



**COUNTERING ONLINE HATE SPEECH THROUGH LEGISLATIVE MEASURE: THE
ETHIOPIAN APPROACH IN A COMPARATIVE PERSPECTIVE**

by Temelso Gashaw

LLM/MA Capstone Thesis

SUPERVISOR: Professor Jeremy McBride

Central European University

© Central European University

5th of June 2020

Abstract

It is almost self-evident that the phenomenon of online hate speech is on the rise. Meanwhile, governments around the world are resorting to legislative measures to tackle this pernicious social problem. This Capstone thesis has sought to offer an overview of the legislative responses to online hate speech in four different jurisdictions and unravel their implication on the right to freedom of expression. Using a comparative methodology, the research describes how the regulation of social networking platforms in relation to hate speech is being approached by the Council of Europe, Germany, the United States of America and Federal Democratic Republic of Ethiopia. It tests the hypothesis that legal frameworks for moderation of user-generated online content can have a more detrimental effect for freedom of speech in semi-democratic countries like Ethiopia than in developed liberal democracies. Findings of this project regarding the recently enacted Ethiopian Hate Speech and Disinformation Prevention proclamation will offer some guidance for The Council of Ministers in the course of adopting implementing regulation.

Table of Contents

Abstract	i
Chapter one: Introduction.....	1
1.1 statement of the problem	1
1.2 General objective	3
1.3 Specific objective	5
1.4 Research Questions	6
1.5 Methodology	6
1.6 Scope.....	7
Chapter two: Conceptual framework	8
2.1 Definition of Hate speech.....	8
2.2 Distinctive features of online hate speech.....	13
2.3 Theoretical framework underlying the interplay between freedom of expression and online hate speech	14
2.4 The interconnection between hate speech, hate crime and social media.....	17
CHAPTER THREE – Regulatory models of online hate speech	20
3.1 The UN framework.....	20
3.2 The Council of Europe.....	21
3.2.2 European Commission against Racism and Intolerance (ECRI), General Policy Recommendation No. 15.....	22

3.2.3 Case laws of the ECtHR	24
3.2.4 Delfi v Estonia	25
3.3 German.....	31
3.3.1 General Background.....	31
3.3.2 The Network Enforcement Act (NetzDG).....	31
3.3.3 Object and main Tenets of the NetzDG.....	32
3.3.4 Scope	33
3.3.5 Transparency mechanisms and Sanctions	33
3.3.6 Its implication on Freedom of expression	34
3.4 The United States of America	35
3.4.1 The Communications Decency Act.....	35
3.5 Ethiopia	37
3.5.1 General Background.....	37
3.5.2 The New Hate Speech and Disinformation Prevention and Suppression Proclamation	39
3.5.3 Its accordance with International standards of freedom of expression	39
3.5.4 Obligations imposed on Social media service providers.....	40
3.5.5 Procedural shortcomings	41
4. Concluding remarks and recommendations.....	42
Bibliography	45

Chapter one: Introduction

1.1 statement of the problem

“Governments of the Industrial World, ... On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather.... We are creating a world where anyone, anywhere may express his or her beliefs, no matter how singular, without fear of being coerced into silence or conformity...”

John Perry Barlow, 1996

Some 20 years after J. P. Barlow's 'Declaration of the Independence of Cyberspace', we are now living in the age of social media. Recent statistics has indicated that, Facebook hosts approximately 2.5 billion monthly active users worldwide.¹ There are also 330 million registered users of Twitter.² The emergence and rapid adoption of these social networking platforms has redefined the concept of communication. Given their inexpensive, high-speed and ubiquitous nature, everyone can be a creator and distributor of contents without going through the traditional gatekeepers. Social media platforms provide an enormous opportunity of accessing and imparting information, amplifying important issues and mobilizing the public to robust socio-economic and politico-legal changes. Had it not been for such platforms, the Arab spring would have hardly yielded results ³, the *me to movement*⁴ and *Black lives matter* might have never been realized.⁵ As a true *marketplace of ideas* or as a “*modern public square*,”⁶ social media platforms can also play

¹J. Clement, ‘Statistics on Number of Facebook users worldwide’ <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/> Accessed 14 March 2020. Along with other application platforms like Messenger, WhatsApp and Instagram, which are owned by Facebook, the total number of Active users climbed to 2.9 billion.

² Ibid.

³ Anita Breuer, Todd Landman & Dorothea Farquhar ‘Social media and protest mobilization: evidence from the Tunisian revolution’, *Democratization*, 22:4, 764-792

⁴ Brünker, Judith and others, ‘The Role of Social Media during Social Movements – Observations from the #metoo Debate on Twitter.’ In *Proceedings of the 52nd Annual Hawaii International Conference on System Sciences (HICSS)*, 2356- 2357

⁵ Cascante, Diana Carolina, "Black Lives Matter: Understanding Social Media and the Changing Landscape of Social Trust" (2019). Theses and Dissertations. 3375.

⁶ Justice Kennedy, in case of *Packingham v. North Carolina*, 137 S. Ct. 1730 (2017).

an important role in enhancing democracy by widening the space for civil deliberation and thoughtful discourses which sometimes accommodates diverse political, cultural and ideological perspectives.

However, as much as they become increasingly important in dispensing incredible and unique communication opportunities, on the downside, social networks can also provide a safe space for the proliferation of illegal contents. Along with their deliberative and democratizing potentials, these platforms can also be an ideal venue for those who want to spread deeply insulting, hateful, racist, xenophobic, misogynistic and other reprehensible messages which causes genuine harm to human dignity. And in some instances, such stigmatizing and dehumanizing content may not stay purely virtual. Recent empirical researches indicated that online hate speech could potentially lead to an outbreak of real-life violence and offline hate crimes.⁷

There have long been debates about how to strike a proper balance between protection of freedom of expression and suppression of hate speech. However, doing this work in digital context and in the era of social media presents some new human right and Constitutional challenges. In addition to the issue of legitimacy, the scale and amount of power that American-born private social media companies have in the public sphere is extremely disproportional to the responsibility that they hold. Given the fact that the reach of internet transcends national borders, identification of governing law and jurisdiction has always been a major obstacle in addressing the issue of online hate speech. Different countries have different Constitutional jurisprudence, socio-political

⁷ See Müller, Karsten & Schwarz, Carlo, 'Fanning the Flames of Hate: Social Media and Hate Crime', (2018) CAGE Online Working Paper Series 373, <<https://warwick.ac.uk/fac/soc/economics/staff/crschwarz/fanning-flames-hate.pdf>> accessed 15 March 2020. See also Andrew Sakowitz, Kevin Dunn and others, *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate* (first edn. Springer 2017). See also L. Shaw, 'Hate Speech in Cyberspace: Bitterness without Boundaries' (2011) *Notre Dame Journal of Law, Ethics and Public Policy*, 25(1), 279-304. See also Mark Austin Walters, *Hate Crime and Restorative Justice: Exploring Causes, Repairing Harms* (Clarendon Studies in Criminology (first edn. Oxford University Press 2014)

values, historical context and cultural sensitivity that informs their commitment to protect or prohibit hate speech. In the United States, for instance, hate speech is constitutionally protected except when it involves “incitement of imminent lawless action”.⁸ One of the firmest theoretical justifications for such strong reluctance in prohibiting hate speech is the normative principle of *marketplace of ideas* which posits that the most effective way of countering problematic speech is more speech.⁹ Placing great significance on the value of human dignity, German and other European countries, on the other hand, adopted an approach which compromises the right to freedom of expression in order to effectively curtail the problem of hate speech.¹⁰ The main difficulty in responding to social media’s facilitation of hate speech stems from such vast divergence of approach and shift of balance in ways of understanding the acceptable limits of freedom of expression.

1.2 General objective

The main purpose of this research will be to explore the legislative responses to online hate speech in different jurisdictions and unravel their implication on the right to freedom of expression. Including the UN framework, the research will describe how the regulation of Internet platforms in relation to hate speech is being approached by Four different Jurisdictions: The Council of Europe, German, United States of America and Federal Democratic Republic of Ethiopia. These jurisdictions have been selected for different reasons. First, despite their global reach, since the most popular private social media companies like Facebook, YouTube and twitter are originally incorporated in the United States, considering US’s regulatory regime will have paramount

⁸ Brandenburg v Ohio [1969] 395 USSC 444

⁹ Abrams v. United States, 250 U.S. 616, 630 (1919) Holmes, J., dissenting

¹⁰ Winfried Brugger, The Treatment of Hate Speech in German Constitutional Law (Part I), 3 German Law Journal (2002), <<http://www.germanlawjournal.com/index.php?pageID=11&artID=21210>> accessed 15 March 2020

significance. The country's pure speech jurisprudence added to its unique approach of regulating social media companies provide an opportunity to put the issue in perspective.

Secondly, addressing the issue of online hate speech requires multi-dimensional consideration as it involves the rights to freedom of expression of social media platforms and their users, the right to access to information of third party readers and the right to have personal dignity and equality of those who were allegedly victimized by the content posted. In the case of *Delfi v Estonia*¹¹ and *MTC v Hungary*,¹² the European Court of Human Rights endeavored to demark the legal boundaries and responsibilities for Member States of council of Europe, for digital platforms and for citizens. Such an exemplary effort to find the right balance between various conflicting interests could be a source of guidance for other Jurisdictions while undertaking legislative measures to effectively tackle illegal content online and, accordingly, worthy of discussion. In addition to this, the European Commission Against Racism and Intolerance (ECRI) plays tremendous role in setting standards of hate speech regulation across 47 member states of the Council of Europe by its General Policy Recommendation (GPR).¹³ Thus, it is important to provide a brief overview of its unique and comprehensive approach of combating hate speech.

Thirdly, as argued by some scholars, Germany has created a global prototype for Intermediary liability regimes as a response to user-generated illegal online contents.¹⁴ Undoubtedly, Germany is serving as a model and setting a precedent for governments around the world. Two years after the coming into effect of Germany Network Enforcement Act (or

¹¹ *Delfi AS v. Estonia* (GC), App no. 64569/09, (ECtHR 2015)

¹² *MTE and index V. Hungary* App no. 22947/2013, (ECtHR, 2016)

¹³ European Commission against Racism and Intolerance, *General Policy Recommendation No. 15 on Combating Hate Speech*. 2015 (CRI (2016) 15) [hereinafter ECRI (2016) 15]

¹⁴ Jacob Mchangama and Joelle Fiss, *The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship*, (first ed 2019 Justitia publisher) 17

NetzDG)¹⁵ more than 14 countries¹⁶ worldwide have adopted or proposed models of intermediary liability which are broadly identical with this very legislation.¹⁷ Thus, since Germany is at the forefront in reversing the tendency of providing safe harbor to private social media companies to eschew liability for hateful content posted by their user, introducing its approach would be necessary to understand the contemporary and ever-growing trends of online hate speech regulation.

1.3 Specific objective

Much research has been conducted on the regulation of problematic online content. However, there has been no comprehensive academic study of the Ethiopian approach in this regard, particularly, after the enactment of the new proclamation for *Prevention and Suppression of Online Hate Speech and Disinformation*¹⁸ which comes to effect on March 23, 2020. My project aims to take a first step to fill this research gap by providing an overview of Ethiopia's current state of affairs in tackling the prevalence of hate speech in the digital ecosystem.

In countries like Ethiopia with polarized political history and strong ethnic tension, the proliferation of online hate speech, unless properly regulated, could have far reaching repercussions. However, given the lack of well-established democratic culture, independent judiciary, robust rule of law and human right protection system, hate speech regulations could also be utilized to silence dissents and suppress legitimate expressions. In view of that, such statute-backed regulatory scheme, be it for online or offline expressions, which fails to contextualize the

¹⁵ Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz) Act to Improve Enforcement of the Law on Social Networks of 1 September 2017 (BGBl. 13352)

¹⁶ France, UK, Australia, Russia, Honduras, Venezuela, Vietnam, Belarus, Kenya, India, Singapore, Malaysia and Philippines and Ethiopia

¹⁷ Jacob and Joelle (no. 14) 17

¹⁸ Hate Speech and Disinformation Prevention and Suppression Proclamation No. 1185 /2020, (March 23rd, 2020) Federal Negarit Gazette No. 26

existing institutional accountability mechanism could be counterintuitive and may entail disproportionate detrimental and chilling effect on the right to freedom of expression. Against such backdrop, this research will examine how should we characterize the Ethiopian approach to legal regulation of digital communications of hate speech, evaluate the challenges and prospects of implementing the newly introduced online hate speech and disinformation proclamation, analyze the implication of this law on the right to freedom of expression, describe why imposing liability on the users of social media/ authors of illegal content is, in and of itself, not sufficient to stamp out the problem of hate speech.

1.4 Research Questions

This research seeks to address the following general and specific questions: -

- What are the implications of legislative responses of different jurisdictions to tackle illegal content online on the right to freedom of expression?
- How do we characterize the Ethiopian regulatory approaches to online hate speech and its implication on freedom of expression?

1.5 Methodology

To answer these research questions and to achieve the aforementioned general and specific objectives, doctrinal, qualitative and critical methods will be employed. In analyzing how the regulation of Internet platforms in relation to hate speech is being approached by the Ethiopian legislator, given the absence of academic literature and the recentness the existing law, more descriptive and explanatory approach will be employed based on interpretive and comparative method. Thus, in addition to **Primary sources** like national legislations, international and regional human right instruments and case laws, **secondary sources** such as books, academic articles, communications, policy recommendations and non-academic sources will also be consulted.

1.6 Scope

Given the breadth of this research topic and the myriad issues it encompasses; it is worth to mention that the subject matters of this project are restricted only to online hate speech. Due to the limitations of time and space, issues related to misinformation, fake news and defamation are not within the scope of this paper. Furthermore, trying to answer the question of “when”, “how” and “who” should regulate the dissemination of hateful content on social networks, evaluating different liability regimes of content providers for third-party's hateful message hosted in their platform, and exploring regulatory mechanism for internet facilitation of hate speech particularly in the context of countries with undeveloped democratic culture needs much comprehensive research. From that perspective, it has to be noted that even though some of the concerns are raised in this project, it will, nevertheless, be relatively brief.

Chapter two: Conceptual framework

2.1 Definition of Hate speech

Because of its illusive and emotive nature¹⁹, there is no universal consensus around the precise definition of hate speech.²⁰ “Despite considerable attempts by national and supranational legislators, international conventions, Courts, civil society organizations, academic scholars and technology companies, determining the objective constitutive element of hate speech has always been a point of contention.”²¹ According to David Kaye – UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, “such lack of consensus around its meaning, may threatens legal certainty and pave the way for the restriction of legitimate expressions.”²² The impreciseness of definition of hate speech may also inhibit stake holders from drawing a clear line between incitement to violence that causes genuine physical or psychological harms and other forms of expressions which may not have the same consequences. In other word, the blurred line between legitimate criticisms and hate speech creates the challenge of “distinguish between harms that justify restrictions from those that do not”.²³ Among the existing diversity of perspectives, in this section, an overview of expressions which are typically identified by different international human right instruments as constitutive elements of hate speech will be provided.

In order to comprehend the notion of hate speech, reference is often made to the Convention on the Prevention and Punishment of the Crime of Genocide which was adopted in 1948 and

¹⁹ Alexander Brown, ‘what is hate speech? the myth of hate speech’ (2017) *Law and Philosophy* 36: pp 419–468 see also Sellars, Andrew ‘Defining Hate Speech’ Berkman Klein Center Research Publication No. 2016-20; Boston Univ. School of Law, Public Law Research Paper No. 16-48.

²⁰ Natalie Alkiviadou, ‘*Regulating Internet Hate: A Flying Pig?*’ (2016) *JIPITEC* 216

²¹ Rotem Medzini and Tehilla Shwartz Altshuler, *dealing with hate speech on social media*, (the Israel Democracy Institute first edn 2019) 5

²² United Nations, Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2018 (A/HRC/RES/38/35).

²³ David Kretzmer ‘Freedom of Speech and Racism’(1987) 8 *Cardozo Law Review* 445-513, 478.

entered into force in 1953.²⁴ Among other things, it has asserted that state parties shall criminalize *public* expressions which could *directly* incite others to commit the crime of genocide. In the case of the *Prosecutor V. Jan Paul Akayesu*²⁵ the International Criminal Tribunal for Rwanda emphasized that to be considered as an incitement the expression has to be *public* and *direct*. To be more precise, the expression must be unambiguous in its intended context to fall under the test of direct and public incitement.²⁶

Even though the phrase “hate speech” is not explicitly mentioned in it, Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD)²⁷ is also another important human right instrument to conceptualize the notion. It provides that “along with condemning all propaganda, state parties are obliged to punish those engaged in the dissemination of ideas based upon racial superiority or hatred, Incitement to racial discrimination, acts of violence or incitement to such acts against any race or group of persons of another color or ethnic origin.”²⁸

The important distinctive features of ICERD is that it imposes a duty on state parties to *criminalize* acts or expressions which are referred in the convention and explanatory recommendation of its committee.(emphasis added) In General recommendation no. 15 and in the case of *Gelle v Denmark* the committee explicitly indicated that in addition to impose criminal

²⁴ UN General Assembly, *Convention on the Prevention and Punishment of the Crime of Genocide*, 9 December 1948, United Nations, Treaty Series, vol. 78 <<https://www.refworld.org/docid/3ae6b3ac0.html>> accessed 16 march 2020

²⁵ ICTR, *The Prosecutor v. Jean-Paul Akayesu*, ICTR-96-4-T, September 1998

²⁶ In the case *Prosecutor v. Nahimana, Barayagwiz and Ngeze* the tribunal again explicitly indicated the distinction between hate speech and direct and public incitement to commit genocide.

²⁷ UN General Assembly, *International Convention on the Elimination of All Forms of Racial Discrimination*, 21 December 1965, United Nations, Treaty Series, vol. 660, <<https://www.refworld.org/docid/3ae6b3940.html>> (accessed 20 March 2020)

²⁸ Ibid.

punishment, states are also required to effectively enforce such punishment.²⁹ ICCPR, in contrast, placed more lenient obligation which needs only *prohibition* by law that may not necessarily be criminalization.

Along with any propaganda of war, Article 20 (2) of the ICCPR prohibits “*advocacy of national racial or religious hatred that constitutes incitement, hostility or violence.*”³⁰ Some argued that this provision is over-broad and comprised ambiguous words like “advocacy” and “incitement” which could be interpreted in different way based on subjective standards. Article 19 of ICCPR has also provided **the three-prong test** to legitimately restrict the right to freedom of expression. In addition to the requirement of legality with sufficient precision and proportionality, the Convention set down lists of legitimate grounds for limiting freedom of expression. The curtailment of hate speech could be justified, *inter alia*, “for protection of the rights or reputations of others.”³¹ At this juncture, it is worth mentioning that the UN human right committee, in its General Comment no. 34, has indicated that article 20 should be interpreted in line with the three tier test provided under Art. 19 (3).³² Therefore, “any limitation relaying on article 20 should also comply with article 19(3) of the convention”³³

Taking into account these different kind restrictions, in 2012, the annual report of UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression classified hate speech into three categories.³⁴ First, the most severe forms of hate

²⁹ Gelle v Denmark, Communication no. 34/2004 (15 March 2006) CERD/C/68/D/34/2004, para. 7.3

³⁰ International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR), art 20(2)

³¹ Ibid. Art. 19 (3)

³² HRC, General Comment No. 34, para. 50 see also Ross v. Canada, Communication No. 736/1997, 18 October 2000

³³ Parmar, S. ‘Uprooting ‘defamation of religions’ and planting a new approach to freedom of expression at the United Nations. In T. McGonagle & Y. Donders (Eds.), *The United Nations and Freedom of Expression and Information: Critical Perspectives* (Cambridge University Press 2015) 373, 427

³⁴ Special Rapporteur to the General Assembly, (2012), A/76/357, para. 2

speech which **must be prohibited** under international law. Second, certain forms of hate speech which **may be prohibited** under international law even if they do not reach the threshold of incitement. And thirdly, the least severe forms of ‘hate speech’ which **must not be subject to legal restrictions** under international law despite its potential to raise concern in terms of intolerance and discrimination.³⁵

The Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred is also considered as another most important soft law instrument to understand the notion of hate speech. Even though it fails to encompass other recognized grounds of discrimination aside from those explicitly provided in Article 20(2) *i.e.* national, racial or religious grounds, the Rabat Plan of Action established the **six-part threshold** test to restrict freedom of expression on the ground of hate speech. These thresholds are (1) “the social and political **context**”, (2) “status of the **speaker**”, (3) “**intent** to incite the audience against a target group”, (4) “**content** and form of the speech”, (5) “**extent** of its dissemination” and (6) “**likelihood** of harm, including imminence.” However, it is the researcher's opinion that these thresholds are too high in a sense that it is possible to encounter virulent and hate-laden expressions which may not fulfill the above cumulative requirements but constitutes incitement to discrimination, hostility or violence.

Within the framework of Council of Europe, Committee of Ministers recommendation no. 97(20) defined hate speech as:

“all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by

³⁵ Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred (11 January 2013) A/HRC/22/17/Add.4 para. 12

aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin”³⁶

In Addition, the European Commission against Racism and Intolerance (ECRI) has produced general policy Recommendation (GPR) No. 15 on combating Hate Speech, which offers the most extensive definition on hate speech. According to ECRI’s GPR No. 15, “hate speech” shall mean:

“The advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons and the justification of all the preceding types of expression, on the ground of ‘race’, colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status”³⁷

The lists of protected characteristics are comparatively long – it includes 10 grounds and it is non-exhaustive. Furthermore, “incitement is not only linked to hatred but also to denigration and vilification of a person or group of persons.”³⁸ Consistent with the Jurisprudence of the ECtHR, the Commission has also recognized that “hate speech may take the form of the public denial, trivialization, justification or condonation of crimes of genocide, crimes against humanity or war crimes which have been found by courts to have occurred, and of the glorification of persons convicted for having committed such crimes”³⁹

The term “hate speech” does not appear in the text of ECHR. It has been explicitly mentioned by the ECtHR for the first time in the case of *Sürek v. Turkey*⁴⁰ in 1999. As it will be discussed in the next chapter, when dealing with cases concerning hate speech, the Strasbourg Court adopted two

³⁶ Council of Europe, Committee of Ministers, *Appendix to Recommendation no. R (97) 20*. 1997 (Recommendation no. (97) 20) [hereinafter Recommendation no. R (97) 20]

³⁷ ECRI (2016) 15 para 6

³⁸ Pejchal Viera, *Hate Speech and Human Rights in Eastern Europe* (London: Routledge 2020). 181 <<https://doi.org/10.4324/9781003005742>>_ (accessed April 2020)

³⁹ ECRI (2016) 15 para 8

⁴⁰ *Sürek v. Turkey*, ECtHR (GC), App no. 26682/95, (1999),

approaches. While in certain cases the Court invoked Article 17 of the Convention (the 'abuse clause') and declare the application inadmissible, in other cases, the Court tend to apply the three-prong test enshrined in article 10(2) of the Convention. Hence, the Court prefers to determine and analyze issues related to hate speech in a case-by-case bases.

2.2 Distinctive features of online hate speech

Although it has been described in different ways with different magnitude, hate speech is not a new phenomenon.⁴¹ It predates the internet. Nevertheless, technological advancement has presented a distinctive challenge as it widened the avenue for the creation of radicalized hate groups.⁴² The first unique feature of online hate speech is its effortlessness of securing **anonymity** of the perpetrator. Some argues that anonymity facilitates the exercise of freedom of expression as it provides a possibility to share ideas without the fear of being persecuted.⁴³ On the other hand, such Anonymity may also seize as an opportunity to disseminate reprehensible and illegal expressions without disclosing once own personality and without any kind of fear of being held accountable.⁴⁴ It may empower haters to utter their harmful content undauntedly.

The other distinctive feature of online hate speech is its **Instantaneousness**.⁴⁵ Because of the very nature Internet, the time delay between having a thought or feeling and expressing it to a particular individual who is located in a long distance away, or to a group of likeminded people or

⁴¹ SELMA 'Hacking Online Hate': Building an Evidence Base for Educators <<http://www.hackinghate.eu/>> accessed April 2020

⁴² Jems Banks, 'European Regulation of Cross-Border Hate Speech in Cyberspace: The Limits of Legislation' (2011) *European Journal of Crime, Criminal Law and Criminal Justice*, 19(1), 1,13. <https://doi.org/10.1163/157181711X553933> (accessed April 2020)

⁴³ Criza veliz 'Online Masquerade: Redesigning the Internet for Free Speech Through the Use of Pseudonyms' (2019) *Journal of Applied Philosophy*, Vol. 36, No. 4, 644

⁴⁴ Alexander Brown, 'What Is so Special about Online (as Compared to Offline) Hate Speech?' (June 2018): *Ethnicities* 18, no. 3, 297–326. <[10.1177/1468796817709846](https://doi.org/10.1177/1468796817709846)> (Accessed 18 march 2020)

⁴⁵ Gagliardone, Ignio and others, 'Countering Online Hate Speech' (UNESCO, 2015) <<https://bit.ly/2HXSnPO>>(Accessed 18 march 2020)

to a mass audience can be a matter of seconds.⁴⁶ Given internet's ease of access, size of audience, boundarylessness and inexpensiveness online hate speech causes grater social and psychological impact on the victim. As the result of its **itinerant** nature deleted content could be re-appear in different platform, account or page when the perpetrator wishes to do so. Hence, hateful content can be immortalized and may entail constant and irreparable damage.

The third distinctive feature is its **multijurisdictional** nature. As indicated in the introductory chapter, internet transcends national borders. Harmful or dehumanizing content created in one jurisdiction may affect vulnerable groups in another jurisdiction. In such a case, identification of proper Court with binding jurisdiction and determination of governing law becomes a challenge to address the issue of online hate speech. The case of *Yahoo!, Inc. V. Lic*⁴⁷ is a textbook example with regard to global reach of internet and its challenge in identification of proper jurisdiction especially when there is a considerable difference in ways of understanding the acceptable limits of freedom of expression.⁴⁸

2.3 Theoretical framework underlying the interplay between freedom of expression and online hate speech

Conceptualization of freedom of expression and demarketing its boundaries has long been contested from legal, philosophical and political perspectives by different scholars.⁴⁹ Principally, there are three theoretical foundations which serves as response to the question of what inherent

⁴⁶ Ibid.

⁴⁷ *Yahoo! Inc. v. La Ligue Contre Le Racisme Et L'antisemitisme and L'union Des Etudiants Juifs De France*

⁴⁸ Okoniewski, Elissa A, 'Yahoo!, Inc. v. LICRA: The French Challenge to Free Expression on the Internet' (2002) *American University International Law Review* 18, no. 1, 295-339

⁴⁹ Oltmann, Shannon M., 'Intellectual Freedom and Freedom of Speech: Three Theoretical Perspectives' (2016). *Information Science Faculty Publications*. 30. 158

value can be found in freedom of expression.⁵⁰ The first and most predominant one is **promotion of marketplace of ideas** or enhancement of the pursuit of truth. In “*On Liberty*”, John Stuart Mill argued that “it is only by uninhibited public discourse and deliberations that theories, knowledge, truth and even society can progress.”⁵¹ His approach adheres that the best way to ensure the advancement of truth is unfettered and free expression of opinion irrespective of its immorality and falsity.⁵² In a way which resonate with formulations from John Stuart Mill, Justice Oliver W. Holmes’s famous dissent in *Abrams v. United States* asserted that “the best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which citizens’ wishes safely can be carried out”.⁵³ However, “Mill's theory provides no justification for an absolute ban on efforts to regulate hate speech.” Accordingly, scholars like Keith N. Hylton contended that “in cases where speech contravenes the law which protect citizens from harassment, threats, and intimidation, Mill's theory implies that a state should be free to enhance penalties for hate speech.”⁵⁴ Furthermore, Mill’s Harm principle⁵⁵ can provide us the moral justifications or philosophical benchmark for prohibition of hate speech.⁵⁶

⁵⁰ George Rutherglen, ‘Theories of free speech’ (1987) *Oxford Journal of Legal Studies* 7, no. 1 115–24 in Pejchal, V. (2020). *Hate Speech and Human Rights in Eastern Europe*. London: Routledge, <<https://doi.org/10.4324/9781003005742>> (accessed April 2020)

⁵¹ John Stuart Mill, *On Liberty* (Penguin Adult, 2010)

⁵² Victoria Nash, *Analyzing Freedom of Expression Online: Theoretical, Empirical, and Normative Contributions*. (Oxford Handbook of Internet Studies 2013).

⁵³ *Abrams v. United States*, 250 U.S.Ct 616 (1919) Holmes, J., dissenting. For further discussion on combating hate speech by uninhibited free expression, See Nadine Strossen, *Hate: Why We Should Resist It with Free Speech, Not Censorship*, Inalienable Rights (New York: Oxford University Press, 2018). Alexander Brown, *Hate Speech Law: A Philosophical Examination*, Routledge Studies in Contemporary Philosophy (New York: Routledge, 2015).

⁵⁴ Hylton Keith ‘Implications of Mill's Theory of Liberty for the Regulation of Hate Speech and Hate Crimes’ (1996) *The University of Chicago Law School Roundtable*: Vol. 3: Iss. 1. 35, 57

⁵⁵ According to Mill’s Harm principle, “the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others.” And since hate speech creates social as well as individual harm its restriction can be morally defensible.

⁵⁶ Viera (no 38) 26

The second theoretical foundation of free speech is its necessity to a healthy functioning of democracy. Placing heavy emphasis on political speech, Alexander Meiklejohn argued that freedom of expression paves the way for citizens to make an informed political decision and to hold their government accountable which ultimately enhance self-governance and participatory democracy.⁵⁷ Nevertheless, should hate speech be protected for the sake of democracy? if we accept democracy as a system of government “that best expresses respect for citizens as free and equal”⁵⁸ or as “a fair mode of decision making in the face of reasonable disagreement about justice,”⁵⁹ it could not be objectionably diminished by laws banning hate speech.⁶⁰

By going beyond instrumental value, the third predominant theoretical justification states that free speech is the vehicle through which individuals assures his or her self-fulfillment.⁶¹ According to this line of argument, personal autonomy cannot effectively be exercised without free expression. In “*A theory of Justice*”, Rawls pointed out that, being an end in itself, freedom of thought and expression underpins the acknowledgment of human autonomy.⁶² However, even though individual autonomy is generally conceived as valuable, “it should always be weighed against other normative commitments, such as equality and dignity of citizens which are the primary focuses of hate speech legislations”⁶³.

⁵⁷ Alexander Meiklejohn, *Political Freedom: The Constitutional Powers of the People*, (Oxford University Press 1965).

⁵⁸ Corey Brettschneider, *When the State Speaks, What Should It Say? How Democracies Can Protect Expression and Promote Equality*, NJ: (Princeton Univ. Press 2012). Cited in Jeffrey W. Howard, ‘Free Speech and Hate Speech’ (2019) Annual. Rev. Political Sci. 22:93–109

⁵⁹ Jeremy Waldron, *Law and Disagreement* (Oxford, UK: Oxford Univ. Press 1999)

⁶⁰ Jeffrey W. Howard, ‘Free Speech and Hate Speech’ (2019) Annual. Rev. Political Sci. 22:93–109
<<https://doi.org/10.1146/annurev-polisci-051517-012343>> (accessed on May 20, 2020)

⁶¹ Alexander Tsesis, ‘Free Speech Constitutionalism’ (February 3, 2014) University of Illinois Law Review, Vol. 20
<<https://ssrn.com/abstract=2390234> or <http://dx.doi.org/10.2139/ssrn.2390234>> (accessed on April 2, 2020)

⁶² John Rawls. *A Theory of Justice* (Harvard University Press 1971)

⁶³ Haward, (no 60) 7

On the other hand, while some contemporary legal philosophers such as Ronald Dworkin argued that “restrictions on speech subverts democratic legitimacy”⁶⁴, others like Jeremy Waldron, attempted to conceptualize free speech by taking Immanuel Kant's perception of human dignity as fundamental component and advocate for the prosecutions of hate speech.⁶⁵ In his book entitled “*the Harm of hate speech*”, Waldron noted that “since there is a public good of inclusiveness that our society should embrace, the inherent value conflict between absolute free speech and human dignity should be resolved in favor of the latter.”⁶⁶

In view of the fact that individual's ability to function in society depends not just on their individualized reputation but the general honor and social standing of the group to which they belong, hate speech against identified group causes emotional pain, distress and intimidation to its members.⁶⁷ Hate speech “undermine the targets’ citizenship, their equal status and their entitlement to basic justice by associating ethnicity, race or religion with conduct or attributes that disqualify them from being treated as members of society in good standing.”⁶⁸ It has also been argued that blanket protection of all speech, irrespective of their social or psychological impact, emanates from the hidden motive of sustaining the *status quo* and the desire to continue to oppress and subordinate vulnerable groups.⁶⁹

2.4 The interconnection between hate speech, hate crime and social media

The destructive effect of hate speech is not restricted to emotional or psychological damage on individual level. It is also self-evident that large-scale atrocities like genocide has always been

⁶⁴ James Weinstein, ‘Hate Speech Bans, Democracy, and Political Legitimacy’ (2017). Constitutional Commentary. 529

⁶⁵ Jeremy Waldron, *The Harm in Hate Speech*. (Cambridge, MA: Harvard University Press, 2012) 96

⁶⁶ Ibid.

⁶⁷ Mari Matsuda, ‘Public Response to Racist Speech: Consider the Victim's Story’, 87 MICH. L. REV. 2320, 2341

⁶⁸ Ibid pp. 5

⁶⁹ Thomas J. Webb, Note, Verbal Poison - Criminalizing Hate Speech: A Comparative Analysis and a Proposal for the American System, (2011) 50 WASHBURN L.J. 445, 4484

accompanied by hateful expressions. The widespread anti-Semitic rhetoric propagated by *derstürmer*, a newspaper founded by Julius Streicher⁷⁰, had led to the extermination of approximately 6 million Jews.⁷¹ Isaiah Berlin described the role of hate speech during the holocaust as follows:

“The Nazis were led to believe by those who preached to them by *word of mouth or printed words* that there existed people, correctly described as sub-human, and that these persons are poisonous creature, ... If you believe it because someone you trust told you that it is true, then you arrive at the state of mind where, it is quite rational to exterminate Jews...”⁷²
(emphasis added)

The cause and effect relationship between hate speech and genocide had also been established by International Criminal Tribunal for Rwanda (ICTR) in the case of *Prosecutor V. Jan Paul Akayesu*.⁷³ “Extremist nationalist speech in the former Yugoslavia was preceded by bloody ethnic conflict in 1992.”⁷⁴ Fear mongering propaganda has also played a significant role in exacerbating the 2007’s post-election violence in Kenya which resulted in persecution of more than 1000 lives and displacement of over 600,000 citizens.⁷⁵ Therefore, it is safe to conclude that there is a direct cause and effect relationship between hate or incitement speech and hate crime.

⁷⁰ He was convicted of “incitement to genocide” and sentenced to death for committing crimes against humanity.

⁷¹ Wibke Kristin Timmermann, ‘The Relationship between Hate Propaganda and Incitement to Genocide: A New Trend in International Law Towards Criminalization of Hate Propaganda’ (2005) *Leiden Journal of International Law*, 18 pp. 257–282

⁷² Andrea Dworkin, *The Jews, Israel and women's liberation* (The free press 2000) 141

⁷³ ICTR, *The Prosecutor v. Jean-Paul Akayesu*, ICTR-96-4-T, September 1998

⁷⁴ András Sajó, *Freedom of Expression* (Institute of Public Affairs, Warsaw 2004) p. 128. In *Milosz Hodun, The Challenges Europe Must Face*” <<http://4liberty.eu/hate-speech-european-union-central-europe-poland/>> (accessed on April 5, 2020)

⁷⁵ ADEAGBO, Oluwafemi Atanda, ‘Post-Election Crisis in Kenya and Internally Displaced Persons: A Critical Appraisal’ (2014) *Journal of Politics and Law* Vol. 4, No. 2; 1

The emergence of social media networks heightened this complex problem to the next level.⁷⁶ Provocative and antagonizing online expressions are arguably on the rise, so does their repercussions. An empirical research conducted in 2019 indicated that hate speech propagated on social media platforms encourage perpetrators to carryout real-life violence in Germany.⁷⁷ Significant correlation has been established between anti-immigrant as well as anti-Muslim posts and actual attacks on Muslims and immigrants.⁷⁸ The findings of the Umati project,⁷⁹ has also indicated the direct casual link between hate speech their on social media and their devastating real-life consequences in Kenya. Indeed, social media facilitation of hate crimes has never been so evident since the 2017's devastating ethnic cleansing against Rohingya Muslim minority in Myanmar.⁸⁰ It was also indisputable that the massacre occurred in Christchurch or in Pittsburgh synagogue has been an extension of perpetrator's online activity.⁸¹

⁷⁶ Yulia A Timofeeva, 'Hate Speech Online: Restricted or Protected? Comparison of Regulations in the United States and Germany' 12 *Journal of Transnational Law and Policy* 2, 255. See also Danielle K. Citron Helen Norton, 'Intermediaries and Hate Speech: Fostering Digital Citizenship for our Information Age', *Boston University Law Review*, Vol. 91, p. 1435, 2011.

⁷⁷ Karsten Müller, and Carlo Schwarz, Fanning the Flames of Hate, 'Social Media and Hate Crime' (November 3, 2019). <<http://dx.doi.org/10.2139/ssrn.3082972>> The causal link between online hate speech and real-life violence has also been empirically studied in Barbara Perry and Patrik Olsson, 'Cyberhate: The Globalization of Hate'(2009) 18(2) *Information & Communications Technology Law*, 185-199 see also James Banks, 'Regulating Hate Speech Online' (2010) *International Review of Law, Computers & Technology* 24(3), 233-239.

⁷⁸ Karsten Müller & Carlo Schwarz, From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment 25-28 (Nov. 2, 2019); Richard A. Wilson & Jordan Kiper, Incitement in an Era of Populism: Updating Brandenburg After Charlottesville (2020) 5 *U. PA. J. L. & PUB. AFFAIRS* <<https://scholarship.law.upenn.edu/jlpa/vol5/iss2/2>>

⁷⁹ Research, Umati Final Report (2013) <<https://preventviolentextremism.info/sites/default/files/Umati%20Final%20Report.pdf>> (accessed on April 2020)

⁸⁰ Alexandra Stevenson, 'Facebook Admits It Was Used to Incite Violence in Myanmar', (The New York times Nov. 6, 2018)

⁸¹ SELMA Hacking hate, when online hate speech goes extreme: The case of hate crimes, 23rd April 2019 Available at <<https://hackinghate.eu/news/when-online-hate-speech-goes-extreme-the-case-of-hate-crimes/>> (accessed on April 2020)

CHAPTER THREE – Regulatory models of online hate speech

3.1 The UN framework

In the previous chapter, despite its double-edged ambiguity, it has been attempted to unpack the notion of hate speech from the perspective of international and regional human right instruments. Here, it is important to point out that there is no binding international treaty and supranational jurisdictional authority which specifically aimed at tackling the problem of online hate speech. Even though the challenge needs multi stakeholder's engagement and responses, "the traditional human rights law has been designed to apply to states and not to private social media companies"⁸² like Facebook and Twitter. However, in a resolution adopted by the UN Human Right Council, it is clearly indicated that *"the same rights that people have offline must also be protected online"*⁸³ In similar manner, in its General Comment No. 34 the Human Right Committee affirms that *"Article 19 of ICCPR protects all forms of expression and the means of their dissemination, including all forms of electronic and internet-based modes of expression"*⁸⁴ From this it is possible to understand that even though these international regulatory frameworks, are adopted in the pre-internet era, since the object of expressions which could legitimately be prohibited remains similar, their applicability to the digital world is warranted. Although it is not binding, the UN Guiding Principles on Business and Human Rights (UNGPs)⁸⁵ provide some

⁸² Taylor, E. (2016). The Privatization of Human Rights: Illusions of Consent, Automation and Neutrality. Global Commission on Internet Governance Paper Series, p. 3.

⁸³ United Nations General Assembly, Human Rights Council, Resolution: The right to privacy in the digital age, adopted on 22 March 2017, A/HRC/34/L.7/Rev.1.

⁸⁴ UN Human Rights Committee, General Comment No. 34 on Article 19 of the ICCPR, adopted at its 102nd session, 11–29 July 2011; <<http://www2.ohchr.org/english/bodies/hrc/docs/GC34.pdf>>. (accessed on April 2020)

⁸⁵ United Nations. (2011). Guiding principles on business and human rights: Implementing the United Nations "Protect, Respect and Remedy" framework. (A/HRC/17/31).

human right standards for companies. Indeed, in his recent thematic report⁸⁶, the UN special rapporteur on the promotion and protection of the freedom of opinion and expression has recommended that “to combat online hate speech, both state and private social media companies should reinforce the standards set by international covenants, their authoritative interpretation by Human Right Committee and the Rabat Plan of Action”.⁸⁷ According to David Kaye, to restrict online proliferation of hate speech countries should observe the three-prong tests provided in article 19 (3) of the ICCPR. Thus, restrictions must fulfill “the requirements of **legality** (specified in a precise, public and transparent law), **legitimacy** (justified to protect rights or reputations of others; national security; public order; public health or morals), and **necessity and proportionality** (the least restrictive means to achieve that aim or the test of strict scrutiny)”.⁸⁸

3.2 The Council of Europe

The Council of Europe has developed treaty based, political standard-setting and policy making strategies of curtailing hate speech while maintaining freedom of expression. As an effort to fill the gaps of Cybercrime Convention⁸⁹, an Additional Protocol has been designed in 2003 concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems.⁹⁰ And apart from the ECHR, this additional protocol is the only treaty-based strategy adopted by the Council of Europe so far.⁹¹ Even though the protocol requires states to

⁸⁶ United Nations, General Assembly, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2019 (A/74/48050).

⁸⁷ Ibid.

⁸⁸ Ibid.

⁸⁹ Council of Europe, **Convention on Cybercrime**, 23 November 2001, available at: <<https://www.refworld.org/docid/47fdfb202.html>> [accessed 3 April 2020]

⁹⁰ McGonagle, Tarlach ‘The Council of Europe against online hate speech: Conundrums and challenges’ (2012) Belgrade: Republic of Serbia, Ministry of Culture and Information <<https://hdl.handle.net/11245/1.407945>> (accessed on 4 April 2020)

⁹¹ Even though their main subject matter is not tackling hate speech, the European Convention on Trans frontier Television and Convention on preventing and combating violence against women and domestic violence does also prohibits certain forms of hate speech.

prohibit various types of expression which could be categorized as hate speech, it has also given discretion to member states not to criminalize such expressions if other effective remedies are available.⁹²

3.2.2 European Commission against Racism and Intolerance (ECRI), General Policy

Recommendation No. 15

Although it is not a Convention-based body, ECRI plays an important role in providing guidelines and setting standards of hate speech regulation which are anchored in the jurisprudence of ECtHR, binding international human right instruments and recommendations like the Rabat plan of action. Other than exclusively relaying on legal(criminal) responses, in its General Policy Recommendation no. 15, the Commission introduced various mechanism of countering hate speech through new areas of interventions. “It provides a coherent and comprehensive approaches including legal and administrative measures; self-regulatory mechanisms; effective monitoring; victim support; awareness raising and educational measures.”⁹³ Recognizing their role in magnifying the impact of hate speech and their importance as a vehicle to challenge it, ICRI encouraged States “to use of regulatory powers with respect to open digital spaces (online media platforms) to promote action to combat the use of hate speech in a way which does not violate the right to freedom of expression”⁹⁴ In this regard, ECRI pointed out the necessity of self-regulation by designing proper code of conduct, setting up content restrictions mechanism including word filtering bots, providing appropriate training for content moderators and adopting compliant mechanisms.”⁹⁵

⁹² McGonagle (no. 90) 21

⁹³ Internet governance forum, ‘Assessing hate speech and self-regulation, who and how?’ Available at <https://www.intgovforum.org/multilingual/content/igf-2018-of-31-assessing-hate-speech-and-self-regulation-who-and-how> (accessed on 4 April 2020)

⁹⁴ ECRI (2016) 15 recommendation 8

⁹⁵ Ibid, recommendation 8

Regarding the extent to which online media can be subjected to some forms of responsibility for the dissemination of illegal content in their platform, ECRI's GPR. No. 15 has made an important distinction between different kinds of remedies. As indicated in *Delfi v. Estonia*, "because of the particular nature, of the Internet, the 'duties and responsibilities' that are to be conferred on an Internet news portal for the purposes of Article 10 may differ to some degree from those of a traditional publisher, as regards third-party content."⁹⁶ In line with ECtHR's understanding, recommendation 8 identifies "the need for specific powers, subject to judicial authorization, to require the deletion of certain hate speech, the blocking of sites using hate speech, the publication of an acknowledgement."⁹⁷ And such kind of remedies are not considered as the violation of Art. 10 of the Convention, so long as they have been used for hate speech of a more serious character. As pointed out in explanatory memorandum of ECRI's GPR. No. 15, being required to take measures identified under recommendation 8 should take into account different factor like "whether or not such online media platforms took an active role on the proliferation hate speech, whether or not they were aware that their facilities were being used for this purpose, whether or not they had and used techniques to identify such use and those responsible for it and whether or not they acted promptly to stop this from continuing once they became aware that this was occurring."⁹⁸ In determining the extent to which platform providers are liable for third party's hateful content the ECtHR considers these factors constantly.

Whereas, criminal sanctions seem to be appropriate measures only in a very restricted manner. ECRI is aware of the danger that such measure could be misused to silence dissenting voices.⁹⁹ Hence, although recommendation 10 envisages the imposition of criminal responsibility,

⁹⁶ *Delfi AS v. Estonia* (GC), Application no. 64569/09, (2015)

⁹⁷ ECRI (2016) 15 recommendation 8

⁹⁸ ECRI (2016) 15, para. 150

⁹⁹ ECRI (2016) 15, para. 171

because of the potential risk that it poses on the right to freedom of expression, such kind of measure should be taken as a remedy of last resort and in line with the principle of proportionality.

3.2.3 Case laws of the ECtHR

The ECtHR has frequently held that freedom of expression constitutes one of the essential foundations of democratic society and basic conditions for the full development of every individual. In the case of *Handyside v. the United Kingdom*, The Strasbourg court decided that article 10 of the Convention protects “not only 'information' or 'ideas' that are favorably received or regarded as inoffensive, but also to those that offend, shock or disturb”.¹⁰⁰ Despite their objectionable character, “such expressions are the demands of pluralism, tolerance and broadmindedness without which there is no democratic society.”¹⁰¹ However, “although free speech is an important value, it is not the only one.”¹⁰² Other normative commitments such as “human dignity, equality, freedom to live without harassment and intimidation, social harmony, mutual respect, and protection of one’s good name and honor are also central to the good life and deserve to be safeguarded”¹⁰³ The *Erbakan v. Turkey* judgment reflects the need to “sanction or even prevent all forms of expression which spread, incite, promote or justify hatred” in order to preserve human dignity and equality.¹⁰⁴

In dealing with the issues of hate speech the Court has adopted two approaches. When the Court confronted with cases like Holocaust denial¹⁰⁵, glamorizing of neo-Nazi ideas and National Socialism,¹⁰⁶ incitement to racial hatred¹⁰⁷ etc., it invokes, *prima facie*, what is mostly known as

¹⁰⁰ *Handyside vs the United Kingdom*, ECtHR, 7 December 1976, Application No. 5493/72,

¹⁰¹ *Sunday Times vs the United Kingdom*, ECtHR, 26 April 1979, Application No. 6538/74

¹⁰² Eric Heinze, *Hate Speech and Democratic Citizenship* (OUP, Oxford 2016) 38.

¹⁰³ *Ibid* Balogh v Hungary App no 47940/99 (ECHR, 20 July 2004)

¹⁰⁴ *Erbakan v. Turkey*, App. No. 59405/00 (ECtHR, 2006)

¹⁰⁵ *Garaudy v. France*, App. No. 65831/01 (ECtHR, Inadmissibility decision 24 June 2003)

¹⁰⁶ *H., W., P. and K. v. Austria*, Appn. No. 12774/87, 62 (ECtHR Inadmissibility decision)

¹⁰⁷ *Norwood vs the United Kingdom*, App. No. 23131/03 (ECtHR, 16 Nov. 2004)

‘the abuse clause’ which is enshrined in Article 17 of the Convention and categorically exclude the applicant’s expression from substantial and procedural guarantees by Article 10 (*guillotine effect*) based on the justification that such expression repudiates the fundamental values of the Convention.¹⁰⁸ In the case of *Leroy vs France*¹⁰⁹, the Court’s reasoning implied that” Article 17 is also applicable in cases of racist, xenophobic, anti-Semitic and Islamophobic expressions”.¹¹⁰ In other cases of hate speech like *Jersild v Denmark*,¹¹¹ *Gündüz v Turkey*,¹¹² *Vejdeland and others v Sweden*¹¹³..., the ECtHR employed the **three-prong proportionality analysis** as provided in Article 10 (2) of the Convention. Having the general overview of ECtHR’s jurisprudence regarding hate speech, it is important to assess where the Court stands when it comes to hate speech disseminated online.

3.2.4 Delfi v Estonia

In the landmark case of *Delfi v. Estonia*¹¹⁴, the Grand Chamber confirmed that imposing liability on online news portal for grossly insulting, hateful and threatening comments hosted in their platform by anonymous user is not a violation of Article 10 of the Convention. *Delfi*, the biggest internet portal in Estonia, published an article entitled ‘*SLK Destroyed Planned Ice Road*’ in 2006.¹¹⁵ This publication attracted many offensive and threatening comments which latter on described as **hate speech** by the Strasbourg Court. The comments stayed for 6 weeks and taken down from the news

¹⁰⁸ W.P. and Others vs Poland, App. No. 42264/98 (ECtHR 2 September 2004); Ivanov vs Russia, App. No. 35222/04 (ECtHR, 20 Feb. 2007)

¹⁰⁹ Leroy vs France, App. No. 36109/03, (ECtHR, 2 October 2008)

¹¹⁰ Hannes Cannie and Dirk Voorhoof, ‘The Abuse Clause and Freedom of Expression in the European Human Rights Convention’ (2011) 29 NQHR 54 pp.63

¹¹¹ Jersild v Denmark, app no. 15890/89, (ECHR 23 September 1994)

¹¹² Gündüz v Turkey App No 35071/97 (ECHR, 4 December 2003)

¹¹³ Vejdeland and others v. Sweden, App. no. 1813/07 (ECtHR, sept 2012)

¹¹⁴ Delfi AS v. Estonia app no. 64669/09, (ECtHR 2015)

¹¹⁵ Ibid. para 16

portal following the request by the ‘victims’ Lawyer.¹¹⁶ By considering the news portal **as a publisher** of user generated contents, the national Court held *Delfi* liable for its failure of removed the impugned comments ‘without delay’.¹¹⁷ In its Chamber judgment the ECtHR accepted the finding of Estonian National Court and concluded that the interference on *Delfi*’s right to freedom of expression was prescribed by law and pursue legitimate aim protecting the reputation and rights of others.¹¹⁸ As the news portal knowingly allowed the possibility of giving comment by anonymous users it had been considered by the Court “to have assumed certain responsibility for such comments.”¹¹⁹ Since the webpage has integrated the reader’s comment into its portal for commercial benefit, and most importantly, since *Delfi* exercised “a substantial degree of control over the comments published on its platform,”¹²⁰ the Chamber asserted that *Delfi* should be considered as content provider rather than purely technical service provider (intermediary), and therefore assume certain level responsibility for an infringement of other persons’ reputation by the comments of its anonymous users.”¹²¹ Accordingly, the Court concluded that the interference was a justified and proportionate as provided in Article 10(2) of the Convention. The Grand Chamber has confirmed this finding by looking at four elements. To put in a nutshell, first, it has pointed out the unique nature of internet and emphasized that “the duties and liabilities of online intermediaries for third- party illegal content may vary from traditional publishers.”¹²² Secondly, the court considered the nature and context of the comment and arrived at the conclusion that “the majority of the impugned comments amounted to hate speech or incitements to violence and as

¹¹⁶ Ibid. para 20

¹¹⁷ Ibid, Para. 65

¹¹⁸ Ibid, Para. 62 and 63

¹¹⁹ Ibid, Para. 65

¹²⁰ Ibid, Para. 65

¹²¹ Ibid, Para. 65

¹²² Ibid, Para. 113

such did not enjoy the protection of Article 10”.¹²³ Thirdly, the Court implied the potential liability of anonymous commenter¹²⁴ (real name policy as a possible solution to exempt news portal from liability for user generated illegal content). Fourthly, the Court has examined necessary measures taken by Delfi as *professional and commercial news portal* to prevent or remove these hateful and threatening remarks posted in its platform. In this regard the Court emphasized that:

“where third-party user comments are in the form of hate speech and direct threats to the physical integrity of individuals, the Court considers that the rights and interests of others and of society as a whole may entitle Contracting States to impose liability on Internet news portals, without contravening Article 10 of the Convention, if they fail to take measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties”¹²⁵

Finally, the Grand Chamber assessed whether the domestic Court had rightly balanced the rights of *Delfi* to impart information via its platform with the protection of personality rights of others based on relevant and sufficient grounds. And by fifteen votes to two, it concluded that despite the functioning of notice-and-take-down facility, the Estonian courts’ finding of liability against *Delfi* has been a justified and proportionate restriction on the portal’s freedom of expression.

As aptly argued by the two dissenting judges, the requirement of removing the alleged comments “without delay and even without notice”, disregarded the concept of “actual knowledge”¹²⁶ and introduced strict liability regime¹²⁷ based on pure promptness”¹²⁸ Such an

¹²³ Ibid para. 140. However, the two dissenting judges observed that such characterization remains” non- specific” and “murky” See Joint Dissenting Opinion of Judges Sajó and Tsotsoria, paras. 28 and 29, respectively.

¹²⁴ Ibid, Para. 147

¹²⁵ Ibid, Para. 153 and para 159

¹²⁶ Joint Dissenting Opinion of Judges Sajó and Tsotsoria, paras 23

¹²⁷ Milica nesiic, online hate speech against strict internet intermediaries’ liability as a solution for ensuring victim’s protection, (2018) Central European University Master’s thesis <http://www.etd.ceu.edu/2019/nesic_milica.pdf >

¹²⁸ Jørgensen, Rikke Frank, (edit) *Human rights in the age of platforms*, Tarlach McGonagle, The Council of Europe and Internet Intermediaries: A Case Study of Tentative Posturing, Cambridge, MA: The MIT Press, [2019] pp 228

approach may lead to Collateral censorship and proactive monitoring of third-party content.¹²⁹ It has also been underscored that the commercial nature of the news portal may not necessarily avoid the potential of comment sections for facilitating individual contributions to public debate.¹³⁰ Indeed, the Grand Chamber restricted the impact of its judgment emphasizing that “the case did not concern to social media platforms or other fora on the Internet where ... the content provider may be a private person running the website or a blog as a hobby.”¹³¹ However, the two dissenting judges described such an approach as a “damage control” stating that “Freedom of expression cannot be a matter of a hobby”.¹³²

Post-Delfi developments

Subsequent to *Delfi*, the ECtHR passed another judgement on similar subject matter in the case of *MTE and Index.hu ZRT v Hungary*.¹³³ As it has done in *Delfi*, the court has applied the five standards to find the right balance between various conflicting interests. These important parameters are: (1) “the context and content of the impugned comments, (2) the liability of the authors of the comments, (3) the measures taken by the website operators and the conduct of the injured party, (4) the consequences of the comments for the injured party, and (5) the consequences for the applicants.”¹³⁴ Nevertheless, unlike *Delfi*, the Court found violation of Article 10 of the Convention basically for two reasons. First, in the case of *MTE* the nature of comment was “notably devoid of the pivotal element of hate speech.”¹³⁵ The court asserted that the alleged

¹²⁹ Joint Dissenting Opinion of Judges Sajó and Tsotsoria para. 17

¹³⁰ Joint Dissenting Opinion of Judges Sajó and Tsotsoria, para 27

¹³¹ *Delfi AS v. Estonia* app no. 64669/09, (ECtHR 2015), Para 116

¹³² Joint Dissenting Opinion of Judges Sajó and Tsotsoria, para 25

¹³³ *MTE and index V Hungary* App no. 22947/2013, (ECtHR, 2016)

¹³⁴ *Ibid*, para.63 It is important to mention that, these parameters are being used by the Court frequently. See for example *Rolf Anders Daniel Pihl v. Sweden*, *Payam Tamiz v. the United Kingdom and Høiness v. Norway*. In addition to this, in its Recommendation CM/Rec (2018)2 the committee of ministers has adopted similar approaches.

¹³⁵ *Ibid* para 77

comments were ‘offensive and vulgar’, but they “did not constitute clearly unlawful speech’ and ‘certainly did not amount to hate speech or incitement to violence.”¹³⁶ The court’s inadmissibility decision in the case of *Pihl v. Sweden*¹³⁷ further demonstrate the importance of distinguishing classic defamatory remarks from hate speech. The court found that an imposition of liability on the news portal for utterances which are not clearly unlawful cannot be justified.¹³⁸ *Secondly*, the Court observed that the “domestic court had failed to strike a proper balance between the applicants right to freedom of expression and real estate website’s right to respect for its commercial reputation.”¹³⁹

Unlike in *Delfi* the Court has not taken into account the commercial and professionally managed character of the news portal in the judgment of *MTE*.¹⁴⁰ It is also important to note that, when the portal is considered as a content provider and involved in modifying user generated comments or undertake considerable editorial function, it can be argued that such portals are best placed to block, filter or remove illegal contents. Hence, as the ECtHR ascertained in *Delfi* (*even though it is difficult to qualify Delfi as an editor of the users’ comment*), imposing liability on such host for its failure to take down harmful contents could be justified. However, when the portal is considered as a mere intermediary internet service provider where there is initially complete freedom for people to post what they want or where the platform provider does not offer any content, as in the cases of social medias, the strict liability regime that the court has introduced in *Delfi* is not applicable. Such an intermediary, which simply provides a podium for sharing can only be held

¹³⁶ Ibid. para 64

¹³⁷ *Pihl v. Sweden*, ECHR (judgment) (no. 74742/14)

¹³⁸ Ibid Para. 81

¹³⁹ Ibid Para. 79

¹⁴⁰ Dirk Voorhoof, Blog Symposium “Strasbourg Observers turns ten” (2): The Court’s subtle approach of online media platforms’ liability for user-generated content since the ‘Delfi Oracle’ Strasbourg observer 10 April 2020, available at <https://strasbourgobservers.com/2020/04/10/the-courts-subtle-approach-of-online-media-platforms-liability-for-user-generated-content-since-the-delfi-oracle/#more-4592> (accessed on 22april 2020)

liable for third party's illegal content after the application of 'notice-and-take-down' rule. *i.e.* when the online platform fails to act expeditiously after it becomes aware of the harmful nature of the content that it hosted. And this is known as limited liability regime. In The first case, state control over internet portals obliges the latter to pre-monitor or filter user generated comments. And as the ECtHR underlined in *MTE and Index.hu Zrt v. Hungary*, this burden will have negative implications since it ultimately creates "excessive and impracticable forethought capable of undermining the right to freedom of expression on the Internet"¹⁴¹ Furthermore, because of their lack of capacity to fulfil an obligation of preemptively monitoring all communications, intermediaries may decide to remove the comment section altogether to avoid liability for third party's illegal comment.

¹⁴¹ MTE and index V Hungary App no. 22947/2013, (ECtHR, 2016)

3.3 German

3.3.1 General Background

Due to philosophical and historical circumstances, the emphasis given to human dignity and the approach of regulating hate speech is quite unique in Germany.¹⁴² Article 1 of Germany's basic law and the jurisprudence of its Constitutional Court gives prominent position to the inviolability of human dignity. In the case of *Luth*¹⁴³, the Constitutional Court asserted the fundamental nature of right to free expression of opinion stating that "...it is one of the foremost human rights of all."¹⁴⁴ Whereas, in *Mephisto*¹⁴⁵ the Court held that "Human dignity is the supreme value that dominates the whole value system of the fundamental rights." The essential framework concerning regulation of hate speech has also been set out by the Constitutional Court in case of *Holocaust Denial*.¹⁴⁶

3.3.2 The Network Enforcement Act (NetzDG)

The unfolding of Germany's Network enforcement Act (NetzDG)¹⁴⁷ is highly attributed to the growing menace of hate speech and right-wing anti-migrant backlash following Angela Merkel's acceptance of over a million Syrian asylum seekers in 2015.¹⁴⁸ Since the government was of the opinion that Social Media companies did not adopt sufficient and effective self-regulatory

¹⁴² Claudia E. Haupt, Regulate Hate Speech – Damned If You Do and Damned If You Don't: Lessons Learned Comparing the German and U.S. Approaches, 23 B.U. INT'L L.J. 299, 309 (2005). see also Michel Rosenfeld, Hate Speech in Constitutional Jurisprudence: A Comparative Analysis, 24 CARDOZO L. REV. 1523, 1529 (2003); Deborah Levine, Sticks and Stones May Break My Bones, But Words May Also Hurt Me: A Comparison of United States and German Hate Speech Laws, 41 FORDHAM INT'L L.J. 1293, 1318 (2018).

¹⁴³ 1 BVerfGE 198 (1958).

¹⁴⁴

¹⁴⁵ 30 BVerfGE 173 (1971).

¹⁴⁶ 90 BVerfGE 241 (1994)

¹⁴⁷ Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz) Act to Improve Enforcement of the Law on Social Networks of 1 september 2017(BGBl. 13352)

¹⁴⁸ Amélie Heldt, *Leibniz Institut für Medienforschung, Hans-Bredow-Institut, Hamburg, Germany accessed at <https://policyreview.info/articles/news/germany-amending-its-online-speech-act-netzdg-not-only/1464>* April 2020

measures for fast removal of unlawful content, it has drafted the NetzDG.¹⁴⁹ The Act was passed in June 2017 and went to full effect on January 2018. Even through many countries throughout the world are grappling with the challenge of countering online proliferation of hate speech, Germany is unique among other western democracies in its ambitious attempt of holding social media gaits accountable for failure to implement effective and transparent complaints management infrastructure. Generally speaking, the restrictions imposed by the NetzDG can be regarded as necessary in a democratic society in the sense of Article 10 (2) of the ECHR. Protection of the right and reputation others who are exposed to hate speech as well as prevention of breaches of public order can be considered as legitimate aims of the law which ate in line with the ECHR.¹⁵⁰

3.3.3 Object and main Tenets of the NetzDG

The Act obliges social networks¹⁵¹ to provide effective, user-friendly and transparent complaint mechanisms¹⁵² which ensures the deletion or blockage of illegal contents published in their platform within a specified timeframe. Once they receive a complaint these social media companies should take immediate note and investigate whether the content is simply ‘unlawful’ or ‘manifestly unlawful’.¹⁵³ According to section 3(2) of the network enforcement Act, while manifestly illegal content must be blocked or taken down within 24 hours of receiving the

¹⁴⁹ Stefan Theil (2019) The Online Harms White Paper: comparing the UK and German approaches to regulation, *Journal of Media Law*, 11:1, 41-51, DOI: [10.1080/17577632.2019.1666476](https://doi.org/10.1080/17577632.2019.1666476)

¹⁵⁰ However, the Act is not narrow in its speech restriction.

¹⁵¹ Social networks are defined as “telemmedia service providers with more than 2 million users which, for profit-making purposes, operate internet platforms which are designed to enable users to share any content with other users or to make such content available to the public. Accordingly, Platforms which provide individualized communication services, such as email or messaging apps, as well as platforms providing editorialized content, such as news websites, are not within the NetzDG).”

¹⁵² Network Enforcement Law s 3(1).

¹⁵³ Network Enforcement Law s 3(2)(2). Thus, Social media platforms are not only obliged to engage in monitoring and reviewing the allegedly illegal content, but also to interpret certain provisions the German penal Code. Some critics described such situation as” Privatization of the law enforcement”

complaint, merely illegal content should be removed within a week subject to some exceptional circumstances.¹⁵⁴ “Both complainants and content generators must immediately be informed of the decision on the complaint, and reasons for the decision must be provided”.¹⁵⁵

3.3.4 Scope

Its scope of application is provided in under sec. 1 (3). Accordingly, NetzDG does targets 21 different criminal offences which are provided in the German Penal Code related to, *inter alia*, “insult, defamation, public incitement to crime, incitement to hatred and dissemination of propaganda material or use of symbols of unconstitutional organizations”. Even though the term hate speech is not explicitly indicated in the Act, the enumerated provisions encompass utterances which are described as hate speech. Here it is also important to point out that the NetzDG does not establish new criminal offences, rather it “merely enforces an existing legal obligations of social media service providers which is proscribed under Sec. 10 of the Telemedia Act.”¹⁵⁶

3.3.5 Transparency mechanisms and Sanctions

In addition to establishing effective complaints management mechanisms, the NetzDG imposes an organizational obligation of transparency or reporting duty on social media companies. The Act provides that the handling of complaints shall be monitored via monthly checks by the social network’s management. “If a platform receives more than 100 complaints per year, it must produce a semi-annual report detailing its content moderation practices and make it available for the public.”¹⁵⁷ Failure to implement such core obligations will entail an administrative fine ranging

¹⁵⁴ Network Enforcement Law s 3(2)(3).

¹⁵⁵ Network Enforcement Law s 3(2)(5).

¹⁵⁶ Thomas Wischmeyer, ‘What is Illegal Offline is Also Illegal Online’ – The German Network Enforcement Act unpublished manuscript (September 27, 2018)

¹⁵⁷ Michel Rosenfeld, Hate Speech in Constitutional Jurisprudence: A Comparative Analysis. (2001) Cardozo Law Review. 24.

from 500,00 to 50 million Euros. However, as we can understand from sec. 4(5) of the Act, “no fines can attach to decisions in individual cases. Instead, fines require a *systemic and persistent* failure in the complaints management mechanism which must be substantiated through content that has been ruled illegal by a court in a separate proceeding.”¹⁵⁸

3.3.6 Its implication on Freedom of expression

According to Article 5 (2) of the German Basic Law, freedom of expression and information can be restricted by general laws meeting the standards enshrined in the Basic Law and developed by the Constitutional Court. Nevertheless, the NetzDG has been criticized by Scholars and Activists for its potentially restrictive effect on free speech.¹⁵⁹ Following its enactment, the UN Special Rapporteur on freedom of opinion and expression raised various concerns and concluded that the Act is incompatible with article 19 of the ICCPR.¹⁶⁰ Given the disproportionality of its potential sanction; some argued that the Act could serve as an *incentive to ‘over-block’* legitimate online speeches.¹⁶¹ Because, social media platforms are more likely to delete or block an alleged hate speech without an in-depth assessment than the risk of being subject to a fifty-million-Euro fine. Such imposition will ultimately lead to over-policing of lawful content driven by risk aversion and creates chilling effects on freedom of expression. Decisions about the legality or illegality of user-generated content needs thorough and comprehensive evaluation which is accompanied by judicial oversight. In this regard, some scholar characterized NetzDG as an “*invitation to private*

¹⁵⁸ Catherine O’Regan, Hate Speech Online: an (Intractable) Contemporary Challenge?, *Current Legal Problems*, Volume 71, Issue 1, 2018, Pages 403–429, <https://doi.org/10.1093/clp/cuy012>

¹⁵⁹ Mark Scott and Janosch Delcker, ‘Free speech vs. censorship in Germany. New rules on online hate speech cause problems for internet giants’ (Politico, 4 January 2018)

¹⁶⁰ Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (OL DEU 1/2017, 1 June 2017). <<https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf>> (accessed april 2020)

¹⁶¹ Wischmeyer (No. 165)

censorship.”¹⁶² However, since the decision of social media companies to remove the content which have been flagged can be challenged in a court of law, such depiction seems unwarranted. The problem might be its failure to compel these intermediaries to prepare a mechanism which enable the ‘publisher’ of an alleged hate speech to challenge or respond on the compliant.

3.4 The United States of America

3.4.1 The Communications Decency Act

The United States stands alone, quite famously, when it comes to providing constitutional protection to hate speech except when it involves “incitement of imminent lawless action” (the standard developed in *Brandenburg v. Ohio*).¹⁶³ In addition to its pure speech Supreme Court jurisprudence, the country's minimalist approach or continued opposition of banning hateful contents has also fortified by its reservation to Article 20(2) of the ICCPR and Article 4 of the CERD. The US is also an outlier in giving safe harbor to giant social media companies.¹⁶⁴ Section 230 of the Communication Decency Act asserted that “no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”¹⁶⁵ Unlike the European system, The CDA does not obliged social media companies to comply with a notice-and-takedown procedure in order to benefit from the liability protection. Some scholars argued that the main purposes of broadly excluding the possibility of liability were not only to protect the free expression of platforms and their users but

¹⁶² Daily Brief” (Human Rights Watch, 15 February 2018), ‘New German Social Media Law Invites Censorship’, <https://www.hrw.org/the-day-in-human-rights/2018/02/15> (accessed April 2020)

¹⁶³ Peter J, Brickhejmer, The haven for hate; the foreign and domestic implication of protecting internet hate speech under the first amendment, southern California law review, vol 75,2002, page1493

¹⁶⁴ Klonick, Kate, The New Governors: The People, Rules, and Processes Governing Online Speech (March 20, 2017). 131 Harv. L. Rev. 1598.

¹⁶⁵ Communication Decency Act, 47 U.S.C. § 230(c) (1996).

also “to foster technological innovation.”¹⁶⁶ On the other hand, while private social media companies are broadly immune from liability for user-generated illegal content posted on their platform, the First Amendment does not bound these online private media companies. Hence, despite the general presumption in favor of free speech, “regulatory actions of private social media companies through their internal term of service and community guidelines would not create affirmative obligations under the First Amendment.”¹⁶⁷ Subsequently, even if the content is constitutionally protected, they are free of liability for removing such contents when they considered it as “obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable.”¹⁶⁸ The term “otherwise objectionable” in section 230 of the CDA is broad enough to include hate speech as a justification for their content restriction. However, recently the immunities of social media companies have been under scrutiny in an especially intense way because of their failure to adopt a clear, transparent and a workable policy that addresses hate speech. On the other hand, since the scheme of control which is supposed to be developed by social media platforms is “not subject to any oversight or monitoring there no guarantee that freedom of expression is properly protected by these platforms.”¹⁶⁹

¹⁶⁶ Samples, John, Why the Government Should Not Regulate Content Moderation of Social Media (April 9, 2019). Cato Institute Policy Analysis, No. 865. Pp

¹⁶⁷ Daithí Mac Síthigh (2020) The road to responsibilities: new attitudes towards Internet intermediaries, *Information & Communications Technology Law*, 29:1, 1-21,

¹⁶⁸ Communication Decency, U.S.C. section 230(c) 2

3.5 Ethiopia

3.5.1 General Background

Ethiopia is on the verge of political transformation. And the country's public protest-driven reforms have been partly facilitated on social media platforms which sometimes serve as an alternative political forum for citizens. However, these platforms have also been weaponized to inflame long-simmering grievances and encourage animosity among communities. Nowadays, the age-old unity and fraternity among the peoples of Ethiopia is being threatened by the prevalence of hate speech on social medias.¹⁷⁰ In 2018 alone, the raise of hate speech had led to “ethnic tension and conflict across the country that have created more than 1.4 million new Internally Displaced Persons”.¹⁷¹ Researchers argues that “hateful speech and disinformation have contributed significantly to the unfolding polarized political climate, religious-based attacks, ethnic violence and displacement in Ethiopia”¹⁷² Historically distinct socio-political dynamics of the country is being instrumentalized to promote hate and tribal intolerance on social medias.¹⁷³ And such situations are becoming a threat to human dignity, diversity and equality, national security and social stability.¹⁷⁴ That being the case, regulation of hate speech through legislative measure is warranted to the country's growing concern. Nevertheless, such statute-backed regulatory schemes may also present, somewhat counter-intuitively, the risk of stifling legitimate expressions and dissenting voices on social medias. There is no doubt that regulation of hate

¹⁷⁰ Addis Fortune, “Protect Free Speech, Tolerate Hate Speech.” Addis Fortune, April 13, 2019.

<<https://addisfortune.news/protect-free-speech-tolerate-hate-speech/>> (accessed may 2020)

¹⁷¹ Abraha, Halefom. ‘The Problems with Ethiopia's Proposed Hate Speech and Misinformation Law.’ Media Policy Project Blog (blog). (June 4, 2019) <<https://blogs.lse.ac.uk/medialse/2019/06/04/the-problems-with-ethiopias-proposed-hate-speech-and-misinformation-law/>> (accessed may 2020)

¹⁷² Ibid.

¹⁷³ Yared Legesse mengstu ‘*Shielding Marginalised Groups from Verbal Assaults without Abusing Hate Speech Laws*’ in Herz and Molnar eds. (2012), the content and context of Hate speech Cambridge, Cambridge University Press pp. 53.

¹⁷⁴ Gagliardone, Iginio, Alisha Patel, and Matti Pohjonen. 2014. Mapping and Analysing Hate Speech Online. Oxford/Helsinki/Addis Ababa: University of Oxford/Addis Ababa University.

speech for preserving public order is more appealing in politically polarized and less democratic countries like Ethiopia than in well-established democracies like Germany and the US. Nonetheless, the legal conundrum surrounding proper regulation, specifically, deciding who should be responsible for determining the boundaries of permissible content? what parameters should be employed? and how can this be achieved without eroding online freedom of speech? remains imperative even in developing countries like Ethiopia whose priority is ensuring peaceful coexistence of communities.

Content regulation rules had existed in plethora of Ethiopian laws even before the enactment of the new hate speech and misinformation suppression proclamation. The TFOP¹⁷⁵ can be considered as the first legislation which aimed at regulating certain type of illegal content on the internet. The ‘Value Added Service License Directive’¹⁷⁶, has also incorporated some provisions which prohibited online contents “that encourage hatred, violence, or discrimination.” The recent effort to adopt an effective countermeasure to the dissemination of problematic online content, however, was the enactment of Computer Crime Proclamation which provides that “sharing any written, video, audio or any other picture that incites violence, chaos or conflict among people shall be punishable with rigorous imprisonment not exceeding three years”¹⁷⁷ Nevertheless, this law failed to explicitly proscribe hate speech, racist and xenophobic content, justification or denial of genocide, and other harmful contents.¹⁷⁸

¹⁷⁵ Telecom Fraud Offence Proclamation, *Federal Negarit Gazeta*, Proclamation No. 761/2012, Article 6(1),

¹⁷⁶ Ministry of Communications and Information Technology, *Value Added Service Licence Directive* No. 3/2011, Article 11

¹⁷⁷ Computer Crime Proclamation, *Federal Negarit Gazeta*, Proclamation No. 958/2016.

¹⁷⁸ Kinfe Micheal Yilma, ‘Some Remarks on Ethiopia’s New Cybercrime Legislation’ (2016) 10(2) *Mizan LR* 448, 458

3.5.2 The New Hate Speech and Disinformation Prevention and Suppression

Proclamation

The new law, which took effect on 23 March 2020 introduced a direct liability regime by imposing fines up to 100,000 Ethiopian birr and imprisonment for up to three years¹⁷⁹ for anyone who “disseminates hate speech by means of broadcasting, print or social media using text, image, audio or video.”¹⁸⁰ Unlike the ECtHR and Germany online speech regulatory approach which focuses on the digital infrastructures, the new Ethiopian hate speech and misinformation law is mainly directed at those who produce the content. What led to the adoption of this proclamation was the evolving concern of various stake holders on the prevalence of hate speech on social medias and the need to prevent individuals from engaging in speech that incite violence under the guise of exercising the right to freedom of expression, and promoting “tolerance, civil discourse and dialogue, mutual respect and understanding”¹⁸¹

3.5.3 Its accordance with International standards of freedom of expression

The Act defines hate speech under article 2(2) as “a deliberate promotion of hatred, discrimination or attack against a person or discernible group of identity, based on ethnicity, religion, race, gender or disability.” Since this definition is nebulous and over-broad, some argued that it may open up loopholes for the arbitrary application of the law to target legitimate dissent. Unlike the US, Germany and other European countries, there is no well-established democratic culture, independent judiciary, robust rule of law and human right protection system in Ethiopia. Hence, the existing system of check and balance in the country does not provide any institutional guarantee

¹⁷⁹ Hate Speech and Disinformation Prevention and Suppression Proclamation No. 1185 /2020, (March 23rd, 2020) Federal Negarit Gazette No. 26

¹⁸⁰ Ibid. Article 3 and Article 7

¹⁸¹ Ibid the preamble

if the new proclamation is being misconstrued by the law enforcement organ. It is also important to note that the proclamation goes beyond the command of Article 20(2) and the limitations on restrictions required by Article 19(3) of the ICCPR. Equally notable is that the new law prioritizes criminal and punitive measures as an effective means of solving the problem. However, in order to avoid a chilling effect on freedom of expression, it is also essential not to lose sight of international human right standard which requires criminal sanction on hate speech to be a remedy of last resort and in line with the principle of proportionality. In that respect, the new proclamation could inadvertently restrict individuals' freedom of expression and silence dissenting voices.

3.5.4 Obligations imposed on Social media service providers

If a certain content has been flagged as hate speech, like the NetzDG, the new Ethiopian proclamation imposed statutory obligation upon social media companies to take down such content within 24 hours. However, the law does not impose any liability on these platforms for their failure to comply with this requirement. Furthermore, there is no clear procedural guidelines that governs the manner by which hateful content could be taken down. It appears that the government does not have the means and leverage to coerce giant social media companies and assure their compliance with the new proclamation. Facebook transparency report, for instance, indicates that the Ethiopian government has never made any requests for problematic content to be taken down from its platform.¹⁸² If we assume that the government can oblige platform providers to develop content moderation units in Ethiopia, to which language this applies? In a country of more than 70 languages, content moderation focusing only on posts in one or two languages could present another problem.

¹⁸²Bethel Ghebru, Mark Lloyd. 'From civic to citizenship education: Toward a stronger citizenship orientation in the Ethiopian CEE curriculum, (2020) *International Journal of Educational Development* 72, pages 102143.

3.5.5 Procedural shortcomings

The new hate speech and misinformation suppression proclamation does not device any mechanism which enable the creator of perceived hateful content to challenge the removal decision when he/she believes that the decision was illegitimate, disproportionate or unjustified. Most notably, it does not require the decision of taking down an alleged illegal content to be backed by judicial authorization. This ultimately creates a situation that allow private social media companies to determine whether a certain content is illegal or not without government oversight. In addition to putting individual's freedom of speech at the mercy of the "heckler's veto"¹⁸³, such delegation of censorship measure to private social media companies can have significant human right ramifications. "While the law has clarified that engaging in ordinary social media activities such as tagging or liking certain content does not entail criminal liability, it's not clear whether re-sharing or re-posting the content would be illegal."¹⁸⁴ It has also failed to provide standing for those who were targeted by hate speech.

¹⁸³ Yohannes Eneyew Ayalew 'Muting sectarianism or muzzling speech'? January 31, 2020

¹⁸⁴ Brhan taye, 'Ethiopia's hate speech and disinformation law: the pros, the cons, and a mystery' May 19 2020

4. Concluding remarks and recommendations

This project aimed to answer the question of how regulation of online communications affects the rights of freedom of expression and to what extent online service providers can be held responsible for hosting third parties' hateful content. To answer those questions, I started by providing some common grounds to comprehend the notion of this evolving phenomenon by taking international and regional human right instruments as a point of departure. After setting theoretical and philosophical foundation for the study, I analyzed the relationship between online hate speech and offline hate crimes and concluded that there is a direct cause and effect relationship between these two occurrences. In the first section of Chapter III, it was observed that apart from the UN Guiding Principles on Business and Human Rights, there is no binding international treaty and supranational jurisdictional authority which specifically aimed at imposing human right obligation on private social media companies. I have also explored how the regulation of Internet platforms in relation to hate speech is being approached by the Council of Europe, Germany and the United States and to what extent social networking platforms can be held liable for the dissemination hateful contents. The response of each jurisdiction is linked to its unique circumstances and informed by the shift of balance in ways of understanding the acceptable limits of freedom of expression.

In the last section of this project I confirmed the hypothesis that because of lack of well-established institutional accountability mechanism, legal frameworks for moderation of user-generated online content can have more detrimental effect for free speech in developing countries like Ethiopia than in developed liberal democracies. As an illustration, it has been shown how the new hate speech and misinformation proclamation of Ethiopia presented a potential risk of unduly

restricting individual's right to free speech and stifling critical public debate on social medias. Therefore, the government of Ethiopia should minimize such dreadful consequences by,

- Affording the institutional and procedural safeguards against misuse of the new hate speech and misinformation proclamation.
- Providing proper mechanism which enable the creator of perceived hateful content to challenge the removal decision when he/she believes that the decision was illegitimate, disproportionate or unjustified.
- Requiring social networking platforms to have judicial authorization before taking down an alleged illegal content
- Commanding judges and public prosecutors to consider the Rabat Plan of Action while implementing the new proclamation

It was argued that the potential impact of hate speech should not be underestimated by Ethiopian policy makers. However, while I understand the arguments regarding the importance of newly introduced proclamation for maintaining peaceful coexistence among communities by proscribing hate speech, in my opinion, the prevalence of this harmful phenomena is a mere manifestation of more profound historical and socio-political problems. In that sense, I argued that resorting to legislative measures alone is not sufficient to bring about real changes.¹⁸⁵ The enactment of the laws would mean little unless the government takes more comprehensive and mutually reinforcing strategies which includes improving media literacy, developing specific educational programmes, conducting evidence-based legal and sociological research which seek

¹⁸⁵ Yet again, because of different reasons highlighted in this project, it is equally important to note that I am rather skeptical about the use of counter-speech as the only strategy to tackle the problem of hate speech in the current Ethiopian context.

to identify the conditions that leads to hate speech, encouraging intercultural dialogue, promoting and exemplifying tolerance, reconciliation, and inter-community respect and *etc.*

Bibliography

Cases

- *Abrams v. United States*, 250 U.S. 616, 630 (1919)
- *Brandenburg v Ohio* [1969] 395 USSC 444
- *Delfi AS v. Estonia (GC)*, App no. 64569/09, (ECtHR 2015)
- *Erbakan v. Turkey*, App. No. 59405/00 (ECtHR, 2006)
- *Garaudy v. France*, App. No. 65831/01 (ECtHR, Inadmissibility decision 24 June 2003)
- *H., W., P. and K. v. Austria*, Appn. No. 12774/87, 62 (ECtHR Inadmissibility decision)
- *Handyside vs the United Kingdom*, Application No. 5493/72, (ECtHR, 7 December 1976)
- *MTE and index V. Hungary* App no. 22947/2013, (ECtHR, 2016)
- *Packingham v. North Carolina*, 137 S. Ct. 1730 (2017).
- *Pihl v. Sweden*, ECHR (judgment) (no. 74742/14)
- *Ross v. Canada*, Communication No. 736/1997
- *Sunday Times vs the United Kingdom*, Application No. 6538/74 (ECtHR, 26 April 1979,)
- *Sürek v. Turkey, ECtHR (GC)*, App no. 26682/95, (1999),
- *Vejdeland and others v. Sweden*, App. no. 1813/07 (ECtHR, sept 2012)
- *Jersild v Denmark*, app no. 15890/89, (ECHR 23 September 1994)
- *ICTR, The Prosecutor v. Jean-Paul Akayesu*, ICTR-96-4-T, September 1998
- *Gündüz v Turkey* App No 35071/97 (ECHR, 4 December 2003)
- *Balogh v Hungary* App no 47940/99 (ECHR, 20 July 2004)
- *Gelle v Denmark*, Communication no. 34/2004 (15 March 2006)
- *Ivanov vs Russia*, App. No. 35222/04 (ECtHR, 20 Feb. 2007)
- *Leroy vs France*, App. No. 36109/03, (ECtHR, 2 October 2008)
- *W.P. and Others vs Poland*, App. No. 42264/98 (ECtHR2 September 2004);
- *Norwood vs the United Kingdom*, App. No. 23131/03 (ECtHR, 16 Nov. 2004)

Legal and semi-legal documents:

- Communication Decency, U.S.C. section 230(c)
- Computer Crime Proclamation, *Federal Negarit Gazeta*, Proclamation No. 958/2016.
- Council of Europe, Committee of Ministers, *Appendix to Recommendation no. R (97) 20*. 1997 (Recommendation no. (97) 20)
- European Commission against Racism and Intolerance, *General Policy Recommendation No. 15 on Combating Hate Speech*. 2015 (CRI (2016) 15)
- Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz) Act to Improve Enforcement of the Law on Social Networks of 1 September 2017 (BGBl. 13352)
- Hate Speech and Disinformation Prevention and Suppression Proclamation No. 1185 /2020, (March 23rd, 2020) Federal Negarit Gazette No. 26
- International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR),
- Ministry of Communications and Information Technology, *Value Added Service Licence Directive* No. 3/2011,
- prohibition of incitement to national, racial or religious hatred (11 January 2013) A/HRC/22/17/Add.4
- Report of the United Nations High Commissioner for Human Rights on the expert workshops on the
- Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (OL DEU 1/2017, 1 June 2017).
- Telecom Fraud Offence Proclamation, *Federal Negarit Gazeta*, Proclamation No. 761/2012, Article 6(1),
- The annual report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression to the General Assembly, (2012).
- The Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence, A/HRC/22/17/Add.4, Appendix, adopted 5 October 2012.

- UN General Assembly, *International Convention on the Elimination of All Forms of Racial Discrimination*, 21 December 1965, United Nations, Treaty Series, vol. 660,
- UN General Assembly, *Convention on the Prevention and Punishment of the Crime of Genocide*, 9 December 1948, United Nations, Treaty Series, vol. 78
- United Nations, General Assembly, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2019 (A/74/48050).
- United Nations, Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2018 (A/HRC/RES/38/35).
- Universal Declaration of Human Rights (adopted 10 December 1948).

Books and Journal Articles

- Rawls J, *A Theory of Justice* (Harvard University Press 1971)
- Brown A, *Hate Speech Laws* (Routledge 2015)
- Heinze E, *Hate Speech and Democratic Citizenship* (Oxford University Press 2016)
- Herz M and Molnar P (eds), *The Content and Context of Hate Speech* (Cambridge University Press 2012)
- Waldron, J. (2012). *The harm in hate speech*. Cambridge, Mass.; London: *Harvard University Press*, 2012.
- McGonagle, T. (2013). *The Council of Europe against online hate speech: Conundrums and challenges. Council of Europe Conference Expert Paper*.
- Meiklejohn, A. (1965). *Political Freedom: The Constitutional Powers of the People*, New York: *Oxford University Press*
- Acob M, and Joelle F, *The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship*, (first ed 2019 Justitia publisher) 17

- Alexander B, 'What Is so Special about Online (as Compared to Offline) Hate Speech?' (June 2018): *Ethnicities* 18, no. 3, 297–326.
- Alexander B, A 'what is hate speech? the myth of hate speech' (2017) *Law and Philosophy* 36: pp 419–468
- Alexander, *Hate Speech Law: A Philosophical Examination*, Routledge Studies in Contemporary Philosophy (New York: Routledge, 2015).
- Alexander T, 'Free Speech Constitutionalism' (February 3, 2014) *University of Illinois Law Review*, Vol.
- Andrea D, *The Jews, Israel and women's liberation* (The free press 2000) 141
- Anita B, Todd Landman & Dorothea Farquhar 'Social media and protest mobilization: evidence from the Tunisian revolution', *Democratization*, 22:4, 764-792
- Banks J, 'European Regulation of Cross-Border Hate Speech in Cyberspace: The Limits of Legislation' (2011) *European Journal of Crime, Criminal Law and Criminal Justice*, 19(1), 1,13.
- Perry B, and Patrik Ol, 'Cyberhate: The Globalization of Hate'(2009) 18(2) *Information & Communications Technology Law*, 185-199 see also James Banks, 'Regulating Hate Speech Online' (2010) *International Review of Law, Computers & Technology* 24(3), 233-239.
- Bethel G, Mark Lloyd. 'From civic to citizenship education: Toward a stronger citizenship orientation in the Ethiopian CEE curriculum, (2020) *International Journal of Educational Development* 72, pages 102143.
- Brhan T, 'Ethiopia's hate speech and disinformation law: the pros, the cons, and a mystery' May 19 2020
- Brünker, Judith and others, 'The Role of Social Media during Social Movements – Observations from the #metoo Debate on Twitter.' In *Proceedings of the 52nd Annual Hawaii International Conference on System Sciences (HICSS)*, 2356- 2357
- Cascante, "Black Lives Matter: Understanding Social Media and the Changing Landscape of Social Trust" (2019). *Theses and Dissertations*. 3375.
- O'Regan o, Hate Speech Online: an (Intractable) Contemporary Challenge?, *Current Legal Problems*, Volume 71, Issue 1, 2018, Pages 403–429,

- Claudia E. Haupt, Regulate Hate Speech – Damned If You Do and Damned If You Don’t: Lessons Learned Comparing the German and U.S. Approaches, 23 B.U. INT’L L.J. 299, 309 (2005).
- Corey Brettschneider, *When the State Speaks, What Should It Say? How Democracies Can Protect Expression and Promote Equality*, NJ: (Princeton Univ. Press 2012).
- Daily Brief’ (Human Rights Watch, 15 February 2018), ‘New German Social Media Law Invites Censorship’,
- Daithí Mac Síthigh (2020) The road to responsibilities: new attitudes towards Internet intermediaries, *Information & Communications Technology Law*, 29:1, 1-21,
- Danielle K. Citron Helen Norton, ‘Intermediaries and Hate Speech: Fostering Digital Citizenship for our Information Age’, *Boston University Law Review*, Vol. 91, p. 1435, 2011.
- David Kretzmer ‘Freedom of Speech and Racism’(1987) 8 *Cardozo Law Review* 445-513, 478.
- Deborah Levine Sticks and Stones May Break My Bones, But Words May Also Hurt Me: A Comparison of United States and German Hate Speech Laws, 41 *FORDHAM INT’L L.J.* 1293, 1318 (2018).
- Displaced Persons: A Critical Appraisal’ (2014) *Journal of Politics and Law* Vol. 4, No. 2; 1
- Elissa A, ‘Yahoo!, Inc. v. LICRA: The French Challenge to Free Expression on the Internet’ (2002) *American University International Law Review* 18, no. 1, 295-339
- Gagliardone, Iginio, Alisha Patel, and Matti Pohjonen. 2014. Mapping and Analysing Hate Speech
- Gagliardone, Ignio and others, ‘Countering Online Hate Speech’ (UNESCO, 2015)
- George Rutherglen, ‘Theories of free speech’ (1987) *Oxford Journal of Legal Studies* 7, no. 1 115–24 in Pejchal, V. (2020). *Hate Speech and Human Rights in Eastern Europe*. London: Routledge,
- Hylton Keith ‘Implications of Mill's Theory of Liberty for the Regulation of Hate Speech and Hate Crimes’ (1996) *The University of Chicago Law School Roundtable*: Vol. 3: Iss. 1. 35, 57

- Jeffrey W. Howard, 'Free Speech and Hate Speech' (2019) Annual. Rev. Political Sci. 22:93–109
- Karsten Müller & Carlo Schwarz, From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment 25-28 (Nov. 2, 2019);
- Karsten Müller, and Carlo Schwarz, Fanning the Flames of Hate, 'Social Media and Hate Crime' (November 3, 2019).
- Kinfé Micheal Yilma, 'Some Remarks on Ethiopia's New Cybercrime Legislation' (2016) 10(2) Mizan LR 448, 458
- Klonick, Kate, The New Governors: The People, Rules, and Processes Governing Online Speech (March 20, 2017). 131 Harv. L. Rev. 1598.
- Mari Matsuda, 'Public Response to Racist Speech: Consider the Victim's Story', 87 MICH. L. REV. 2320, 2341
- Mark Scott and Janosch Delcker, 'Free speech vs. censorship in Germany. New rules on online hate speech cause problems for internet giants' (Politico, 4 January 2018)
- McGonagle, Tarlach 'The Council of Europe against online hate speech: Conundrums and challenges' (2012)
Belgrade: Republic of Serbia, Ministry of Culture and Information
- Michel Rosenfeld, Hate Speech in Constitutional Jurisprudence: A Comparative Analysis, 24 CARDOZO L. REV. 1523, 1529 (2003);
- Michel Rosenfeld, Hate Speech in Constitutional Jurisprudence: A Comparative Analysis. (2001) Cardozo Law Review. 24.
- Nadine Strossen, *Hate: Why We Should Resist It with Free Speech, Not Censorship*, Inalienable Rights (New York: Oxford University Press, 2018).
- Nash, V *Analyzing Freedom of Expression Online: Theoretical, Empirical, and Normative Contributions*. (Oxford Handbook of Internet Studies 2013).
- Natalie Alkiviadou, 'Regulating Internet Hate: A Flying Pig?' (2016) JIPITEC 216
- Oltmann, Shannon M., 'Intellectual Freedom and Freedom of Speech: Three Theoretical Perspectives' (2016). Information Science Faculty Publications. 30. 158
- Parmar, S. 'Uprooting 'defamation of religions' and planting a new approach to freedom of expression at the United Nations. In T. McGonagle & Y. Donders (Eds.), *The United*

Nations and Freedom of Expression and Information: Critical Perspectives (Cambridge University Press 2015) 373, 427

- Pejchal Viera, *Hate Speech and Human Rights in Eastern Europe* (London: Routledge 2020). 181
- Peter J, The haven for hate; the foreign and domestic implication of protecting internet hate speech under the first amendment, southern California law review, vol 75,2002, page1493
- Richard A. Wilson & Jordan Kiper, Incitement in an Era of Populism: Updating Brandenburg After Charlottesville (2020) 5 U. PA. J. L. & PUB. AFFAIRS Alexandra Stevenson, 'Facebook Admits It Was Used to Incite Violence in Myanmar', (The New York times Nov. 6, 2018)
- Rotem Medzini and Tehilla Shwartz Altshuler, *dealing with hate speech on social media*, (the Israel Democracy Institute first edn 2019) 5
- Sajó, A *Freedom of Expression* (Institute of Public Affairs, Warsaw 2004) p. 128
- Samples, John, Why the Government Should Not Regulate Content Moderation of Social Media (April 9, 2019). Cato Institute Policy Analysis, No. 865. Pp
- Sellars, Andrew 'Defining Hate Speech' Berkman Klein Center Research Publication No. 2016-20; Boston Univ. School of Law, Public Law Research Paper No. 16-48.
- SELMA 'Hacking Online Hate': Building an Evidence Base for Educators
- Shaw, L 'Hate Speech in Cyberspace: Bitterness without Boundaries' (2011) Notre Dame Journal of Law, Ethics and Public Policy, 25(1), 279-304.
- Stefan T, (2019) The Online Harms White Paper: comparing the UK and German approaches to regulation, Journal of Media Law, 11:1, 41-51,
- Taylor, E. (2016). The Privatization of Human Rights: Illusions of Consent, Automation and Neutrality. Global Commission on Internet Governance Paper Series, p. 3.
- Thomas J. Webb, Note, Verbal Poison - Criminalizing Hate Speech: A Comparative Analysis and a Proposal for the American System, (2011) 50 WASHBURN L.J. 445, 4484

- Thomas Wischmeyer, ‘What is Illegal Offline is Also Illegal Online’ – The German Network Enforcement Act unpublished manuscript (September 27, 2018)
- Veliz, c ‘Online Masquerade: Redesigning the Internet for Free Speech Through the Use of Pseudonyms’ (2019) Journal of Applied Philosophy, Vol. 36, No. 4, 644
- Wibke Kristin Timmermann, ‘The Relationship between Hate Propaganda and Incitement to Genocide: A New Trend in International Law Towards Criminalization of Hate Propaganda’ (2005) Leiden Journal of International Law, 18 pp. 257–282
- Winfried Brugger, The Treatment of Hate Speech in German Constitutional Law (Part I), 3 German Law Journal (2002)
- Yared Legesse mengstu ‘*Shielding Marginalised Groups from Verbal Assaults without Abusing Hate Speech Laws*’ in Herz and Molnar eds. (2012), the content and context of Hate speech Cambridge, Cambridge University Press pp. 53.
- Yohannes Eneyew Ayalew ‘Muting sectarianism or muzzling speech’? January 31, 2020
- Yulia A Timofeeva, ‘Hate Speech Online: Restricted or Protected? Comparison of Regulations in the United States and Germany’ 12 Journal of Transnational Law and Policy 2, 255

Online resources:

- Addis Fortune, “Protect Free Speech, Tolerate Hate Speech.” Addis Fortune, April 13, 2019. <<https://addisfortune.news/protect-free-speech-tolerate-hate-speech/>> (accessed may 2020)
- Abraha, Halefom. ‘The Problems with Ethiopia’s Proposed Hate Speech and Misinformation
- Law.’ Media Policy Project Blog (blog). (June 4, 2019)
<<https://blogs.lse.ac.uk/medialse/2019/06/04/the-problems-with-ethiopias-proposed-hate-speech-and-misinformation-law/>> (accessed may 2020)