts' (2019) 18 *Journal of Information Warfare* 15

Massey, CR 'Hate speech, cultural diversity, and the foundational paradigms of free expression' (1992) 40 *UCLA Law Review* 103

Matsuda, MJ 'Public response to racist speech: Considering the victim's story' (1989) 87 *Michigan Law Review* 2320

McGlynn, C, Rackley, E & Houghton, R 'Beyond 'Revenge Porn': The continuum of image-based sexual abuse (2017) 25 *Feminist Legal Studies* 26

McKay, S, Tenove, C 'Disinformation as a threat to deliberative democracy' (2020) 74 *Political Research Quarterly* 703

Mohney, S 'The great power origins of human rights' (2014) 35 *Michigan Journal of International Law* 828

© University of Pretoria

Moran, M 'Talking about hate speech: A rhetorical analysis of American and Canadian approaches to the regulation of hate speech' (1994) *Wisconsin Law Review* 1428

Morris, HF 'A history of the adoption of codes of criminal law and procedure in British Colonial Africa, 1876-1935' (1974) 18 *Journal of African Law* 23

Morris, HF 'How Nigeria got its Criminal Code' (1970) 14 *Journal of African Law* 137-154

Mubangizi, JC 'A human rights-based approach to development in Africa: Opportunities and challenges (2014) 39 *Journal of Social Sciences* 67

Mueller, M Mathiason, J and Klein, H 'The Internet and global governance: Principles and norms for a new regime' (2007) 13 *Global Governance: A Review of Multilateralism and International Organizations* 237

Murray, R 'International human rights: Neglect of perspectives from African institutions' (2006) 55 *The International and Comparative Law Quarterly* 193

Naidoo, K 'The origins of hate-crime laws' 22 *Fundamina* 53-66; J Botha & A Govindjee (n 106 above)117-155

Naidoo, K 'Factors which influenced the enactment of hate-crime legislation in the United States of America: *Quo Vadis* South Africa' (2016) *Journal of South African Law* 705

Nasiritousi, N, Hjerpe, M & Bäckstrand, K 'Normative arguments for non-state actor participation in international policymaking processes: Functionalism, neocorporatism or democratic pluralism?' (2016) 22 *European Journal of International Relations* 920

Naughton, J 'The evolution of the Internet: From military experiment to General Purpose Technology' (2016) 1 *Journal of Cyber Policy* 5

Nkrumah, B 'Words that wound: Rethinking online hate speech in South Africa' (2018) 23 *Alternation Journal* 118-123

Nkomo, NZ, Kandiro, A & Bigirimana, S 'The viability of the print newspaper in the digital era in Zimbabwe: A digital strategy perspective' (2017) 5 *European Journal of Business and Innovation Research* 44

© University of Pretoria

Nwogu, GAI 'Democracy: Its meaning and dissenting opinions of the political class in Nigeria: A philosophical approach' (2015) 6 *Journal of Education and Practice* 131

Nyhan, B & Reifler, J 'When corrections fail: The persistence of political misperceptions' (2010) 32 *Political Behaviour* 303

Nyokabi, DN, Diallo, N, Ntesang, NW, White, TK & Ilori, T 'The right to development and internet shutdowns: Assessing the role of information and communications technology in democratic development in Africa' (2019) 3 *Global Campus Human Rights Journal* 147

Ochi, IB & Mark, KC 'Effect of the #EndSARS protests on the Nigerian economy' (2021) 9 *Global Journal of Arts, Humanities and Social Sciences* 2

Ofori-Parku, SS & Moscat, D 'Hashtag activism as a form of political action: A qualitative analysis of the #BringBackOurGirls Campaign in Nigerian, UK, and US Press' (2018) 23 *International Journal of Communication* 2480

Ogbondah, CW & Onyedike, EU 'Origins and interpretations of Nigerian press laws' (1991) 5 *Africa Media Review* 59

Ogunleye, J 'The concepts of predictive analytics' (2014) 2 International Journal of Knowledge, Innovation and Entrepreneurship 83

Okebukola, EO 'The application of international law in Nigeria and the façade of dualism' (2020) 11 *Nnamdi Azikiwe University Journal of International Law* 21

Okolie, EQ 'Extent of the latitudes and limits of social media, and freedom of expression within the confines of the law in Nigeria' (2019) 83 *Journal of Law, Policy and Globalization* 162-167

Okoth-Ogendo, HWO 'Some issues of theory in the tenure relations in African agriculture' (1989) 59 *Africa: Journal of the International African Institute* 6

Oliva, TD 'Content moderation technologies: Applying human rights standards to protect freedom of expression' (2020) *Human Rights Law Review* 607-640

Papaevangelou, C 'The existential stakes of platform governance: A critical literature review' (2021) *Open Research Europe* 1

Penney, W 'Internet access rights: A brief history and intellectual origins' (2011) 38 *William Mitchell Law Review* 23

© University of Pretoria

Pielemeier, J 'Disentangling disinformation: What makes regulating disinformation so difficult' (2020) 4 *Utah Law Review* 921

Pohjonen, M 'A comparative approach to social media extreme speech: Online hate speech as media commentary' (2019) 13 *International Journal of Communication* 3091

Prakash, G 'Postcolonial criticism and Indian historiography' (1992) 31/32 *Social Text* 8

Price, L 'Platform responsibility for online harms: towards a duty of care for online hazards' (2022) 13 *Journal of Media Law* 238-261

Pyżalski, J 'From cyberbullying to electronic aggression: Typology of the phenomenon' (2012) 17 *Emotional and Behavioural Difficulties* 305

Redish, M 'The value of free speech' (1982) 130 *University of Pennsylvania Law Review* 605

Redish, MH 'Self-realisation, democracy and freedom of expression: A response to Professor Baker' (1981) 130 *University of Pennsylvania Law Review* 678

Rice, MF 'Information and communication technologies and the global digital divide' (2003) 1 *Comparative Technology Transfer and Society* 72

Roberts, L 'Jurisdictional and definitional concerns with computer-mediated interpersonal crimes: An Analysis on Cyber Stalking' (2008) 2 *International Journal of Cyber Criminology* 272

Rochefort, A 'Regulating social media platforms: A comparative policy analysis' (2020) 25 *International and Comparative Perspectives on Communication Law* 225

Ronel, Y 'Human rights obligations of territorial non-state actors' (2013) 46 *Cornell International Law Journal* 21

Roy, A 'Postcolonial theory and law: A critical introduction' (2008) 29 *Adelaide Law Review* 278

Kakungulu-Mayambala, R & Rukundo, S 'Digital activism and free expression in Uganda' (2019) 19 *African Human Rights Law Journal* 167-192

© University of Pretoria

Rutenberg, I, Zalo, M, Sugow, A 'Appraising the impact of Kenya's cyber-harassment law on the freedom of expression' (2021) 1 *Journal of Intellectual Property and Information Law* 91 114

Salau, AO 'Social media and the prohibition of 'false news': Can the free speech jurisprudence of the African Commission on Human and Peoples' Rights provide a litmus test?' (2020) 4 *African Human Rights Yearbook* 231

Sanders, B 'Freedom of expression in the age of online platforms: The promise and pitfalls of a human rights-based approach to content moderation' (2020) 43 *Fordham International Law Journal* 939

Santarelli, N 'Non-state actors' human rights obligations and responsibility under international law' (2008) 15 *Revista Electronica De Estudios Internacionales* 1

Scanlon, T 'A theory of freedom of expression' (1972) 1 *Philosophy and Public Affairs* 204

Sewpaul, V 'The West and the rest divide: Human rights, culture and social work' (2016) 1 *Journal of Human Rights and Social Work* 31

Shariff, S & Hoff, D 'Cyber bullying: Clarifying legal boundaries for school supervision in cyberspace' (2007) 1 *International Journal of Cyber Criminology* 84

Snyder, F 'Customary law and the economy' (1984) 28 *Journal of African Law* 34
Sunstein, CR 'Fifty shades of manipulation' (2016) 1 *Journal of Marketing Behaviour* 213

Sive, D & Price, A 'Regulating expressions on social media' (2019) 136 *South African Law Journal* 51-83

Suzor, N 'A constitutional moment: How we might reimagine platform governance' (2020) 36 *Computer Law & Security Review* 2

Tar, UA 'The challenges of democracy and democratisation in Africa and Middle East' (2010) 3 *Information, Society and Justice* 88

Tadeg, MA 'Freedom of expression and the media landscape in Ethiopia: Contemporary challenges' (2020) 5 *University of Baltimore Journal of Media Law and Ethics* 69-99

Ten, CL 'Was Mill a liberal?' (2002) 1 *Politics, Philosophy & Economics* 355-370

© University of Pretoria

Tokunaga, RS 'Following you home from school: A critical review and synthesis of research on cyberbullying victimization' (2010) 26 *Computers in Human Behaviour* 279

Tucker, J, Guess, A, Barberá, P, Vaccari, C, Siegel, A, Sanovich, S, Stukal, D, & Nyhan, B 'Social media, political polarisation, political disinformation: A review of scientific literature' (2018) *Social Science Research Network* 3

Udofa, IJ 'Right to freedom of expression and the law of defamation in Nigeria' (2013) 2 *International Journal of Advanced Legal Studies and Governance* 75

Viljoen, F 'Africa's contribution to the development of international human rights and humanitarian law' (2001) 1 *African Human Rights Law Journal* 19

Viljoen, F 'Contemporary challenges to international human rights law and the role of human rights education' 44 (2011) *De Jure* 209-220

Vindex 'The suggested repeal of Roman-Dutch law in South Africa' (1901) 18 *South African Law Journal* 153

Vollenhoven, WJ 'The right to freedom of expression: The mother of our democracy' (2015)18 *Potchefstroom Electronic Law Journal* 2302

Wachter, S & Mittelstadt, B 'A right to reasonable inferences: Re-thinking data protection law in the age of Big Data and AI' (2019) 2 *Columbia Business Law Review* 494

Wasserman, H 'Fake news from Africa: Panics, politics and paradigms' (2020) 21 *Journalism* 3

Weeks, BE & Garrett, RK 'Electoral consequences of political rumours: Motivated reasoning, candidate rumours and vote choice during the 2008 US presidential election' (2014) 26 *International Journal of Public Opinion Research* 402

Welch, CE 'The African Charter and the freedom of expression in Africa' (1998) 4 *Buffalo Human Rights Law Review* 112

West, SM 'Thinking beyond content in the debate about moderation' (2020) 9 *Internet Policy Review: Journal of Internet Regulation* 15

Wilson, RA & Land, MK 'Hate speech on social Media: Towards a context-specific content moderation policy' (2020) 52 *Connecticut Law Review* 47

© University of Pretoria

Wodajo, K 'Mapping (in)visibility and structural injustice in the digital space' (2022) 9 *Journal of Responsible Technology* 1-8
Woolley, SC & Howard, PN 'Political communication, computational propaganda, and autonomous agents' (2016) 10 *International Journal of Communication* 4882

Zankova, B & Dimitrov, V 'Social media regulation: Models and proposals' (2020) 10 *Journalism and Mass Communication* 75

Zarsky, T 'Social justice, social norms and the governance of social media' (2015) 35 *Pace Law Review* 154

Ziegler, C 'International dimensions of electoral processes: Russia, the USA, and the 2016 elections' (2018) 55 *International Politics* 557

Reports

'Draft National Action Plan: 2021 - 2025' (2021) https://www.nigeriarights.gov.ng/files/nap/NAP%20for%20final%20Review%20July%209%20(3)-converted.pdf (accessed 15 August 2021)

Africa Centre for Strategic Studies 'Domestic disinformation on the rise in Africa' (2021) https://africacenter.org/spotlight/domestic-disinformation-on-the-rise-in-africa/ (accessed 13 October 2021)

AfriForum 'Comments on the Draft Regulations of the Internet Censorship Amendment Act' (2020) https://afriforum.co.za/wp-content/uploads/2020/08/AfriForum-commentary-Internet-Censorship-Amendment-Bill.pdf (accessed 18 August 2021)

AfriForum 'Submission on the Prevention and Combating of Hate Crimes and Hate Speech Bill' (2018) https://www.afriforum.co.za/wp-content/uploads/2019/09/Afriforum-Submission-B9-2018-MO-003.pdf (accessed 18 August 2021)

Alt Advisory & Research ICT Africa 'Domestic Violence Amendment Bill: Joint submission by Research ICT Africa and Alt Advisory' (2021) https://altadvisory.africa/wp-content/uploads/2021/07/ALT-Advisory-Research-ICT-

© University of Pretoria

Africa-Joint-Submissions-Domestic-Violence-Amendment-Bill.pdf (accessed 15 August 2021)

ARTICLE 19 & Electronic Frontier Foundation (EFF) 'Necessary and proportionate: International principles on the application of human rights law communication surveillance – Background and supporting international legal analysis' (2014) https://www.article19.org/data/files/medialibrary/37564/N&P-analysis-2-final.pdf (accessed 24 August 2021)

ARTICLE 19 'Social Media Councils: Consultation paper' (2019) https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf (accessed 16 June 2019)

ARTICLE 19 'Social Media Councils: Consultations' (2019) https://www.article19.org/resources/social-media-councils-consultation/ (accessed 15 June 2020)

ARTICLE 19 'Uganda: Government must safeguard freedom of expression after arrest and attack' (2017) https://www.article19.org/resources/uganda-government-must-safeguard-freedom-of-expression-after-arrest-and-attack/ (accessed 15 January 2018)

Badenhorst, C 'Legal responses to cyberbullying and sexting in South Africa' (2001) *Centre for Justice and Crime Prevention*

Bossey, C 'Report of the United Nations Working Group on Internet Governance' (2005) https://www.wgig.org/docs/WGIGREPORT.pdf (accessed 14 February 2021)

Bosch, T & Roberts, T 'South Africa digital rights landscape report' in T Roberts (ed) *Digital rights in closing civic space: lessons from ten African countries* (2021) 143 https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/15964/South_Africa_Report.pdf (accessed 1 December 2021)

Bradshaw, S & Howard, PN 'Challenging truth and trust: A Global Inventory of Organized Social Media Manipulation' (2018) *Oxford Internet Institute*, *University of Oxford*

© University of Pretoria

Brown, H 'Violence against vulnerable groups' *Council of Europe* https://www.corteidh.or.cr/tablas/r25587.pdf (accessed 15 January 2020)

Business for Social Responsibility (BSR) 'A human rights-based approach to content governance' (2021) https://www.bsr.org/reports/A_Human_Rights-Based_Approach_to_Content_Governance.pdf (accessed 2 April 2021)

Chala, E 'How the murder of musician Hachalu Hundessa incited violence in Ethiopia: Part II' (2020) *Global Voices* https://globalvoices.org/2020/08/07/how-the-murder-of-musician-hachalu-hundessa-incited-violence-in-ethiopia-part-ii/ (accessed 23 September 2020)

Chapelle, B 'Multistakeholder governance: Principles and challenges of an innovative political paradigm' (2011) in Kleinwachter, W (ed) *Multistakeholder Internet Dialogue Collaboratory Discussion* Paper Series No 1 http://dl.collaboratory.de/mind/mind_02_neu.pdf (accessed 15 February 2021)

Coliver, S 'Commentary on the Johannesburg principles on national security, freedom of expression and access to information' (1999) https://www.right2info.org/exceptions-to-access/resources/publications/CommentaryontheJohannesburgPrinciples.pdf10-11 (accessed 24 July 2021)

Department of Communications 'Electronic Communications and Transactions Act (25/2002): Guidelines for recognition of industry representative bodies of information system service providers' (2006) https://www.ellipsis.co.za/wp-content/uploads/2011/02/IRB-Regulations-Gazette-29474.pdf (accessed 13 August 2021)

Digital, Culture, Media and Sport Committee 'Disinformation and 'fake news': Interim Report: Government's Response to the Committee's Fifth Report of Session 2017-2019' https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1630/1630.pdf (accessed 20 September 2020)

Donahoe, E & Hampson, FE 'Governance innovation for a connected world protecting free expression, diversity and civic engagement in the global digital ecosystem' (2018)

© University of Pretoria

Centre for International Governance Innovation: Special Report https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/stanford_special_report_web.pdf (accessed 15 August 2020)

Donovan, J 'Navigating the tech stack: When, where and how should we moderate content' (2019) in *Models for Platform Governance CIGI Essay Series* https://www.cigionline.org/sites/default/files/documents/Platform-gov-WEB_VERSION.pdf (accessed 20 February 2021)

Duncan, J 'Monitoring and defending freedom of expression and privacy on the internet in South Africa' (2012) *Global Information Society Watch* 14 https://giswatch.org/sites/default/files/southafrica_gisw11_up_web.pdf (accessed 1 December 2021)

Geneva Academy 'Defending the boundary: Constraints and requirements on the use of autonomous weapon systems under international humanitarian and human rights law' **(**2016**)** https://www.geneva-academy.ch/joomlatools-files/docman-files/Briefing9_interactif.pdf (accessed 20 August 2021)

Global Network Initiative (GNI) 'Content regulation and human rights' (2020) https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf (accessed 23 February 2021)

Graham, B & MacLellan, M 'Overview of the challenges posed by Internet platforms: Who should address them and how' (2018) in Donahoe, E & Hampson, FO (eds) *Governance innovation for a connected world protecting free expression, diversity and civic engagement in the global digital ecosystem*

Guerriero, M 'The impact of Internet connectivity on economic development in sub-Saharan Africa' *Economic and Private Sector (*2015) https://assets.publishing.service.gov.uk/media/57a0899b40f0b652dd0002f4/The-impact-of-internet-connectivity-on-economic-development-in-Sub-Saharan-Africa.pdf (accessed 21 February 2021)

Helen Suzman Foundation 'Submission in response to the Prevention and Combating of Hate Crimes and Hate Speech Bill (Gazette No 41543 of 29 March 2018)' (2019)

© University of Pretoria

https://hsf.org.za/publications/submissions/hsf-submission-hate-crimes-and-hate-speech-bill.pdf (accessed 14 October 2021)

Hoffmann, S, Taylor, E & Bradshaw, S 'The market of disinformation' (2019) *Oxford Internet Institute, University of Oxford* https://oxtec.oii.ox.ac.uk/wp-content/uploads/sites/115/2019/10/OxTEC-The-Market-of-Disinformation.pdf (accessed 28 September 2020)

Institute for Multi-Stakeholder Initiative 'Not Fit-for-Purpose: The grand experiment of multi-stakeholder initiatives in corporate accountability, human rights and global governance (2020)https://www.msi-integrity.org/wp-content/uploads/2020/07/MSI_Not_Fit_For_Purpose_FORWEBSITE.FINAL_.pdf (accessed 15 July 2020)

International Telecommunications Union (ITU) 'Measuring digital development: Facts and figures 2020' https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf (accessed 18 February 2021)

Internet & Jurisdiction Policy Network 'Toolkit Cross-border Content Moderation' (2021) https://www.internetjurisdiction.net/uploads/pdfs/Internet-Jurisdiction-Policy-Network-21-104-Toolkit-Cross-border-Content-Moderation-2021.pdf (accessed 15 August 2021)

ISPA 'ISPA submissions on the draft Films and Publications Regulations, 2020' (2020) https://ispa.org.za/wp-content/uploads/2020/08/ISPA-Submission-Draft-Films-and-Publications-Regulations-20200817.pdf (accessed 14 August 2021)

ISPA 'ISPA submissions on the Hate Crimes and Hate Speech bill' (31 January 2019) https://ispa.org.za/wp-content/uploads/2019/02/ISPA-Hate-Crimes-and-Hate-Speech-Bill-31-January-2019.pdf (accessed 14 August 2021)

ISPA 'Submissions on the draft online content regulation policy' (2015) https://ispa.org.za/wp-content/uploads/2012/06/Internet-Service-Providers-Association-ISPA.pdf (accessed 14 August 2021)

© University of Pretoria

Iyer, N, Nyamwire, B & Nabulega, S 'Alternate realities, alternate internets: African feminist research for a feminist Internet' (2020) *Pollicy* https://www.apc.org/sites/default/files/Report_FINAL.pdf (accessed 15 October 2020)

Jack, C 'Lexicon of lies: Term for problematic information' (2017) *Data & Society*

Jackson, D 'Distinguishing disinformation from propaganda, misinformation and "fake news"' (2018) *National Endowment for Democracy*

Livingstone, S & Bulger, M 'A global agenda for children's rights in the digital age recommendations for developing UNICEF's research strategy' (2013) *UNICEF* https://www.unicef-irc.org/publications/pdf/lse%20olol%20final3.pdf (accessed 27 October 2020)

Luminate 'Data and digital rights in Nigeria: Assessing the activities, issues and opportunities' (2021) https://luminategroup.com/storage/1361/Data-%26-Digital-Rights-in-Nigeria-Report-%5BFINAL%5D.pdf (accessed 15 August 2021)

Madebo, A 'Social media, the diaspora, and the politics of ethnicity in Ethiopia' (2020) *Democracy in Africa* http://democracyinafrica.org/social-media-the-diaspora-and-the-politics-of-ethnicity-in-ethiopia/ (accessed 23 October 2020)

Madung, O & Obilo, B 'Inside the shadowy world of disinformation for hire in Kenya' (2021) *Mozilla* https://foundation.mozilla.org/en/blog/fellow-research-inside-the-shadowy-world-of-disinformation-for-hire-in-kenya/ (accessed 15 October 2021)

Malhotra, N 'End violence: Women's rights and safety online' https://www.apc.org/sites/default/files/end_violence_malhotra_dig.pdf (accessed 15 October 2021)

Manilla principles on Internet intermediaries 'Background paper' https://www.eff.org/files/2015/07/08/manila_principles_background_paper.pdf (accessed 15 March 2021)

Marari, D 'Of Tanzania's cybercrimes law and the threat to freedom of expression and information' (2015) *AfricLaw* https://africlaw.com/2015/05/25/of-tanzanias-

© University of Pretoria

cybercrimes-law-and-the-threat-to-freedom-of-expression-and-information/ (accessed 20 January 2020)

Media Defence 'Mapping digital rights and online freedom of expression litigation in East, West and Southern Africa' 1 October 2021 https://www.mediadefence.org/resource-hub/wp-content/uploads/sites/3/2021/08/Media-Defence-Mapping-digital-rights.pdf (accessed 30 October 2021)

Morar, D & Riley, C 'A guide for conceptualising the debate over Section 230' (2021) *Brookings* https://www.brookings.edu/techstream/a-guide-for-conceptualizing-the-debate-over-section-230/ (accessed 15 April 2021)

Nigerian Communications Commission (NCC) 'Final report: Study on young children and digital technology' September 2021 https://www.ncc.gov.ng/accessible/documents/1005-young-children-and-digital-technology-a-survey-across-nigeria/file (accessed 1 December 2021)

NSRP 'How-to guide: Mitigating dangerous speech: Monitoring and countering dangerous speech to reduce violence' (2017) http://www.nsrp-nigeria.org/wp-content/uploads/2017/12/NSRP-How-to-Guide-Mitigating-Hate-and-Dangerous-Speech.pdf accessed (15 September 2020)

Nwaodike, C & Naidoo, N 'Fighting violence against women online: A comparative analysis of legal frameworks in Ethiopia, Kenya, Senegal, South Africa, and Uganda' (2020) *Pollicy*

Office of the Special Representative of the Secretary-General on Violence Against Children 'Ending the torment: Tackling bullying from the schoolyard to the cyberspace' (2016) https://violenceagainstchildren.un.org/sites/violenceagainstchildren.un.org/files/documents/publications/tackling_bullying_from_schoolyard_to_cyberspace_low_res_fa.pdf (accessed 25 July 2020)

Paradigm Initiative & OONI 'Tightening the noose on freedom of expression: 2018 Status of Internet Freedom in Nigeria' (2019) https://ooni.org/documents/nigeria-report.pdf (accessed 15 August 2021)

© University of Pretoria

Pauwels, E 'The anatomy of information disorders in Africa' (2020) *Konrad-Adenauer-Stiftung* https://www.kas.de/documents/273004/10032527/Report+-+The+Anatomy+of+Information+Disorders+in+Africa.pdf/787cfd74-db72-670e-29c0-415cd4c13936?version=1.0&t=1599674493990 (accessed 15 October 2021)

Perset, K, West, J, Winickoff, D & Wyckoff, A 'Moving "upstream" on global platform governance' (2019) in *Models for platform governance: A CIGI essay series* https://www.cigionline.org/sites/default/files/documents/Platform-gov-WEB_VERSION.pdf

Phyfer, J; Burton, P & Leoschut, L 'South African Kids Online: Barriers, opportunities and risks. A glimpse into South African children's internet use and online activities' (2016) *Centre for Justice and Crime Prevention* http://globalkidsonline.net/wp-content/uploads/2016/06/GKO_Country-Report_South-Africa_CJCP_upload.pdf (accessed 1 December 2021)

Plan International 'Free to be online? Girls and young women's experiences of online harassment' (2020) https://plan-international.org/publications/freetobeonline (accessed 1 December 2021)

Reventlow, JN, Penney, J, Johnson, A, Junco, R, Tilton, C, Coyer, K, Dad, N, Chaudhri, A, Mutung'u, G, Benesch, S, Lombana-Bermudez, A, Noman, H, Albert, K, Sterzing, A, Oberholzer-Gee, F, Melas, H, Zuleta, L, Kargar, S, Matias, JN, Bourassa, N & Gasser, U 'Perspectives on harmful speech online' (2016) *Berkman Klein Center for Internet & Society Research* Publication https://dash.harvard.edu/bitstream/handle/1/33746096/2017-08_harmfulspeech.pdf?sequence=5&isAllowed=y (accessed 23 July 2019)

Roberts, T (ed) 'Digital rights in closing civic space: Lessons from ten African countries' (2021) *Institute of Development Studies* https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/15964/Digital_Rights_in_Closing_Civic_Space_Lessons_from_Ten_African_Countries.pdf?sequence=4&isAllowed=y 114, 122, 153, 164 (accessed 23 May 2021)

© University of Pretoria

Rule of Law 'Submission to the Portfolio Committee on Justice and Correctional Services on the Prevention and Combating of Hate Crimes and Hate Speech bill, 2018' (2021) https://www.freemarketfoundation.com/dynamicdata/documents/20210927-submission-on-hate-speech-bill.pdf (accessed12 October 2021)

SAHRC 'Findings of the South African Human Rights Commission regarding certain statements made by Mr Julius Malema and another member of the Economic Freedom Fighters' (2019) https://www.sahrc.org.za/home/21/files/SAHRC%20Finding%20Julius%20Malema%20&%20Other%20March%202019.pdf (accessed 15 June 2021)

SAHRC 'Revised strategic plan for the fiscal years 2015 to 2020' https://www.sahrc.org.za/home/21/files/SAHRC%20Revised%20Strategic%20Plan%202015%20-%202020.pdf 18 (accessed 15 August 2021)

SAHRC 'Terms of Reference: Develop a draft Social Media Charter for the South African Human Rights Commission https://www.sahrc.org.za/home/21/files/Terms%20of%20reference%20-%20Social%20Media%20Charter%20-%20Final.doc (accessed 15 August 2021)

Salanova, R 'Social media and political change: The case of the 2011 revolutions in Tunisia and Egypt' (2012) *ICIP* Working *Papers 2012/7*

Strickling, LE & Hill, JF 'Multi-stakeholder governance innovations to protect free expression, diversity and civility online' (2018) in Donahoe, E & Hampson, FO (eds) *Governance innovation for a connected world: Protecting free expression, diversity and civic engagement in the global digital ecosystem* https://apo.org.au/sites/default/files/resource-files/2018-11/apo-nid203391.pdf (accessed 20 August 2021)

Taye, B & Pallero, J 'Ethiopia's hate speech predicament: Seeking antidotes beyond a legislative response' (2020) *Access Now* https://www.accessnow.org/open-letter-to-facebook-protect-ethiopians/ (accessed 15 August 2020)

The Women's Legal and Human Rights Bureau 'End violence: Women's rights and safety online from impunity to justice: Domestic legal remedies for cases of

© University of Pretoria

technology-related violence against women' (2015) https://www.genderit.org/sites/default/files/flow_domestic_legal_remedies_0.pdf (accessed 15 October 2019)

Thorson, E 'Identifying and correcting policy misperceptions' (2016) https://www.americanpressinstitute.org/wp-content/uploads/2015/04/Project-2-Thorson-2015-Identifying-Political-Misperceptions-UPDATED-4-24.pdf (accessed 27 September 2020)

UNICEF 'Global kids online report' (2019) https://www.unicef-irc.org/publications/pdf/GKO%20LAYOUT%20MAIN%20REPORT.pdf (accessed 15 October 2020)

Wardle, C & Derakhshan, H 'Information disorder: Toward an interdisciplinary framework for research and policy making' (2017) *Council of Europe Report* https://rm.coe.int/information-disorder-report-2017/1680766412 (accessed 20 September 2020)

Wardle, C 'Information disorder: The essential glossary' (2018) *Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy* https://firstdraftnews.org/wp-content/uploads/2018/07/infoDisorder_glossary.pdf (accessed 20 April 2020)

Wardle, C 'Understanding information disorder' (2019) *First Draft* https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76701 (accessed 15 September 2020)

Woolley, SC & Howard, PN 'Computational propaganda worldwide: Executive Summary' (2017) *Oxford Internet Institute, University of Oxford* https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf (accessed 1 July 2019)

© University of Pretoria

News and website articles

'Chad slows down internet to curb hate speech on social media' *AL JAZEERA* 4 August 2020 https://www.aljazeera.com/news/2020/8/4/chad-slows-down-internet-to-curb-hate-speech-on-social-media (accessed 15 October 2021)

'Facebook shuts Uganda accounts ahead of vote' *Yahoo!* 11 January 2020 https://news.yahoo.com/facebook-shuts-uganda-accounts-ahead-110022184.html?guccounter=1 (accessed 12 January 2020)

'Kemi Olunloyo & Samuel Welson bail application hearing postponed over judge's absence' *Naijagists* 24 March 2017 https://naijagists.com/kemi-olunloyo-samuel-welson-bail-application-hearing-postponed-judges-absence/ (accessed 14 October 2021)

'OAU Charter' https://au.int/sites/default/files/treaties/7759-file-oau_charter_1963.pdf (accessed 20 March 2020)

'Summary Sheet' https://bit.ly/3qECucH (accessed 16 September 2020)

Access Now 'Open letter to Facebook on violence-inciting speech: Act now to protect Ethiopians' *Access Now* 27 July 2020 https://www.accessnow.org/open-letter-to-facebook-protect-ethiopians/ (accessed 15 April 2021)

Achieng, G 'How harassment keeps women politicians offline in Uganda' *restofworld* 1 September 2021 https://restofworld.org/2021/women-politics-social-media-uganda/ (accessed 19 October 2021)

ARTICLE 19 'Our mission' https://www.article19.org/about-us/ (accessed 15 June 2020)

African Commission on Human and Peoples' Rights 'Declaration of Principles on Freedom of Expression and Access to Information in Africa 2019' https://www.achpr.org/presspublic/publication?id=80 (accessed 1 December 2021)

African Commission on Human and Peoples' Rights 'Documentation Centre' https://www.achpr.org/documentationcenter (accessed 1 December 2021)

© University of Pretoria

Athumani, H 'Ugandan government restores social media sites, except Facebook' *Voice of America* 10 February 2021 https://www.voanews.com/a/africa_ugandan-government-restores-social-media-sites-except-facebook/6201864.html (accessed 16 April 2021)

Au, Y, Howard, PN & Ainita, P 'Profiting from the pandemic moderating COVID-19 lockdown protest, scam, and health disinformation websites' (2020) https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/127/2020/12/Profiting-from-the-Pandemic-v8-1.pdf (accessed 5 December 2020)

Ayalew, YE 'Uprooting hate speech: The challenging task of content moderation in Ethiopia' *Centre for International Media Assistance* (CIMA) 27 April 2021 https://www.cima.ned.org/blog/uprooting-hate-speech-the-challenging-task-of-content-moderation-in-ethiopia/ (accessed 15 October 2021)

Azelmat, M 'Can social media platforms tackle online violence without structural change?' *Association for Progressive Communications* 17 August 2021 https://genderit.org/node/5511 (accessed 23 August 2021)

Belli, L & Zingales, N 'Glossary of platform law and policy terms' https://cyberbrics.info/wp-content/uploads/2020/11/Glossary-on-Platform-Law-and-Policy-CONSOLIDATED17472-1.pdf 106 (accessed 10 November 2020)

Belsey, B 'Cyberbullying: An emerging threat to the "always on" generation' (2005) https://billbelsey.com/?p=1827 (accessed 12 February 2021)

Benesch, S 'Proposals for improved regulation of harmful online content' https://dangerousspeech.org/wp-content/uploads/2020/06/Proposals-for-Improved-Regulation-of-Harmful-Online-Content-Formatted-v5.2.1.pdf (accessed 24 August 2021)

Bietti, E 'A genealogy of digital platform regulation' 3 June 2021 https://bit.ly/3j5YX0W (accessed 15 October 2021)

Campbell, J 'Nigerian President Buhari clashes with Twitter Chief Executive Dorsey' *Council on Foreign Relations* 8 July 2021 https://www.cfr.org/blog/nigerian-president-buhari-clashes-twitter-chief-executive-dorsey (accessed 16 July 2021)

© University of Pretoria

Cassim, F 'Addressing the growing spectre of cybercrime in Africa: Evaluating measures adopted by South Africa and other regional role players' (2011) https://core.ac.uk/download/pdf/79170924.pdf (accessed 23 July 2021)

Cavoukian, A 'Privacy by design: The 7 foundational principles' (2011) https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf (accessed 5 February 2021)

Centre for Human Rights 'Centre for Human Rights participates in the Southern African sub-regional consultation on the revision of the draft Declaration of Principles on Freedom of Expression and Access to Information in Africa' 3 October 2020 https://www.chr.up.ac.za/expression-information-and-digital-rights-news/1855-centre-for-human-rights-participates-in-the-southern-african-sub-regional-consultation-on-the-revision-of-the-draft-declaration-of-principles-on-freedom-of-expression-and-access-to-information-in-africa (accessed 15 February 2021)

Centre for Human Rights 'Democracy, Transparency and Digital Rights Unit participates in Francophone consultation on freedom of expression and access to information' 22 October 2019 https://www.chr.up.ac.za/expression-information-and-digital-rights-news/1872-democracy-transparency-and-digital-rights-unit-participates-in-francophone-consultation-on-freedom-of-expression-and-access-to-information (accessed 15 February 2021)

Cheeseman, N & Fisher, J 'How colonial rule committed Africa to colonial rule' *Quartz Africa* 2 November 2019 https://qz.com/africa/1741033/how-colonial-rule-committed-africa-to-fragile-authoritarianism-2/ (accessed 17 June 2020)

Chen, Y, Conroy, NJ & Rubin, VL 'Misleading online content: Recognizing clickbait as "false news"' (2015) http://dx.doi.org/10.1145/2823465.2823467 (accessed 30 July 2020)

Committee to Protect Journalists 'Nigerian journalists charged with criminal defamation, breach of peace' 29 October 2019 https://cpj.org/2019/10/nigerian-journalists-charged-with-criminal-defamat/ (accessed 15 July 2021)

© University of Pretoria

Committee to Protect Journalists 'Two journalists charged with sedition over presidential jet story' 27 June 2006 https://cpj.org/2006/06/two-journalists-charged-with-sedition-over-preside/ (accessed 12 October 2021)

De Streel, A 'Online platforms' moderation of illegal content: Law, practices' (2020) https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU652718_EN.pdf (2020) (accessed 15 October 2020)

Debré, E 'The Facebook Oversight Board has made its first rulings' *Slate* 28 January 2021 https://slate.com/technology/2021/01/facebook-oversight-boards-content-moderation-rulings.html (accessed 3 June 2021)

Digital Rights Forensic Lab 'Nigerian government-aligned Twitter network targets #EndSARS protests' *Medium* 20 November 2020 https://medium.com/dfrlab/nigerian-government-aligned-twitter-network-targets-endsars-protests-5bb01a96665c (accessed 21 November 2020)

Duncan, J 'Monitoring and defending freedom of expression and privacy on the internet in South Africa' (2011) https://www.apc.org/sites/default/files/SouthAfrica_GISW11_UP_web.pdf (accessed 18 August 2021)

Dvoskin, B 'International human rights law is not enough to fix content moderation's legitimacy crisis' *Berkman Klein Center Collection* 16 September 2020 https://medium.com/berkman-klein-center/international-human-rights-law-is-not-enough-to-fix-content-moderations-legitimacy-crisis-a80e3ed9abbd (accessed 20 February 2021)

Dwyer, M & Molony, T 'How social media is changing politics in Africa' 23 February 2021 *Democracy in Africa* https://democracyinafrica.org/socialmedia/ (accessed 16 November 2021)

Edelman, G 'On Social Media, American-Style Free Speech Is Dead' *Wired* 27 April 2021 https://www.wired.com/story/on-social-media-american-style-free-speech-is-dead/ (accessed 29 April 2021)

© University of Pretoria

Ekdale, B & Tully, M 'Cambridge Analytica in Africa – what do we know?' *Democracy in Africa* 10 January 2020 http://democracyinafrica.org/cambridge-analytica-africa-know/ (accessed 16 August 2020)

Ekwealor, V 'Nigeria's president refused to sign its digital rights bill, what happens now?' *Techpoint* 27 March 2019 https://techpoint.africa/2019/03/27/nigerian-president-declines-digital-rights-bill-assent/ (accessed 20 August 2021)

Essa, A 'China is buying African media's silence' *Foreign Policy* 14 September 2018 https://foreignpolicy.com/2018/09/14/china-is-buying-african-medias-silence/ (accessed 23 June 2020)

Essien, H 'Installing Twitter seditious under Penal Code of Northern Nigeria, AGF Malami tells Court' *The People Gazette* 21 September 2021 https://gazettengr.com/installing-twitter-seditious-under-penal-code-of-northern-nigeria-agf-malami-tells-court/ (accessed 14 October 2021)

Facebook 'Global feedback and input on the Facebook Oversight Board for content decisions' 27 June 2019 https://about.fb.com/news/2019/06/global-feedback-on-oversight-board/ (accessed 13 October 2021)

Facebook 'Government request to remove content' https://transparencyreport.google.com/government-removals/overview?hl=en (accessed 13 February 2020)

Facebook 'Oversight Board Charter' (2019) 5 https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf (accessed 1 December 2021)

Facebook 'What is a legal restriction on access to content on Facebook' https://www.facebook.com/help/1601435423440616?helpref=related (accessed 15 June 2019)

Facebook Community Standards 'Hate Speech' https://www.facebook.com/communitystandards/hate_speech (accessed 19 November 2019)

© University of Pretoria

Oxford Pro Bono Publico 'Comparative hate speech law: Memorandum' (2012) https://www.law.ox.ac.uk/sites/files/oxlaw/1._comparative_hate_speech_-_lrc.pdf (accessed 15 June 2021)

Gacengo, B 'Online child sexual exploitation' *Council of Europe* 15 August 2020 https://rm.coe.int/3148-afc2018-ws9-ecpat-manifestations/16808e85b8 (accessed 12 February 2021)

Giles, C & Mwai, P 'Africa internet: Where and how are governments blocking it?' *BBC News* 14 January 2021 https://www.bbc.com/news/world-africa-47734843 (accessed 15 January 2020)

Global Internet Forum to Counter Terrorism https://gifct.org (accessed 15 March 2021)

Gorwa, R 'The shifting definition of platform governance' *Centre for International Governance Innovation* 23 October 2019 https://www.cigionline.org/articles/shifting-definition-platform-governance (accessed 21 March 2021)

Government of Uganda 'Presidency warns Facebook and Twitter' *Twitter* 12 January 2021 https://twitter.com/govuganda/status/1349060384490213377?s=12 (accessed 12 January 2021)

Guerini, M & Staiano, J 'Deep feelings: A massive cross-lingual study on the relation between emotions and virality' https://arxiv.org/abs/1503.04723v1 (accessed 30 July 2020)

Heilweil, R & Ghaffary, S 'How Trump's internet built and broadcast the Capitol insurrection' *VoxMedia* 8 January 2021 https://www.vox.com/recode/22221285/trump-online-capitol-riot-far-right-parler-twitter-facebook (accessed 10 January 2021)

Hirsch, J 'Where Facebook's Oversight Board falls short' *Centre for International Governance Innovation* 22 October 2019 https://www.cigionline.org/articles/where-facebooks-oversight-board-falls-short (accessed 4 May 2021)

Hollyer, JR, Rosendorff, BP & Vreeland, JR 'Fake news is bad news for democracy' *The Washington Post* 5 April 2019

© University of Pretoria

https://www.washingtonpost.com/politics/2019/04/05/fake-news-is-bad-news-democracy/ (accessed 15 December 2019)

Ibrahim, N 'Kano court sentences singer to death for blasphemy' *The Premium Times* 10 August 2020 https://www.premiumtimesng.com/news/headlines/407936-kano-court-sentences-singer-to-death-for-blasphemy.html (accessed 13 October 2021)

Ilori, T 'A socio-legal analysis of Nigeria's Protection from Internet Falsehoods, Manipulations and other Related Matters Bill' *AfricLaw* 5 December 2019 https://africlaw.com/2019/12/05/a-socio-legal-analysis-of-nigerias-protection-from-internet-falsehoods-manipulations-and-other-related-matters-bill/ (accessed 24 August 2021)

Ilori, T 'Content moderation is particularly hard in African countries' *Slate* 21 August 2020 https://slate.com/technology/2020/08/social-media-content-moderation-african-nations.html (accessed 21 August 2020)

Ilori, T 'Facebook's censorship of the #EndSARS protests shows the price of its content moderation errors' *Slate* 27 October 2020 https://slate.com/technology/2020/10/facebook-instagram-endsars-protests-nigeria.html (accessed 13 March 2021)

Internet Society 'Full IP access timeline' https://www.internetsociety.org/wp-content/uploads/2017/09/history_internet_africa.pdf (accessed 17 February 2021)

ISPA 'Press release: ISPA recognised as an Industry Representative Body' 20 May 2009 https://ispa.org.za/press_releases/ispa-recognised-as-an-industry-representative-body/ (accessed 13 August 2021)

Jacinto, L '#Jan25 hashtags resurfaces twenty years after Egypt's revolution' *FRANCE 24* 25 January 2021 https://www.france24.com/en/africa/20210125-a-hashtag-resurfaces-10-years-after-egypt-s-revolution-and-the-posts-are-bittersweet (accessed 20 February 2021)

Karanicolas, M 'Moderate globally, impact locally: A series on content moderation in the Global South' *Yale Law School* 5 August 2020 https://law.yale.edu/isp/initiatives/wikimedia-initiative-intermediaries-and-

© University of Pretoria

information/wiii-blog/moderate-globally-impact-locally-series-content-moderation-global-south (accessed 20 August 2021)

Karanicolas, M 'The countries where democracy is most fragile are test subjects for platforms' content moderation policies' *Slate* 16 November 2020 https://slate.com/technology/2020/11/global-south-facebook-misinformation-content-moderation-policies.html (accessed 15 December 2020)

Karombo, T 'Tanzania has blocked social media, bulk SMS as its election polls open' *Quartz Africa* 28 October 2020 https://qz.com/africa/1923616/tanzanias-magufuli-blocks-twitter-facebook-sms-on-election-eve/ (accessed 16 April 2021)

Kaye, D 'Ethiopia, the scourge of 'hate speech' & American social media' 9 December 2019 https://dkisaway.medium.com/ethiopia-the-scourge-of-hate-speech-american-social-media-952c9228e21c (accessed 2 August 2021)

Kazeem, Y 'How a youth-led digital movement is driving Nigeria's largest protests in a decade' *Quartz Africa* 13 October 2020 https://qz.com/africa/1916319/how-nigerians-use-social-media-to-organize-endsars-protests/ (accessed 16 February 2021)

Kivuva, M 'Online violence in times of COVID-19' *KICTANET* 29 May 2020 https://www.kictanet.or.ke/online-violence-in-times-of-covid-19/ (accessed 15 October 2021)

Kleinwächter, W 'History of Internet governance and challenges of tomorrow' (2016) http://3.15.112.233/wp-content/uploads/2018/01/History-of-Internet-Governance-and-Challenges-of-Tomorrow-Wolfgang-Kleinwächter-9-August-2016.pdf (accessed 12 February 2021)

Li, C & Lalani, F 'Why so much harmful content has proliferated online - and what we can do about it' *World Economic Forum* 13 January 2020 https://www.weforum.org/agenda/2020/01/harmful-content-proliferated-online/ (accessed 17 October 2020)

Manilla principles on Internet intermediaries https://manilaprinciples.org/principles.html (accessed 15 March 2021)

© University of Pretoria

Marwick, E 'Are there limits to online free speech?' *Medium* 5 January 2017 https://points.datasociety.net/are-there-limits-to-online-free-speech-14dbb7069aec (accessed 3 February 2021)

McCullagh, D 'Google's chastity belt too tight' *CNET NEWS* 23 April 2004 http://news.cnet.com/2100-1032_3-5198125.html (accessed 12 February 2021)

Media Defence 'Module 7: Cybercrimes' (2020) https://www.mediadefence.org/ereader/wp-content/uploads/sites/2/2020/12/Module-7-Cybercrimes.pdf (accessed 20 July 2021)

Media Foundation for West Africa 'Nigeria's cybercrime law being selectively applied' *IFEX* 9 October 2020 https://ifex.org/nigerias-cybercrime-law-being-selectively-applied/ (accessed 24 July 2021)

Merriam Webster 'Harm' https://www.merriam-webster.com/dictionary/harm (accessed 24 July 2020)

Merriam Webster 'Internet' https://www.merriam-webster.com/dictionary/Internet (accessed 24 July 2020)

Merriam Webster 'Online' https://www.merriam-webster.com/dictionary/online (accessed 24 July 2020)

Mokone, O 'The colonial-era laws that still govern African journalism' *AL JAZEERA* 10 March 2019 https://www.aljazeera.com/programmes/listeningpost/2019/03/colonial-era-laws-govern-african-journalism-190310080903941.html (accessed 17 June 2020)

Monye, C 'Children's online safety in Nigeria: the government's critical role' 12 September 2018 https://blogs.lse.ac.uk/parenting4digitalfuture/2018/09/12/childrens-online-safety-in-nigeria/ (accessed 1 December 2021)

Mumbere, D 'Fake news fuels xenophobic tensions in South Africa' *Africa News* 6 September 2019 https://www.africanews.com/2019/09/06/fake-news-fuels-xenophobic-tensions-in-south-africa// (accessed 15 July 2020)

© University of Pretoria

NHRC 'What are human rights' https://www.nigeriarights.gov.ng/about/nhrc-mandate.html (accessed 15 August 2021)

O'Flaherty, M 'Limitations on freedom of opinion and expression: Growing consensus or hidden fault lines' (2012) 106 *Proceedings of the Annual Meeting (American Society of International Law: Confronting Complexity* 348

O'Reilly, T & Battelle, J 'Web squared: Web 2.0 five years on' (2009) http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper (accessed 15 February 2021)

O'Reilly, T 'What is web 2.0' (2005) www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1 (accessed 12 February 2021)

Ogbondah, CW 'Nigerian Press under Imperialists and Dictators: 1903-1985' *Paper presented at the International Division of the AETMC conference at Portland, Oregon, July 2, 1988* https://files.eric.ed.gov/fulltext/ED296319.pdf (accessed 24 July 2021)

OHCHR 'Moderating online content: Fighting harm or silencing dissent' 23 July 2021 https://www.ohchr.org/EN/NewsEvents/Pages/Online-content-regulation.aspx (accessed 20 August 2021)

OHCHR 'National human rights institutions: History, principles, roles and responsibilities' (2010) https://www.ohchr.org/Documents/Publications/PTS-4Rev1-NHRI_en.pdf (accessed 24 August 2021)

Ojo, J 'Did Twitter fund #EndSARS protests as Lai claimed?' *The Cable* 5 June 2021 https://www.thecable.ng/fact-check-did-twitter-fund-endsars-protests-as-lai-claimed (accessed 10 June 2021)

Okunola, A & Mlaba, K 'From #EndSARS to #AmINext: How young Africans used social media to drive change in 2020' 23 December 2020 *Global Citizen* https://www.globalcitizen.org/en/content/endsars-aminext-young-african-social-movement-2020/ (accessed 16 November 2021)

© University of Pretoria

Olojo, A & Allen, K 'Social media and the state: challenging the rules of engagement' 24 June 2021 Institute for Security Studies <https://issafrica.org/iss-today/social-media-and-the-state-challenging-the-rules-of-engagement> (accessed 16 November 2021)

Open Rights Group 'ORG policy responses to Online Harms White Paper' (2019) https://modx.openrightsgroup.org/assets/files/reports/report_pdfs/ORG_Policy_Lines_Online_Harms_WP.pdf (accessed 13 July 2021)

Ordway, D 'Information disorder: The essential glossary' *Journalists' Resource* 23 July 2018 https://journalistsresource.org/studies/society/internet/information-disorder-glossary-fake-news/ (accessed 15 July 2020)

Organisation for Security and Co-operation in Europe (OSCE) 'Joint Declarations' https://www.osce.org/fom/66176 (accessed 1 December 2021)

Organisation for Security and Co-operation in Europe 'Joint declaration on freedom of expression and 'fake news', disinformation and propaganda' 3 March 2017 https://www.osce.org/files/f/documents/6/8/302796.pdf (accessed 24 August 2021).

Oversight Board
'Ensuring respect for free expression, through independent judgement'
https://oversightboard.com (accessed 15 February 2021)

Oversight Board 'Meet the Board' https://www.oversightboard.com/meet-the-board/ (accessed 1 December 2021)

Owen, T 'Introduction: Why platform governance?' *Centre for International Governance Innovation* 28 October 2019 https://www.cigionline.org/articles/introduction-why-platform-governance (accessed 4 February 2021)

Papenfuss, M 'Twitter agrees to block Tweets critical of India government's COVID-19 response *HuffPost* 25 April 2021 https://www.huffpost.com/entry/india-twitter-covid-surge-cremations_n_6084cfa3e4b02e74d21a6ef2 (accessed 2 May 2021)

© University of Pretoria

Pavelka, J 'Descriptions, prescriptions, and the limits of knowledge' (2020) http://www.knowledgedefinition.com/CH6TheoryOfDefinition0416.pdf (accessed 15 September 2020).

Pen International 'Blogger arrested over critical posts, held incommunicado' 30 October 2008 https://ifex.org/blogger-arrested-over-critical-posts-held-incommunicado/ (accessed 14 October 2021)

Perišin, D & Opić, S 'Connection between exposure to Internet content and violent behaviour among students' (2013) *The 1st International conference "Research and education challenges toward the future* https://files.eric.ed.gov/fulltext/ED565463.pdf (accessed 15 June 2020)

Rickcard, B 'Words that started a riot: An appraisal of the law against sedition & criminal libel in Kenya' (2019) https://bit.ly/3ALym00 (accessed 19 October 2020)

Rozen, J 'Colonial and apartheid-era laws still govern press freedom in southern Africa' *Quartz Africa* 7 December 2018 https://qz.com/africa/1487311/colonial-apartheid-era-laws-hur-southern-africas-press-freedom/ (accessed 17 June 2020)

Sagasti, F 'A human rights approach to democratic governance and development' in UNOHRC (ed) *Realizing the right to development: Essays in commemoration of 25 years of the United Nations Declaration on the Right to Development* (2013) https://www.ohchr.org/Documents/Issues/Development/RTDBook/PartIIChapter9.pdf (accessed 15 February 2021)

Santa Clara Principles https://www.eff.org/files/2015/07/08/manila_principles_background_paper.pdf (accessed 15 March 2021)

Scheffler, A 'The inherent danger of hate speech legislation: A case study from Rwanda and Kenya on the failure of a preventative measure' (2015) *fesmedia Africa series* https://library.fes.de/pdf-files/bueros/africa-media/12462.pdf (accessed: 25 June 2021)

Sidner, S & Simon, M 'Heading 'into a buzzsaw': Why extremism experts fear the Capitol attack is just the beginning' *CNN* 18 January 2021

© University of Pretoria

https://edition.cnn.com/2021/01/16/us/capitol-riots-extremism-threat-soh/index.html (accessed 19 January 2021)

Singh, S 'Everything in moderation: An analysis of how Internet platforms are using artificial intelligence to moderate user- generated content' *New America* 22 July 2019 https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/ (accessed 15 March 2021)

Smith, D & Torres, L 'Timeline: A history of free speech' *The Guardian* 5 February 2006 https://www.theguardian.com/media/2006/feb/05/religion.news (accessed 23 May 2020)

Solomon, S 'Cambridge Analytica played roles in multiple African elections' *Voice of America* 22 March 2018 https://www.voanews.com/africa/cambridge-analytica-played-roles-multiple-african-elections (accessed 18 August 2020).

Stubbs, J 'French and Russian trolls wrestle for influence in Africa, Facebook says' *Reuters* 15 December 2020 https://www.reuters.com/article/facebook-africa-disinformation/french-and-russian-trolls-wrestle-for-influence-in-africa-facebook-says-idUKL8N2IV3NR?edition-redirect=uk (accessed 2 January 2021)

The Observer 'Museveni warns Facebook ahead of elections' *Twitter* 12 January 2021 https://twitter.com/observerug/status/1349059958885785601?s=12 (accessed 12January 2021)

Tuerk, M 'Africa is the next frontier for the Internet' *Forbes* 9 June 2020 https://www.forbes.com/sites/miriamtuerk/2020/06/09/africa-is-the-next-frontier-for-the-internet/?sh=f1e169749001 (accessed 10 August 2021).

Twitter 'About country withheld content' https://help.twitter.com/en/rules-and-policies/tweet-withheld-by-country (accessed 12 February 2020);

Twitter 'Hateful Conduct Policy' https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy (accessed 19 November 2019)

© University of Pretoria

Vetten, L 'Domestic violence in South Africa' (2014) https://issafrica.s3.amazonaws.com/site/uploads/PolBrief71.pdf (accessed 24 June 2021)

Ward, C, Polglase, K, Shukla, S, Mezzofiore, G & Lister, T 'Russian election meddling is back – via Ghana and Nigeria – and in your feed' *CNN* 11 April 2020 https://edition.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html (accessed 12 July 2020)

We are social 'Digital 2021: Digital overview report' (2021) https://wearesocial-cn.s3.cn-north-1.amazonaws.com.cn/common/digital2021/digital-2021-global.pdf (accessed 13 October 2021)

Wessing, T 'Online harms: The regulation of internet content' (2019) https://www.taylorwessing.com/download/article-online-harms.html (accessed 15 August 2020)

Woodhouse, J 'Regulating online harms' *House of Lords* 21 August 2021, https://researchbriefings.files.parliament.uk/documents/CBP-8743/CBP-8743.pdf 31-32 (accessed 23 August 2021)

Woollacott, E 'Russian trolls outsource disinformation campaigns to Africa' *Forbes* 13 March 2020 https://www.forbes.com/sites/emmawoollacott/2020/03/13/russian-trolls-outsource-disinformation-campaigns-to-to-africa/?sh=7ec387d1a263 (accessed 13 October 2021)

York, J 'The global impact of content moderation' *ARTICLE 19* 7 April 2020 https://www.article19.org/resources/the-global-impact-of-content-moderation/ (accessed 20 August 2021)

YouTube 'Hate Speech Policy' https://support.google.com/youtube/answer/2801939?hl=en (accessed 19 November 2019)

© University of Pretoria

Unpublished thesis

Unpublished: Abdulrauf, LA 'The legal protection of privacy in Nigeria: Lessons from Canada and Nigeria' unpublished LLD Thesis, University of the Pretoria, 2015 https://repository.up.ac.za/bitstream/handle/2263/53129/Abdulrauf_Legal_2015.pdf?sequence=1&isAllowed=y (accessed 15 June 2020)

Unpublished: Adjei, WE 'The protection of freedom of expression in Africa: Problems of application and interpretation of Article 9 of the African Charter on Human and Peoples' Rights' unpublished PhD thesis, University of Aberdeen, 2012

Unpublished: Jeffrey, AJ 'Media freedom in an African state: Nigerian law in its historical and constitutional context' unpublished PhD thesis, University of London, 1983

Unpublished: Saloojee, N 'The prevalence of traditional bullying and cyberbullying among University students' unpublished Masters thesis, University of Witwatersrand, 2019

Unpublished: Ugangu, W 'Normative media theory and the rethinking of the role of the Kenyan media in a changing social economic context' unpublished PhD thesis, University of South Africa, 2012 http://uir.unisa.ac.za/bitstream/handle/10500/8606/thesis_ugangu_w.pdf;sequence=

International instruments and documents

ACHPR 'Resolution to modify the Declaration of Principles on Freedom of Expression to include Access to Information and Request for a Commemorative Day on Freedom of Information' *ACHPR/Res.222(LI)2012* 2 May 2012

ACHPR 'Declaration of Principles on Freedom of Expression in Africa' (2002) https://www.achpr.org/presspublic/publication?id=3 (accessed 15 March 2020)

'African Charter on Human and Peoples' Rights' (adopted 27 June 1981, entered into force 21 October 1986) (1982) 21 ILM 58

© University of Pretoria

African Commission 'Special Rapporteur on Freedom of Expression and Access to Information' https://www.achpr.org/specialmechanisms/detail?id=2 (accessed 15 March 2020)

Commission on Human Rights 'Resolution 1993/45' 5 March 1993

Türk, D & Joinet, L 'The right to freedom of opinion and expression' *Final report, UN Doc No E/CN.4/Sub.2/1992/9/Add.1* 14 July 1992

'Declaration of Principles on Freedom of Expression and Access to Information in Africa' (2019)

'Freedom of Opinion and Expression - Annual reports' https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Annual.aspx (accessed 15 February 2020).

'Joint declaration on freedom of expression and 'fake news', disinformation and propaganda' https://www.osce.org/files/f/documents/6/8/302796.pdf (accessed 24 March 2021)

'Joint Declarations' https://freedex.org/resources/joint-declarations/ (accessed 15 April 2020)

OHCHR 'Guiding principles on business and human rights' https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf (accessed 24 August 2021).

'Provisions of the Declaration on Principles of Freedom of Expression and Access to Information Online' (2019)

'African Declaration on Internet Rights and Freedoms' (2014) https://africaninternetrights.org/sites/default/files/African-Declaration-English-FINAL.pdf (accessed 115 October 2020)

United Nation General Assembly (UNGA) 'Resolution 59(1)' 14 December 1946

© University of Pretoria

United Nations, Economic and Social Council, *Final act of the United Nations Conference on Freedom of Information* 21 April 1948 https://digitallibrary.un.org/record/3806839?ln=en (accessed 20 June 2019)

United Nations General Assembly 'Expert workshops on the prohibition of incitement to national, racial or religious hatred', para 12, A/HRC/22/17 (11 January 2011), http://undocs.org/en/ A/HRC/22/17 (accessed 1 December 2021)

United Nations Educational, Scientific and Cultural Organisation (UNESCO) 'The right to information in the time of crisis' (2020) https://en.unesco.org/sites/default/files/unesco_ati_iduai2020_english_sep_24.pdf (accessed 23 August 2021)

UNGA 'Contemporary challenges on freedom of expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' *A/71/373* 6 September 2016 http://undocs.org/en/A/71/373 (accessed 26 August 2021)

UNGA 'Disinformation and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' *A/HRC/47/25* 13 April 2021 http://undocs.org/en/A/HRC/47/25 (accessed 26 August 2021)

UNGA 'Hate speech, incitement to hatred and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Rreedom of Opinion and Expression' *A/67/357* 9 October 2019 http://undocs.org/en/A/67/357 (accessed 22 August 2020)

UNGA 'Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' *A/HRC/RES/7/36* 28 March 2008 http://undocs.org/en/A/HRC/RES/7/36 (accessed 26 July 2020)

UNGA 'Online content regulation and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' *A/HRC/38/35* 6 April 2018 http://undocs.org/en/A/HRC/38/35 (accessed 15 October 2021)

© University of Pretoria

UNGA 'Gender justice and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' *A/76/258* 30 July 2021 http://undocs.org/en/A/76/258 (accessed 26 August 2021)

'United Nations Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework' (2011)

UNGA 'General Comment No 34' *CCPR/C/GC/34* 12 September 2011 http://undocs.org/en/CCPR/C/GC/34 (accessed 26 June 2021)

UNGA 'General Comment No 35' *CCPR/C/GC/35* 26 September 2013 http://undocs.org/en/CCPR/C/GC/34 (accessed 26 June 2020)

UNGA 'Online hate speech and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' *A/74/486* 7 September 2012 http://undocs.org/en/A/74/486 (accessed 26 August 2020)

UNGA 'Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography (CRC-OP-SC)' *A/RES/54/263* 18 January 2002 http://undocs.org/en/A/RES/54/263 (accessed 26 August 2021)

United Nations, General Assembly, *The right to freedom of opinion and expression exercised through the Internet*, A/HRC/66/290 (10 August 2011), http://undocs.org/en/A/66/290 paras 20-36 (accessed 1 December 2021)

National legislation and documents

Computer Crime Proclamation No 958/2016, Ethiopia

Computer Misuse Act, 2011, Uganda

Computer Misuse and Cybercrime Act, 2018, Kenya

Constitution of the Federal Republic of Nigeria, 1996 (as amended)

Constitution of the Republic of South Africa, 1996.

© University of Pretoria

Criminal Code Act, Cap C38 Laws of the Federation of Nigeria 2004 (Criminal Code Act)

Cybercrime Act, 2015, Tanzania

Cybercrime (Prohibition, Prevention etc) Act, 2015, Nigeria

Digital Act 'Online harms White Paper: Seven expert perspectives' 8 April 2019 https://www.politico.eu/wp-content/uploads/2019/04/Seven-expert-perspectives-on-the-UK-online-harms-White-Paper-.pdf?utm_source=POLITICO.EU&utm_campaign=723cb52285-EMAIL_CAMPAIGN_2019_04_10_05_07&utm_medium=email&utm_term=0_10959edeb5-723cb52285-189780761 (accessed 15 October 2021).

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the internal market https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32000L0031

Domestic Violence Act, 1998, South Africa

Electronic Communications Act, 2008, South Africa

Electronic Transactions and Cybersecurity Act, 2016, Malawi

Films and Publications Amendment Act, 2019, South Africa

Government of Canada 'Discussion guide' https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/discussion-guide.html (accessed 21 August 2020)

Hate Speech and Disinformation Prevention and Suppression Proclamation, 2019, Ethiopia

LAW n° 2020-766 of June 24, 2020 aimed at combating hateful content on the internet, France

Mass Media and Freedom of Information Proclamation, 195), Ethiopia

© University of Pretoria

National Human Rights Commission (Amendment) Act, 2010, Nigeria

Network Enforcement Act 2017, Germany

Penal Code (Northern States) Federal Provisions Act, Cap P3, Laws of the Federation of Nigeria (Penal Code)

Penal Code Act of 1950, Uganda

Penal Code of Ethiopia, 1957

Penal Code of Tanzania, 2019

Protection from Internet Falsehoods and Manipulation and Other Related Matters (PIFM) Bill, 2019, Nigeria

Singapore's Protection from Online Falsehoods and Manipulation Act 2019 <https://sso.agc.gov.sg/Acts-Supp/18-2019/Published/20190625?DocDate=20190625> (accessed 5 February 2021)

South African Human Rights Commission Act, 2013

UK Online Harms White paper https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973939/Online_Harms_White_Paper_V2.pdf, Canada (accessed 21 August 2020)

Case law

African Commission

Constitutional Rights Project v Nigeria (2000) AHRLR 191 (ACHPR 1998)

Constitutional Rights Project, Civil Liberties Organisation and Media Rights Agenda v Nigeria (2000) AHRLR 227 (ACHPR 1999)

Egyptian Initiative for Personal Rights and INTERIGHTS v Egypt I (2011) AHRLR 42 (ACHPR 2011)

ECOWAS Court of Justice

© University of Pretoria

Amnesty International Togo vs The Togolese Republic (ECW/CCJ/APP/61/18) (2020)

Laws and Awareness Initiatives v Federal Republic of Nigeria (ECW/CCJ/JUD/16/20) (2020)

Facebook Oversight Board decisions

Case decision 2020-002-FB-UA https://oversightboard.com/decision/FB-I2T6526K/

Case decision 2020-003-FB-UA https://oversightboard.com/decision/FB-QBJDASCV/

Case decision 2020-004-IG-UA https://oversightboard.com/decision/IG-7THR3SI1/

Case decision 2020-005-FB-UA https://oversightboard.com/decision/FB-2RDRCAVQ/

Case decision 2020-006-FB-FBR https://oversightboard.com/decision/FB-XWJQBU9A/

Case decision 2020-007-FB-FBR https://oversightboard.com/decision/FB-R9K87402/

Case decision 2021-002-FB-U https://oversightboard.com/decision/FB-S6NRTDAJ/

Case decision 2021-003-FB-UA https://oversightboard.com/decision/FB-H6OZKDS3/ (accessed 25 February 2021)

Oversight Board 'Case decision 2021-011-FB-UA' https://oversightboard.com/decision/FB-TYE2766G/ (accessed 6 October 2021)

Kenya

Jacqueline Okuta & another v Attorney General & 2 others (2017) Petition 397 of 2016 eKLR (accessed 24 October 2020)

Nigeria

Paradigm Initiative & Others v Attorney General of the Federation & Others (CA/L/556/2017) (2018)

South Africa

© University of Pretoria

EFF & Others v Manuel 2021 (3) SA 425 (SCA) (17 December 2020)

Heroldt v Wills (2013 (2) SA 530 GSJ))

Mwanele Manyi v Mcebo Freedom Dhlamini Case number 36077/18 (un reported) heard in the High Court of RSA Guateng Division, Pretoria

RM v RB (2015) (1) SA 270 (KZP)

The Citizen 1978 (Pty) Ltd and others v. McBride (Johnstone and others as amicus curiae) 2011 (8) BCLR 816 (CC)

Qwelane v South African Human Rights Commission and Another 2021 SA 22 (CC)

United States

Stratton Oakmont, Inc. v. Prodigy Services Co 23 Media Law Report 1794

© University of Pretoria