JOURNAL OF JUSTICE

The Challenges of Regulating Hate Speech on Social Media in light of the Theory of Freedom of Expression

Dr. Seydi Çelik* & Biruk Paulos**

Abstract

Although international human rights instruments declare that everyone has the right to freedom of thought and expression, many countries today have laws that censor or limit certain forms of expression on social media, including speech that fuels hatred and violence. Some free speech advocates prefer an open market of ideas where no expression is restricted thinking that the best answer to harmful speech is discussions that allow different ideas to challenge any speech freely. Still some others, on the other hand, argue that restrictions on hate speech on social media are vital for protecting minority communities from the harms caused by this speech. Hate speech raises a very complex question that tests the limits of free speech. Thus, based on the theories that underpin freedom of expression, this piece of work strives to analyze the regulatory dilemma on hate speech on social media by exploring the arguments and challenges that frame the regulatory process.

Keywords: Freedom Of Expression, Hate Speech, Social Media

* Dr. Öğr. Üyesi, *Kocaeli Üniversitesi Hukuk Fakültesi,* seydi.celik@kocaeli.edu.tr, Orcid: 0000-0003-1224-9877

** Biruk Paulos (LL.B, LL.M) is a lecturer of International Law at Dilla University School of Law, brookcanada2017@gmail.com. ORCID ID: https://orcid.org/0000-0001-6059-972X

İfade Özgürlüğü Teorisi Işığında Sosyal Medyada Nefret Söylemini Düzenlemenin Zorlukları

Öz

Uluslararası insan hakları belgeleri herkesin düşünce ve ifade özgürlüğü hakkına sahip olduğunu ilan etmesine rağmen, bugün birçok ülkede nefret ve şiddeti körükleyen konuşma da dahil olmak üzere sosyal medyada belirli ifade biçimlerini sansürleyen veya sınırlayan yasalar bulunmaktadır. Bazı serbest konuşma savunucuları, zararlı konuşmaya en iyi yanıtın, farklı fikirlerin herhangi bir konuşmaya özgürce meydan okumasına izin veren tartışmalar olduğu düşüncesiyle hiçbir ifadenin kısıtlanmadığı açık bir fikir pazarını tercih eder. Yine de bazıları, sosyal medyada nefret söylemi üzerindeki kısıtlamaların, azınlık topluluklarını bu konuşmanın neden olduğu zararlardan korumak için çok önemli olduğunu savunuyor. Nefret söylemi, özgür konuşmanın sınırlarını test eden çok karmaşık bir soruyu gündeme getiriyor. Bu nedenle, ifade özgürlüğünü destekleyen teorilere dayanarak, bu makale, sosyal medyada nefret söylemi üzerindeki düzenleyici ikilemi, düzenleyici süreci çerçeveleyen argümanları ve zorlukları inceleyerek analiz etmeye çalışmaktadır.

Anahtar Kelimeler: İfade Özgürlüğü, Nefret Söylemi, Sosyal Medya

1. Introduction

Hate speech in social media is a legal problem that has become a controversial issue at both international and national levels (Walker, 2018). At one extreme, some argue that hate speech is freedom of expression that requires legal protection, while others argue that hate speech should not fall within the scope of freedom of expression. Striking the balance between hate speeches on the one hand and freedom of expression on the other has been a challenge for governments trying to regulate hate speech (Apalara, 2017). Countries around the world are faced with challenges in maintaining the balance between the freedom of expression necessary for the realization of a democratic society and preventing harm to individuals or minority groups as a result of hate speech (Silvana Neshkovska, 2018). In many parts of the world, widespread atrocities and other human rights violations continue to threaten, especially those belonging to minority groups and indigenous peoples (CeaseFire Centre for Civilian Rights, 2019). The growing number of attacks against immigrants and other minorities have raised new concerns about the link between hate speech on social media and violent acts (Pejchal, 2015). The increasing trends in hate speech and hate crimes around the world reflect changes in the political climate and increasing social media violence (George, 2020). While hate speech is increasing on social media in many countries, the risks of violence, political repression and mass murder targeting certain

segments of society have emerged or the risk of such things happening is getting high (European Parliament, 2018). While social media has greatly benefited the global community in many ways, it has also been used as a platform to spread hate speech. In Myanmar, dehumanizing language and overt incitement against mass murder have been reinforced through Facebook and Twitter, contributing mainly to the widespread targeting of the Muslim Rohingya minority (Irving, 2018). Following the comprehensive political reform in Ethiopia since 2018, the spread of hate speech and misinformation on social media targeting some ethnic and religious minorities has resulted in widespread violent conflict, mass murder and internal displacement of millions of people (Maas, 2020). In the Central African Republic (CAR), online hate speech and provocation have played a role in fueling cycles of brutality between Christians and Muslims in recent years (OHCHR, 2019). In Turkey, hate speech against religious and ethnic minorities has pervaded the mainstream media threatening social peace in the country (Arcan, 2013; and Dural, 2012). Şükrü and Abdulkadir (2012) also have shown how hate speeches, being dominant news discourse in Turkish news Medias, have reinforced anti-Semitism, homophobia and racism in Turkey (GÖLCÜ, 2012). Hate crimes against dissident groups are on the rise in Turkey, and political hate speech might have a triggering effect on this trend (Barbara Perry et al., 2020). In India, where social media has portrayed an aggressive and privileged Hindu nationalism, rumors spread over WhatsApp and hate speech has

led to a series of gang attacks against Muslims (Paula Thompson, 2019; Samarth Mishra, 2019). In Sri Lanka, similarly, rumors on social media sparked anti-Muslim gang attacks in Sri Lanka last year, which were rekindled following April 2019 bombings of the Easter Sunday church and hotel by ISIS (AFP, 2020). In Somalia, Al-Shabaab has used Twitter and Facebook as a tool to broadcast hate speech propaganda and recruit fans in an environment where mobile phone use has steadily increased in recent years (HRW, 2018). In Sudan, social media, which is banned but still accessible via virtual private networks (VPNs), has played an important mobilizing and organizing role by gathering protesters around an iconic video of a young female activist who went viral around the world (US Department of State, 2019). In the United States, the perpetrators of recent white supremacist attacks have roamed online racist communities and also embraced social media to publicize their actions (ALA, 2017). The Charleston church killer, who killed nine black clergy and worshipers in June 2015, embarked on a process of self-learning that led him to believe that the goal of white supremacy required an act of violence (Laub, 2019). In Germany, hate speech on social media has caused serious violence against some minority groups, especially immigrants and refugees (Schwarz, 2019). Social media was involved in the far-right terrorist attack in Christchurch, New Zealand in March 2019 (Matthew L. Williams P. B., 2019). In the past, the spread of hateful, threatening and biased messages was traditionally done through radio, print media, SMS messages, and

even lyrics (Tsesis, 2002). Today, social media such as Facebook, Twitter, YouTube, Instagram and various internet news portals are used as platforms to spread hate speech in a whole new way (Silvana Neshkovska, 2018). Violence related to hate speech in social media has revealed many legal problems such as what constitutes hate speech, who should regulate social media and to what extent hate speech should be regulated. While some countries criminalize all kinds of hate speech and have laws regulating social media, some countries authorize social media companies to adopt their own regulatory policies to control hate speech on their platforms (Solmone, 2018). However, the issue of what constitutes hate speech and to what extent it should be limited remains a befuddling legal issue in many countries. While excessive regulation of hate speech on social media endangers freedom of expression and silences political dissidence, lenient regulations will also result in violence and crimes (Radu Meza, 2019; and Tsesis, 2002). In situations where hate speech on social media remain unregulated, the speeches translate in to violent acts on the ground creating insecurity for some group of people. On the contrary, in those systems where the government strictly regulates online hate speech, such regulations are vehemently rejected by some people who contend that such laws are meant to silence political dissidences and to repress some ethnic or religious minorities. Thus, this article attempts to explore the challenges of regulating hate speech on social media in light of the theory of freedom of expression.

2. Theoretical Analysis on 'Freedom of Expression' and 'Hate Speech'

2.1 The Theory of Freedom of Expression

2.1.1 The Truth Theory

According to John Stuart Mill, truth can be discovered through public debate (Appiagyei-Atua, 2005). Therefore, society should not place any limits on the exercise of free public debate so as not to hinder the attainment of the truth. In other words, government alone is not the key to all the facts in setting the right policies for its citizens. Such correct policies can only come from citizens through tight public debate. In line with this view, the spread of false views should also be tolerated. It is important to remember that, as Daniel Jacobson (2000) points out, Mill does not intend to restrict freedom of expression simply because someone is harmed by the statements of others. Mill distinguishes between legitimate and illegitimate damages and he says hate speech can only be limited or regulated when the conversation causes a direct and explicit infringement (Hylton, 1996). Therefore, in Mill's opinion, speech cannot be limited only to prevent harm. The harm justifies restricting speech only if it involves a direct and obvious violation of rights. Conversely, the conversation should not be restricted or regulated if it does not cause a direct and apparent violation of the rights of others. Hence, in the light of this theory, hate speech can be considered as freedom of expression as long as it does not violate the rights of others.

2.1.2 The Marketplace of Ideas

The marketplace of ideas is a justification for freedom of expression based on an analogy to the economic concept of the free market (Schultz, 2017). This logic assumes that freedom of expression is important because in a marketplace of ideas, better ideas will eventually prevail through competition. To justify a legal right to freedom of expression, referring to the public sphere as a "marketplace of ideas" is a familiar metaphor (Lee, 2010). The idea market increases social value by providing easier access to the truth or reality. However, arguments based on this logic are stronger to protect statements that are capable of being true or false than statements that are not (Lombardi, 2019). This is important because many kinds of arguments are not linked to any truth-value like art, and even less linked to obscenity and pornography (Macedo, n.d). Therefore, this theory explains that hate speech must pass the test of social acceptability or rejection in the market of ideas. In addition, some proponents of this theory support the idea of regulating hate speech by likening such regulation to regulation of market imperfections (Lee, 2010).

2.1.3 The Democracy Theory

Alexander Meiklejohn was the first political theorist to find a defense of freedom of expression on purely democratic theory (JMC, 2020). According to him, freedom of expression and freedom of the press are the most inclusive and therefore the best public debate. For this reason, Meiklejohn places freedom of

expression above almost all other rights as the basis or the most fundamental condition of the democratic process. Accordingly, a good justification for freedom of expression comes from self-government and participation in the democratic process. According to this logic, political speech is at the top of the hierarchy of freedom of expression. However, the logic of political self-government extends political protection to only some expressions while others (such as pornography) fall outside of these parameters. The self-government logic clearly states that freedom of expression is justified by democratic theories and justification. One argument against restrictions on hate speech is that they undermine the democratic legitimacy of the State (Weinstein J. , 2017). A contradictory view argues that hate speech itself harms the democratic process and therefore can rightly be limited to the defense of democracy (Reid, 2019). Where equal political status is a necessary condition for legitimate policy making, hate speech can undermine the legitimacy of this process. Therefore, for the advocates of the theory of freedom of self-government (democracy), if hate speech is compatible with logic, rationality and democratic values, it can be counted as freedom of expression. According to this theory, freedom of expression holds the government responsible and sensitive. Therefore, the theory of democracy suggests measures other than legal restrictions to prevent the negative consequences of hate speech in a democratic society (Weinstein J. , 2017).

Rather than regulating hate speech, it is better to compensate for the damage caused by hate speech by other means.

2.1.4 Autonomy Theory

According to this theory, freedom of expression is essential for self-realization and development (Oltmann, 2016). The main idea is that the government should not place any restrictions on the assumption that speech only leads to the formation of harmful beliefs or triggers harmful actions as a result of those beliefs (Scanlon, 1972). This theory advocates freedom of expression based on the assumption that freedom of expression has an intrinsic value. The free expression of one's beliefs is part of being autonomous. In this view, it is important for individuals to be autonomous regardless of whether this leads to good results, i.e. whether individual autonomy has an external value or not. Being able to express oneself freely is part of what it means to be autonomous. This theory argues that freedom of expression is intrinsically necessary as autonomy is intrinsically necessary. There are apparently two types of harmful speeches that should not be prohibited by law. As for that, speech cannot be banned simply because it has one of the following harmful consequences: First, if it causes the audience to have false beliefs; or secondly, if it causes harmful consequences by causing agents to believe (or increase their tendency to believe) that they should perform harmful actions (Scanlon, 1972). The importance of freedom of

expression for the pursuit of autonomy is twofold: self-reflection and talent (Peonidis, 1998). First, if freedom of expression is restricted, people are deprived of valuable information about the decisions they make (or can make) and the tools to review their beliefs and desires. Second, since many actions are carried out through many verbal, written or symbolic communication, people are prevented from exercising their autonomy when restrictions are placed on freedom of expression. The government has a responsibility to create conditions in which individuals can freely make their own choices about their lives (the good life) and realize themselves (Macedo, 2017). The theory of autonomy explains hate speech as freedom of expression, unless such speech or expression contradicts the principle of harm, as John Stuart Mill and Scanlon argued. The self-actualization argument for free expression thus provides a strong foundation for the preservation of all types of speech, including political speech, literature and art, and even radical, non-intellectual forms of expression such as nude dance, pornography or vulgarity, and commercial speech (Weiner, 2012). While the autonomy theory claims that freedom of expression will have full legal protection in order to fully protect the autonomy of the individual, some researchers argue that hate speech is harmful and must be regulated. The logic of autonomy will not be convincing if the harm caused by hate speech outweighs the ban on hate speech, which significantly contributes to the occurrence of serious violence (Weinstein I. H., 2009).

2.2 The definition of 'Hate Speech'

In recent years, especially with regard to freedom of expression on social media, many scholars have tried to define hate speech (Gelashvili, 2018). Hate speech has various nomenclatures: 'incitement to hatred' (OSCE-ODIHR, 2010, p. 2); 'antagonistic speech' (Radu Meza, 2019); 'cyber hate' (Roiha, 2016, p. 5); and 'inflammatory rhetoric' (Silvana Neshkovska, 2018, p. 2). Despite the variations of nomenclatures, the term 'hate speech' will be used for the purpose of this article. One reason why hate speech is difficult to define is that several types of related expression exist at the intersection. For example, it is very difficult to draw a line between hate speech and discriminatory language that humiliates people on the basis of a shared identity; cliché generalization based on a common identification of people, whatever the intent; dangerous speech that provokes violence (hate crime); symbolic expression, e.g. Swastikas, redesigned emojis and memes (Donovan, 2019). Therefore, there is no consensus on the definition of "hate speech". Searching for the perfect definition of the term "hate speech" is also beyond the scope of this article. However, it is imperative to provide a working definition of hate speech in this article, at least to make a clear analysis. Hate speech is typically defined as the act that disparages a person or people on the basis of a number of characteristics that may include and not limited to: race, ethnicity, sexual orientation, gender, religion, sexual orientation, caste,

gender identity, serious disease or disability and nationality (Al-Dossari, 2019; Radu Meza, 2019; Quinn, 2018; and Leandro Silva, 2016). The act of hate speech can take place either online or offline through any media. Online hate speech is defined as any technology-mediated speeches or digital communications that harass, discriminate, or humiliate a person (s) on the basis of a group-defining characteristic such as race, ethnicity, gender, nationality, sexual orientation, religion, age, disability and others (Melhuish, 2018). Online hate speech can also be aimed at spreading anti-Semitic, racist, bigoted, extremist or terrorist messages or information (Audrin, 2019). Hate speech on social media is created through harsh discourses which are expressed in the form of racism, xenophobia, discrimination, sexism, and homophobia (Yanik, 2017). Hate speech includes speech that is degrading, insulting, discriminatory, provocative, or even incites or encourages violence or causes violent reactions (Yadav, 2018). Therefore, for the purpose of this article, hate speech on social media is defined as "any internet-mediated expression either through words, symbols, codes, abbreviations, pictures, and any other digital graphics that discriminates, dehumanizes or incites violent attacks against group-defining characteristics such as race, ethnicity, gender, nationality, religion, age, disability, and other similar traits".

2.3 Current Academic Debates on 'Hate Speech'

There are multifaceted arguments concerning the regulation of hate speech on social media. The arguments can be categorized in to three groups: liberalist (postmodernist) arguments, progressivist arguments and multiculturalist debates. Actually, these arguments are not mutually exclusive. The center of these arguments focuses on striking the balance between protecting freedoms of expression and regulating hate speech (Jacobson, 2016). Due to censorship, the liberalists argue that freedom of expression can never be an absolute right (Alkiviadou, 2018). In their view, although an absolute protection of freedom of expression cannot be attained in all circumstances, liberal and non-liberal ideas should be tolerated. They argue for the adoption of free and neutral doctrine for the promotion of the right to freedom of expression even extending the legal protection for non-liberal ideas and expressions (Romero, 2016). Thus, in line with this argument, sometimes it is good to outlaw and punish some expressions for some good reasons within the limits of the law (Yaraghi, 2018). If freedom of expression means the freedom to express any opinion or idea, speech acts such as incitement to violence and other derogatory words should fall out of the ambit of freedom of expression (Juanatey, 2016). Accordingly, scholars are divided on the issue of regulating hate speech. While some scholars argue that the regulation of hate speech can severely restrict individual freedom of expression (Samples, 2019), others

see such restrictions as necessary to counter the harm, discrimination and subjugation that may result from hate speech (Altman, 1993). On the other spectrum, the progressivists argue that equality or social justice should prevail over freedom of expression, equality should be attained even at the cost of freedom of expression (Demaske, 2019). Accordingly, collective rights should be prioritized over individual liberty and the government should take measures to control hate speech to protect the collective rights of the society (Kaniklidis, 2015). However, most progressivists reject violence, deception or totalitarianism even though such techniques could be used to achieve social justice (Seidman, 2018). On the other hand, multiculturalists argue that any view or expression that provokes violence against minorities or marginalized groups of people must not be legally protected as freedom of expression (Chong, 2006). They justify the regulation of hate speech from the perspective of protecting minorities. They concur with the postmodernists that censorship is an important measure to avoid social harm. Thus, according to this view, only minorities or marginalized people are considered to be vulnerable to hate speech (Benesch, 2014; and Bell, 2009). The literature on hate speech is highly pervaded with arguments for social justice. Accordingly, the vast majority of people are very skeptical about freedom of expression and are intolerant of differences and disagreements (Traum, 2014). A brief review of the hate speech literature and its link to freedom of expression reveals a wide

variety of opposing views. At one extreme, some scholars consider hate speech as freedom of expression (Quent, 2019), while others at the other end consider hate speech as a highly harmful phenomenon that should be avoided or punished by law (Scheffler, 2015). At the mid-point of this spectrum, some scholars contend that only those hate speeches that intend to induce harm to those protected characteristics (race, gender, religion, ethnic origin, sexual orientation, etc.) that should be restricted and banned (Silvana Neshkovska, 2018). According to this spectrum, freedom of expression is distinguished from speeches or expressions that incite violence or harmful effects (Holmes, 2018). Freedom of expression does not include the right to express anything that harms others (Romero, 2016). Hate speech of all kinds, even though potentially harmful, do not have immediate harmful effects. Therefore, when the actions of individuals cause damage to or affect the rights or interests of others, their absolute freedom should be restricted (Wang, 2018).

3. Legal Regimes on "Freedom of Expression" and "Hate Speech"

3.1 International Human Rights Instruments

While the concept of freedom of expression is found in major international human rights documents[1], it is expressly provided as

[1] The concept of freedom of expression is expressed in major international human rights documents including Universal Declaration of Human Rights;

fundamental right in the Universal Declaration of Human Rights and International Convention on Civil and Political Rights.

UDHR emphasizes freedom of expression in its preface[2]. In article 19, the document provides for the right to freedom of expression. It provides freedom of expression as a right that constitutes the freedom to keep ideas free from infringement and to seek, receive and share information. Also the ICCPR, provides for freedom of expression in a similar way to UDHR. As per ICCPR, freedom of expression includes the freedom to seek, receive and disseminate ideas or information through any manner and any media[3]. According to this provision, freedom of expression can be used by any media, both offline and online. Although freedom of expression is clearly provided as the most

International Covenant on Civil and Political Rights (ICCPR); International Covenant on Economic, Social and Cultural Rights (ICESCR); International Convention on the Elimination of All Forms of Racial Discrimination (ICERD); Convention on the Elimination of All Forms of Discrimination against Women; ILO Convention No. 135, Workers' Representatives Convention; General Comment 10 (Article 19) of the Human Rights Committee (CCPR/C/21/Rev.1 of 19 May 1989); General Comment 11 (Article 20) of the Human Rights Committee (CCPR/C/21/Rev.1 of 19 May 1989); and the public's right to know: Principles on Freedom of Information Legislation. Annex II Report E/CN.4/2000/63.

[2] It reads "whereas disregard and contempt for human rights have resulted in barbarous acts which have outraged the conscience of mankind, and the advent of a world in which human beings shall enjoy freedom of speech and belief and freedom from fear and want has been proclaimed as the highest aspiration of the common people".

[3] The provision reads "freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of a person's choice".

important part of human rights and fundamental freedoms, some restrictions have been placed on it. Article 19 (3) (a) and (b) of the ICCPR imposes certain restrictions on the right to freedom of expression. Accordingly, freedom of expression can be restricted by law if deemed necessary. Reasons for such restrictions include respect for the rights or reputations of others; protection of national security or public order or public health or morals. The ICCPR appears to provide very limited scope for legal restrictions on freedom of expression. The question here is whether hate speech could be one of such grounds for restricting freedom of expression or not. At the international level, only the ICCPR includes rules on hate speech and incitement on the subject in question. Article 20 (2) of the Convention stipulates that any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law. This article clearly states that the advocacy of hate against national, racial or religious identities with the aim of inciting discrimination, hostility or violence will be the basis for prohibiting such acts by law. Thus, States are at liberty or responsibility to adopt domestic laws that prohibit hate speech in line with international standards. Under the ICCPR, if any speech constitutes discrimination, hostility or violence against national, racial or religious groups, it is considered hate (hate speech). In fact, this definition of hate speech seems to exclude speech targeting other types of identity (traits) such as age, gender, legal status (e.g. refugees) and physical disability. Despite gaps in the

definition of what constitutes hate speech, the ICCPR appears to have provided a legal basis for States to regulate hate speech by law. However, there is still a debate about who should regulate hate speech on social media; Government, social media companies or any other international organization? This question has not been properly addressed yet (Gelashvili, 2018). If hate speech is considered an exception to freedom of expression, it is self-explanatory, with which the government can intervene to legally regulate such malicious conversations in both online and offline media. The obligation to regulate hate speech is directed at States that require them to prohibit such behavior under domestic law, rather than social media companies that are not directly bound by human rights treaties as local institutional actors (Brannon, 2019). However, while companies may not have international legal standing, efforts have been made to subject companies to human rights law standards in recent years (Irving, 2018). For instance, the UN Guiding Principles on Business and Human Rights (UN GPBHR)[4] requires social media companies to respect human rights. Some of the responsibilities listed in the document include the responsibility of companies to avoid negative human rights impacts of their activities[5] and the responsibility to diligently detect potential harms that can ensue

[4] UN Office of the High Commissioner for Human Rights, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, UN Doc. HR/PUB/11/04 (Mar. 21, 2011) (hereinafter UN Guiding Principles).

[5] Id. Principle 13.

from the activities of the company.[6] The normative working environment of social media companies has changed significantly after the adoption of the UN Guidelines. The UN Guideline deliberately uses the term "responsibility" rather than "obligation" to convey the message that human rights are not a set of legal obligations for companies, but the expected standard of behavior. In addition, there is not a single agreement yet imposing an obligation on social media companies to respect human rights as part of their legal responsibilities under international law.[7] Therefore, according to international human rights law, social media companies do not have the legal responsibility to suppress hate speech and incitement to violence on their platforms. However, according to the UN Guiding Principles, all internet intermediaries (social media companies) share their responsibilities to respect human rights. Therefore, governments can enact laws in their jurisdictions to ensure that social media companies comply with human rights laws in terms of avoiding hate speech.

3.2 Regional Human Rights Instruments

Certain regional developments are relevant for the current discussion here. The African Charter on Human and Peoples' Rights (the Banjul Charter) provides for the right to freedom of

[6] Id. Principle 17.

[7] For information on this initiative, see the work of the Open-ended intergovernmental working group on transnational corporations and other business enterprises with respect to human rights.

expression[8]. In 2002, the African Commission on Human and Peoples' Rights (ACHPR) issued the Resolution on the Adoption of the Declaration of Principles on Freedom of Expression in Africa[9]. The Declaration of Principles refer to the ACHPR's concerns by reaffirming Article 9 of the African Charter on Human and Peoples' Rights (the Banjul Charter), which recognized freedom of expression as two rights: "the right to receive information" and "the right to express and disseminate one's opinions within the law." The Declaration of Principles was generally based on the freedom of expression provisions of the UDHR, ICCPR and the Banjul Charter. Uniquely, the Declaration of Principles specifically highlighted the importance of freedom of expression and oral traditions to African culture. The Declaration of Principles described the freedom of expression as "a fundamental and inalienable human right and an indispensable component of democracy."[10] Furthermore, it declared that "any restrictions on freedom of expression shall be provided by law, serve a legitimate interest and be necessary and in a democratic society"[11]. Although the Declaration makes it clear that restrictions can be imposed on freedom of expression, it does not provide the reasons for such restrictions. In 2014, the African

[8] See article 9 of the Charter.

[9] Refer Declaration of Principles on Freedom of Expression in Africa, African Commission on Human and Peoples' Rights, 32nd Session, 17 - 23 October, 2002: Banjul, The Gambia.

[10] See article I (1) of the Declaration of Principles.

[11] Article II (2) of the Declaration of Principles.

Union adopted the Malabo Protocol[12], which will create an international criminal law section within the African Court of Justice and Human Rights[13]. This Court has jurisdiction to prosecute corporations for a number of international crimes, including crimes against humanity and genocide, each of which encompass hate speech and incitement under certain circumstances[14]. Pursuant to the Malabo Protocol, the intention of a company to commit a crime can be determined by "proofing that it is the company's policy to do the crime that actually constitutes the crime". Assuming that the Malabo Protocol eventually enters into force[15], prosecuting social media companies for unlawful content posted on their platforms is a theoretical possibility. However, considering that the senders of illegal content are third parties rather than social media companies, it will be almost impossible to show that the company's policy is "to commit the criminal act". It is enough not to remove the problematic content even with the knowledge that the content is illegal. Therefore, the Malabo Protocol does not offer an effective way to persuade social media companies to suppress hate speech and incitement on their platform. In Europe,

[12] Protocol on Amendments to the Protocol on the Statute of the African Court of Justice and Human Rights, June 27, 2014 [hereinafter Malabo Protocol].

[13] Protocol on the Statute of the African Court of Justice and Human Rights, July 1, 2008.

[14] Jurisdiction over corporations is provided for in Article 46 C of the Malabo Protocol.

[15] The Malabo Protocol requires fifteen ratifications to enter into force. As of March 2020, more than fifteen states have signed it but none has ratified it.

article 10 of the European Human Rights Convention provides for the right to freedom of expression in a very similar manner as the ICCPR. However, the restrictions on the right seem relatively wider than the ones provided under the ICCPR. The restriction of freedom of expression for sake of protecting others privacy (secrecy) and for the protection of judicial independence are additional grounds for limiting the right under the European Convention[16]. Hate speech is not expressly provided under the terms of the Convention as a ground for restricting freedom of expression. In 2003, the Organization for Security and Co-operation in Europe (OSCE) adopted or published the Amsterdam Recommendations on Freedom of the Media and the Internet. The document reflected a recent application in support of the freedom of expression with regard to contemporaneously methods of expression, i.e. the internet (social media). The preface immediately identified the tone that supports freedom of expression, saying, "The fundamental constitutional value of the freedom of the media must not be questioned." The Amsterdam Recommendations went on to propose a number of measures related to numerous topics, including freedom of expression (Wehbé, 2018). Importantly, the first recommendation under freedom of expression stated that there is a balance between the free flow of information and the abuse of social media and therefore illegal content should be prosecuted in the country of

[16] Refer article 10 (2) of the European Convention on Human Rights.

origin.[17] The Amsterdam Recommendations went on to state that "the right to disseminate and to receive information is a basic human right," and "new forms of censorship must not be developed". In addition, in 2016, the European Commission presented Twitter, Facebook, YouTube, and Microsoft with a Code of Conduct on Countering Illegal Hate Speech Online.[18] The rules in this code focus around procedures for identifying, reporting, and removing hate speech from social media platforms. Accordingly, they undertake to establish clear and effective processes for the review of reported content and to remove unlawful hate speech within twenty-four hours after notification. In the Code of Conduct, hate speech is not defined but it is understood as incitement to violence towards particular groups. In the 2016 press release introducing this code, the Commission stressed the importance of member states' compliance with EU law to punish hate speech in local legal systems (Irving, 2018). However, with regard to social media companies, the Code of Conduct tries to use the word "commitments" instead of "obligations." The code is intended to create a normative compliance environment, as in the UN Business and Human Rights Guidelines, but it is not intended to create legally binding

[17] Recommendations from the conference on freedom of the media and the Internet, Amsterdam, 13-14 June 2003.

[18] European Commission Press Release IP/16/1937, European Commission and IT Companies Announce Code of Conduct on Illegal Online Hate Speech (May 31, 2016). More companies signed up in 2018. See European Commission, Countering Illegal Hate Speech Online #NoPlace4Hate (Oct. 18, 2018).

obligations for social media companies. Subsequent developments at the EU level, while more specific, detailed, and forceful in their language, essentially continue this voluntary approach for regulating hate speech on social media[19]. Therefore, while it is stressed that freedom of expression shall not be subjected to censorship, the international and regional legal developments impose the duty on States to regulate hate speech on any media be it online or offline media. Besides, social media companies are required to take commitments to remove hate speech contents from their media platforms in line with human rights standards.

3.3 National Human Rights Legislation

The right to freedom of expression is constitutionally provided in Ethiopia. According to the Ethiopian Constitution, the right to "freedom of expression" includes freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any media of his choice[20]. The Constitution stipulates some limitations to this right in accordance with international human

[19] European Commission, Tackling Illegal Content Online: Towards an Enhanced Responsibility of Online Platforms, COM (2017) 555 final (Sept. 28, 2017). With respect to terrorist online content specifically, which may constitute hate speech and incitement but which is a much broader category of content, the European Union has initiated steps to impose binding measures on service providers to remove content. Council of the European Union Press Release, Terrorist Content Online: Council Adopts Negotiating Position on New Rules to Prevent Dissemination (Dec. 6, 2018).

[20] Refer article 29 (2) of the FDRE Constitution.

rights standards. Accordingly, legal limitations can be laid down in order to protect the well-being of the youth, and the honor and reputation of individuals. Besides, any propaganda for war as well as the public expression of opinion intended to injure human dignity shall be prohibited by law[21]. It is not clear if hate speech constitutes one of these limitations to freedom of expression. However, it seems that the expression "public expression of opinion intended to injure human dignity" encompasses the idea of hate speech as a ground for restricting freedom of expression. As it is demonstrated above, hate speech is considered as any expression with the intention to inflict injury to the dignity of others. The limitations to freedom of expression are provided in the Ethiopian Constitution in a slightly different manner from other international human rights instruments. But this does not mean that the limitations fall short of the international standards. In addition to the Constitutional provision, Ethiopia has adopted a law that have bearing on freedom of expression. The recent law on disinformation and hate speech on social media is worth discussing here. Following the current political reform in Ethiopia after 2018, hate speech and disinformation has pervaded the social media resulting in wide spread ethnic tensions and violence throughout the country (JMC, 2020). Thus, Ethiopia has adopted

[21] See article 29(6) of the Federal Democratic Republic of Ethiopia Constitution.

a new proclamation[22] to prevent and suppress hate speech and disinformation. The proclamation considers hate speech as a threat to social harmony, political stability, national unity, human dignity, diversity and equality. The regulation of hate speech in Ethiopia has raised a fiery debate in the academia and the media particularly about its effect on the freedom of expression. Although the law was adopted with extensive deliberations and the involvement of both the academia and civil societies in the drafting process, there are still some legal criticisms and concerns on this law (Taye, 2020). To begin with, the proclamation defined hate speech as "speech that deliberately promotes hatred, discrimination or attack against a person or a discernable group of identity, based on ethnicity, religion, race, gender or disability"[23]. This definition lacks objective clarity in that it fails to provide what constitutes hatred, discrimination and attack. Thus, it is too vague for any effective implementation of the law without affecting the individuals' right to freedom of expression. Secondly, while there are many other mechanisms to control hate speech (such as digital literacy campaigns, education, and public messaging), the law excessively criminalizes and imposes punishments not proportional to the act. For instance, punishing or imprisoning people only for sharing hate speech on social media is extremely unfair and does not comply with

[22] See Hate Speech and Disinformation Prevention and Suppression Proclamation No. 1185/2020.

[23] Article 2 (2) of the Proclamation.

international standards and recommendations. Thirdly, the proclamation contains provisions that increase the penalty for individuals and online groups which have more than 5,000 follower[24].This provision can potentially repress political dissidence as most politicians and activists with more than 5000 followers reflect their political ideologies on social media. Finally, the government is expected to prepare directives to help implement this legislation. This directive can identify some terms that are missing in the proclamation, thereby limiting their negative impact on freedom of expression. However, the chances of threatening people's rights can be further increased by institutionalizing the mass surveillance of social media users to control the information that people share online. Therefore, the Ethiopian approach on regulating hate speech on social media is very problematic as it significantly endangers freedom of expression (Brown, 2017).

4. The Regulatory Issues on 'Hate speech' on Social Media

4.1 The Rationales for Legal Control on 'Hate Speech'

Today, most democracies in the world prohibit hate speech, because such provocative words do not only harm but also disrupt public order by causing violent consequences such as hate crimes

[24] See article 7(4) of the Proclamation.

among other violent consequences (Yadav, 2018). The "incitement to violence" standard appears to be a more concrete basis for prohibiting speech through legislation. Incitement to violence demands a greater level of harm to be demonstrated in comparison to other forms of hate speech and therefore justifiably be the subject of censure by criminal law (Yadav, 2018). As far as criminalization of speech is concerned, it remains a debatable issue with the legal scholars divided between what kind of speech should ideally be criminalized; should only a certain type of hate speech be banned and whether all hate speech be made punishable by criminal law or it can be dealt under civil law. However, hate speech, which has been shown to "provoke violence", is considered to be a serious case and deserves a tough action to prevent further damage (Brannon, 2019). Here, some advocate that criminal sanctions can optimally be implemented to reduce hate. Sometimes hate speech is directed at a particular individual, but sometimes at a group of people, or even an entire community. In consequence, the gravity of the repercussions of hate speech are in a direct correlation with and heavily depend on its respected target. On an individual level, for instance, bearing in mind that people have this inherent need to be respected and treated with dignity, when they become a target of hate speech, the harm done invariably is in some form of immediate psychological injury, i.e. emotional distress, with the possibility of physical attacks and bodily harm not being entirely excluded either. In the best case scenario, individuals who are victims of hate speech may feel

galvanized to stand up to those who hurt them and advocate their dignity, integrity and well-being. But in the worst case scenarios, harm can lead to a variety of psychosomatic conditions and disorders, one of which is absolutely depression, which can completely destroy not only the victim's life, but also the lives of the closest and loved ones to them. When hate speech is directed at a group of people or the whole community, it poses a much more serious threat than it can lead to immediate violation of peace and increased violence and deaths (Silvana Neshkovska, 2018). Unfortunately, the world has so far seen such bloody and devastating events, initiated by hate speech, among other things. Employing the 'Pyramid of Hate Tool' developed by the Anti-Defamation League, Roiha (2016) has shown that hate speech can contribute to stigmatizing, marginalizing, and intimidating members of distinct and vulnerable groups in a country. The Pyramid shows biased behaviors, growing in complexity from the bottom to the top (general bias, individual acts of prejudice, discrimination, bias-motivated violence, and genocide). Although the behaviors at each level negatively impact individuals, groups and society, moving up the pyramid the behaviors have more life-threatening consequences, so that the upper levels are supported by the lower levels. If people or institutions treat lower-level behavior as acceptable or "normal", the next-level behavior becomes more acceptable. However, it is largely unknown how hateful online content promotes and shapes negative attitudes, prejudices, and criminal / hateful behavior towards certain

identifiable individuals and groups (Rohlfing, 2017). It is widely believed that the dissemination of hate speech on social media particularly by politicians can result in a public disturbance, violence, and communal killings (Samarth Mishra, 2019). The key rationales for the regulation of hate speech on social media are therefore based on two justifications: incitement to violence and social harm (harm to some groups in the community). Currently, many countries are adopting laws that control hate speech, and incitement to violence and the protection of certain categories of people, particularly minorities, are the underlying justifications for enacting such laws. Although the regulation of hate speech focuses, in some countries, on the protection of some groups and minorities, it seems that the Turkish approach focuses more on the protection of state unity than on vulnerable groups and minorities (Kınık, 2018).

4.2 The Challenges of Regulating 'Hate Speech'

4.2.1 Excessive restriction or Over-criminalization

Hate speech is extremely regulated in some nations, or excessively criminalized in some instances. There have been instances of serious violent hate speech in India, for example, which have turned into violence, mob murders and lynching of some individuals and groups (Yadav, 2018). Consequently, by issuing serious laws that excessively limit and overly criminalize hate speeches, the government took legal measures. The problem in the country could not, however, be prevented by excessive

legal control. Instead, the cases of hate speech have increased all the more with the increased legal constraints. As a solution to this problem, Yadav (2018) proposes that instead of issuing excessive legal restrictions and criminalization of hate speeches, 'Alternative Dispute Resolution (ADR)'[25] and 'counter-speech'[26] can be effective responses to the problem of hate speech in the country. Many countries today have political debates on anti-hate speech laws and censorship of hate speech on social media (L.Hundson, 2019). Some human rights advocates argue that excessive restrictions on hate speech on social media can significantly erode an individual's right to freedom of expression (Brannon, 2019). Besides, some politicians and think tanks argue that excessive speech restriction and media censorship can be manipulated to silence political opposition in some countries (Repucci, 2019; and Ulusoy, 2013) The German parliament, for example, passed an anti-online hate speech law which threatens providers of online platforms such as Facebook with fines up to

[25] ADR shifts the focus from court-centered formal legal proceedings to the settlement of the dispute between parties by way of negotiation, mediation, arbitration and/or conciliation.

[26] Counter-speech is, simply put, a response to hateful speech which might alternately call for violence, promote hate or uses incendiary words to provoke or defame others. Counter-speech is a definitive and exacting answer to such speech and is solely aimed at undoing the damage wrought by the hate speech in the first place. This responsive speech may take myriad forms depending upon the medium used to propagate hateful speech: it may be a direct answer in the form of true facts to a hateful message concocted out of a false claim or fake news; clarification of any dubious claims made or even using sarcasm, humor and cartoons, memes and caricatures to counter and defuse the tension that a hate message aims to promote.

EUR 50 million for failing to delete "criminal" content that is "obviously unlawful" (Schwarz, 2019). It is obvious that excessive regulation and criminalization of online hate speech will invite policy makers to employ social media censorship. This social media censorship can in tern restrain the right to freedom of expression and political dissidence as it has been witnessed in Turkey (Altıparmak, 2015). Instead of excessive criminalization, other measures such as adopting ethical code of conduct and taking disciplinary measures can be employed to reduce the negative consequences of hate speech on social media (TAŞ, 2017).

4.2.2 Issues of Jurisdiction

Some countries are taking legislative measures to control hate speech on social media. The measures include criminalizing hate speech and imposing civil liabilities in the form of paying compensation to be paid by the offender to the victim (O'Regan, 2018). However, concerning the criminal investigation and prosecution of hate crimes, there are issues relating to the location, which concerns the place where the crime was committed (e.g. posted) and to the legal position of the country in which the service is provided (e.g. the Internet server's host country). Although a hateful expression on social media might constitute an online hate crime in the country where it was committed, it may not be regarded as such in the country through which the service was provided. The specific character of the

Internet poses serious challenges to any attempts to regulate content, particularly while respecting freedom of expression. Unlike traditional media, it is often very difficult to identify the authors of content available online. Material which originates in one country is copied, edited, and shared across national borders, and can be hosted in different countries, subject to different legislation (OSCE-ODIHR, 2010). Besides the challenges posed by jurisdiction and legislation, there are also practical challenges involved. For instance, once any material is posted on the Internet, it is usually shared on several servers or online repositories, which makes it difficult if not impossible to remove all copies completely. Due to global nature of social media platforms, the major obstacle in addressing hate speech is the issue of indeterminacy regarding the applicable law and jurisdiction (Gelashvili, 2018). International human rights law considers the concept of legality a vital precondition to limit freedom of expression. This is primarily evidenced in the requirements of Article 19 of the ICCPR. The requirement of legal certainty is one of the key requirements of Article 19(3) of the ICCPR, i.e. it must be provided by law in order to justify the limitation. In the digital space, however, as social media platforms do not operate within national borders, the requirement of legality becomes problematic. The grounds for limitation prescribed in law in one country might not constitute unlawful hate speech in another.

4.2.3 The Conceptual Fluidity of 'Online Hate Speech'

The lack of a global legal consensus of what constitutes an online hate speech and what lies within the domain of freedom of speech/expression has made it extremely difficult to regulate online hate speech (Sellars, 2016). While the EU along with popular social media companies such as Facebook, Twitter, YouTube, Google and Microsoft have recently released a code of conduct to tackle and prevent online hate speech, the US has somehow given protection to hate speech by the First Amendment (Rohlfing, 2017). For instance, there is no a solid line distinction between the concepts of hate speech and hate crime in the US. Hate speech will qualify to be a hate crime only if such a speech is considered to be a clear and immediate threat to other individual or groups (Beausoleil, 2019). This has resulted in the inability of the law enforcement to prosecute most cases. For example, whilst there can be prosecutions where there is a clear and immediate threat to individuals and/or groups in both online and offline contexts, what exactly constitutes a threat is harder to prove in online contexts. Defining a hate speech on social media is a rather complicated process. Since there is no agreement on what constitutes hate speech, the work of detecting and filtering hate speech from online dialogues is a very difficult task (Leandro Silva, 2016). Therefore, the line between freedom of expression and hate speech, a speech potentially motivating violent acts, should be more clearly defined in the literature and in the context

of criminal law for the purpose of legal regulation (Barbara Perry et al., 2020). In addition to this, it is imperative to distinguish offline hate speech from online hate speech. Hate speech in social media has emerged with new challenges that require new definitions for the term for effective regulation (Onbaşi, 2015). Although there is no fundamental difference in effect, there are distinctive features such as permanence, itinerancy, anonymity and transnationality in online hate speech (Roiha, 2016). First, hate speech can stay online for a long time and can remain in different forms across different platforms and be linked over and over again (Roiha, 2016; and Levin, 2002). In this sense, the effect of online hate speech can be more devastating and lasting than offline hate speech. Second, even when content is removed, it can be revived elsewhere, on the same platform as another name, or in other online areas (Udoh-Oshin, 2017; and Roiha, 2016). Third, with the possibility of being anonymous, people tend to feel more comfortable expressing hate as they will not be discovered or have to deal with any consequences (Roiha, 2016; and MEŽNAR, 2014). Anonymity can also hamper prosecution, particularly on social media. Fourth, transnational internet access increases the effects of hate speech and also creates problems with legal mechanisms to combat online hate speech (Roiha, 2016; and Matthew L. Williams P. B., 2015).

4.2.4 Divergence of Regulatory Models

Even though hate speech has been identified as a threat to individuals and groups in many countries across the world, there is a sharp divergence in the regulatory model each country follows to regulate online hate speech. While some countries enact stringent laws that criminalize and severely punish the act, other countries simply impose lenient civil liabilities on the perpetrators. Whether hate speech should be punished by criminal law or it should simply be subjected to civil liability has been highly contested as well (Richards, 2009). On the basis of the universality of human rights, the right to freedom of expression should be taken under equal protection in all countries of the world. Likewise, the regulation of hate speech should be harmonized accordingly (Alkiviadou, 2018). In the European Union, for example, the efforts of the Council of Europe to achieve greater unity in member states' legislation on hate crimes have largely yielded results (Jourová, 2016). Nevertheless, if the national laws of the European countries are compared with the ones of the United States of America, many striking differences will be noted (Silvana Neshkovska, 2018). The policy of the United States is highly influenced by the First Amendment of the federal Constitution. Hate speech, which is commonly considered a political speech, falls most of the time under its protection (Beausoleil, 2019). In contrast, this is not the case with many European countries. In Denmark, France, the UK and Germany, people have been facing trial for crimes involving hate speech on

the Internet (OBC Transeuropa, 2018). There are some more extreme examples of national laws in other parts of the world, especially when it comes to hate speech towards religious groups. For example, in Bangladesh, a person can face up to ten years in jail for defaming a religion (Rabbee, 2018).

4.2.5 The Legal Status of Social Media Companies

The legal status of social media companies at domestic laws determines the government policy towards the regulation of hate speech (George, 2015). If social media companies are delegated by the government to control their platforms, they are playing the regulatory roles of the state. In this case, the companies shall respect the constitutional guarantee of freedom of expression of individual users of their platforms. If they are considered as corporate persons such as news agencies or broadcasting media only, then the government shall protect the rights of such social media platforms to edit or disseminate the content in line with applicable laws. Taking lessons from the US experience, there are at least three possible frameworks for analyzing governmental restrictions on social media sites' ability to regulate user content (Brannon, 2019)First, using the company town analogue, social media sites could be treated as state actors who, when they regulate protected speech, are themselves bound to follow the First Amendment. If social media sites were treated as state actors under the First Amendment, then the Constitution itself would constrain their conduct, even in the absence of legislative

regulation. In the second major framework, social networking platforms can be seen as equivalent to special industries such as common carriers or broadcast media. In this framework, if special aspects of social media sites threaten the use of the medium for communicative or expressive purposes, courts might approve of content-neutral regulations intended to solve those problems. The third analogy considers social media outlets as news reporters, who, when making editorial decisions, normally enjoy the full protection of the First Amendment. If social media sites were deemed equivalent to newspaper editors when making decisions on whether and how to present user content, then those editorial decisions would receive the broadest protections under the First Amendment. Any government regulation that changes the editorial preferences of social media sites by forcing them to host content that they would not otherwise transmit, or by forcing them to remove content they want to host, may be subject to scrutiny. Which of these three frameworks will apply depends largely on the specific action that is being regulated. Under US law, for example, social media platforms may be more likely to receive constitutional protection when they apply more content discretion in delivering user-generated content rather than impartially transmitting all such contents (Hawkins, 2012). In addition, some types of speech receive less protection under the constitution. Thus, courts are more likely to support regulations that target categories of discontent speech, such as obscene or violence-promoting speech. Finally, if a law targets the behavior of the

social media site rather than speech, it may never trigger the constitutional protection of free speech.

4.3 Approaches on the Regulation of 'Hate Speech' on Social Media

4.3.1 State-Imposed Regulation (Domestic Regulation)

At domestic level, some countries are imposing legal obligations on social media companies to eliminate some contents including hate speech (O'Regan, 2018). This type of state-imposed laws and regulations, however, continues to remain the exception rather than the norm. In order to moderate online content, the general position at international, regional and domestic levels is to rely on voluntary commitments and on self-regulation (Irving, 2018). Germany´s Network Enforcement Act (or NetzDG) is perhaps the most widely debated and controversial initiative enacted by a liberal democracy to counter illegal online content including hate speech (Fiss, 2019). The NetzDG imposes intermediary liability for social media networks with over two million registered users. These platforms need to remove illegal content such as hate speech and "blackening of religions" marked by individuals. All content that is "clearly illegal" should be removed within a 24-hour period. The deadline for all other "illegal content" is seven days. Failure to remove illegal content can be punishable with a fine up to 50m euros. The scope and nature of NetzDG has been widely debated, and the law has been

criticized for being vague and over-inclusive, "privatizing" online censorship with little transparency or due process, and encouraging "over-implementation" by providing an incentive to err on the side of caution rather than free expression. Moreover, illegal content categories include criminal insults and insults, as well as a "defamation" that violates international human rights standards. Many states have been directly or indirectly inspired by German efforts to tighten its intermediary obligation. Since the adoption of the NetzDG, at least 13 countries[27] (as identified in this paper) in addition to the EU, have adopted or proposed models of intermediary liability broadly similar to the NetzDG matrix.

4.3.2 Self-Regulation by Social Media Companies

Social media have correctly recognized how salient their role in the distribution of hate messages is, and, in response to that they have agreed to lay out terms of agreement, hoping to regulate and restrict such potentially hazardous activities (Silvana Neshkovska, 2018). Thus, for instance, Facebook in its terms of agreement specifies that "content that attacks people based on

[27] According to Freedom House's assessment of freedom on the Internet (2019) , five of those countries are ranked as being "not free" (Honduras, Venezuela, Vietnam, Russia and Belarus), five are ranked "partly free" (Kenya, India, Singapore, Malaysia and Philippines), and only three ranked "free" (France, UK and Australia). With the exception of India, Kenya, Vietnam, and Australia all these countries – as well as the EU – have explicitly referred to the NetzDG as an inspiration or justification for their models of intermediary liability.

their actual or perceived race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or disease is not allowed" to be posted on their platform (Tuba Ciftci et al., 2018). Twitter also forbids to "publish or post direct, specific threats of violence against others". YouTube's policy is also strongly against permitting "any speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status and sexual orientation/gender identity" to be disseminated via their website. Evidently, social media, overall, seem to be in agreement when it comes to prohibiting hate speech. In that respect, it is important to mention that in May 31, 2016, many of them (Facebook, Google, Microsoft, and Twitter) jointly agreed to follow a European Union code of conduct which obligates them to review "the majority of valid notifications for removal of illegal hate speech posted on their services within 24 hours"[28]. Apps and platforms on social media are also responsible for restricting freedom of speech. In recent times, cases of continuous monitoring of user activities and censoring their posts via networking sites have also occurred (Samarth Mishra, 2019). WhatsApp is misused to deliver misinformation, which has led to the false killing of individuals due to suspicion by the mob. Fake news, especially related to politics has up-surged on such

[28] Information extracted from "Facebook, YouTube, Twitter and Microsoft sign EU hate speech code". *The Guardian*. Retrieved at *https://www.theguardian.com/technology/2020/march/31/facebook-youtube-twittermicrosoft-eu-hate-speech-code.*

platforms. Offensive messages created by corrupt users are shared and transmitted blindly. As the dangerous side of social media becomes more evident, states and academics ask how this phenomenon can and should be addressed. In particular, it turns to the attention of social media companies about what they should do or take to suppress hate speech and provocation published on their platforms. Until now, social media companies have been left to regulate themselves. This approach has come under increasing criticism in light of data breaches, election intervention, hate speech, and provocation scandals. In response to this criticism, some of the prominent social media companies, Facebook increasingly acknowledged their role in these events, stating that they have not fulfilled their responsibilities until now (Irving, 2018). Major social media platforms are also using Artificial Intelligence (AI) and computational methods to detect and remove online hate speech from their services (Budek, 2019). However, there are concerns about the bias of computer-based assessments of what online content is acceptable or how the learning of an AI system can be used in unintended and unanticipated ways (Melhuish, 2018). Despite the use of some soft-wares to detect online hate speech by the social media companies, the detection of such contents has never been an easy task. The design of social media posts generates one degree of difficulty. In other words, social media textual material is rarely well-formed and sometimes includes paralinguistic elements, such as URLs, emoticons and other special characters (Simonite,

2020). Another degree of difficulty stems from the naturally unstable nature of hate speech detection, because most social media posts do not contain hate speech or profanity (Robert Gorwa et al., 2020). Lastly, a third degree of difficulty emerges from the subjective nature of hate speech labeling. One good example to the subjective nature of decisions about hate speech and offensive content is how people react very differently to the use of the word "fuck" when it is used as part of a hashtag, as opposed to when it is used without a hashtag (Pedro Alonso, 2019).

5. Conclusion

The issue of hate speech regulation has attracted the attention of many scholars from legal and philosophical dimensions. The vast majority of this attention is focused on presenting and critically evaluating arguments for and against hate speech prohibitions. Although hate speech has been identified as one of a growing threat to human rights both at international and local levels, imposing legal restrictions on it has faced many challenges. This article has identified some of the major challenges of regulating hate speech. Firstly, the lack of a global legal consensus of what constitutes an online hate speech and what lies within the domain of freedom of speech/expression has made it extremely difficult to regulate online hate speech. Secondly, the over-criminalization of hate speech and excessive media censorship in some countries

has resulted in restraining of the right to freedom of expression and political dissidence. Thirdly, due to global nature of social media platforms, the major obstacle in addressing hate speech is the issue of indeterminacy regarding the applicable law and jurisdiction. Fourthly, while some countries enact stringent laws that criminalize and severely punish the act, other countries simply impose lenient civil liabilities on the perpetrators creating divergence in the regulatory model. Finally, the discrepancy on the policies as to who should regulate social medias; self-regulation by the social media owners or government-imposed regulation has not been adequately addressed in conformity with the international standards set by human rights instruments.

References

AFP. (2020, May 13). *Hate speech in focus again: Facebook apologises for role in 2018 Sri Lanka unrest*. Retrieved from https://www.deccanchronicle.com/technology/in-other-news/130520/hate-speech-in-focus-again-facebook-apologises-for-role-in-2018-sri-l.html

ALA. (2017, December). *American Library Association*. Retrieved from Hate Speech and Hate Crime: http://www.ala.org/advocacy/intfreedom/hate

Al-Dossari, A. A.-H. (2019). Detection of Hate Speech in Social Networks: A Survey On Multilingual Corpus. Computer Science & Information Technology (CS & IT).

Alkiviadou, N. (2018). Hate speech on social media networks: towards a regulatory framework? *Information & Communications Technology Law*. doi:10.1080/13600834.2018.1494417

Altıparmak, Y. A. (2015). Silencing Effect on Dissent and Freedom of Expression in Turkey in Journalism at Risk: Threats, Challenges and Perspectives, Council of Europe Publishing, 2015, 145-173. *Council of Europe Publishing*, 145-173. Retrieved June 25, 2020, from https://www.researchgate.net/publication/323511383_Silencing_Effect_on_Dissent_and_Freedom_of_Expression_in_Turkey_in_Journalism_at_Risk_Threats_Challenges

_and_Perspectives_Council_of_Europe_Publishing_2015_145-173

Altman, A. (1993). Liberalism and Campus Hate Speech: A Philosophical Examination. *An International Journal of Social, Political, and Legal Philosophy, 103*(2), 302-317. Retrieved June 3, 2020, from https://www.journals.uchicago.edu/doi/abs/10.1086/293497?journalCode=et

Apalara, R. (2017). Striking a Balance: Freedom of Expression and the Prohibition of Hate Speech and Offensive Remarks. *Social Science Research Network (SSRN)*, 1-76. Retrieved June 7, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035410

Appiagyei-Atua, K. (2005). A Review of Theories of Expression in the Context of the Development Argument. *University of Ghana Law Journal, XXIV*, 197-227. Retrieved March 22, 2020, from https://www.researchgate.net/publication/275271275_A_REVIEW_OF_THEORIES_OF_EXPRESSION_IN_THE_CONTEXT_OF_THE_DEVELOPMENT_ARGUMENT/link/553651a20cf20ea35f11f91f/download

Arcan, H. E. (2013). Interrupted Social Peace: Hate Speech in Turkish Media. *The IAFOR Journal of Media, Communication and Film, 1*(1), 43-56.

Audrin, C. B. (2019). Toward an Understanding of the Characteristics of Secondary School Cyberhate Perpetrators. *Front. Educ*. doi:https://doi.org/10.3389/feduc.2019.00046

Barbara Perry, D. A. (2020). Planting Hate Speech to Harvest Hatred: How Does Political Hate Speech Fuel Hate Crimes in Turkey? *International Journal for Crime, Justice and Social Democracy .Advance online publication*, 1-17. doi:https://doi.org/10.5204/ijcjsd.v9i4.1514

Beausoleil, L. E. (2019). Free, Hateful, and Posted: Rethinking First Amendment Protection of Hate Speech in a Social Media World. *Boston College Law Review, 60*(7). Retrieved June 12, 2020, from https://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3797&context=bclr

Bell, J. (2009). Restraining the Heartless: Racist Speech and Minority Rights. *Indiana Law Journal, 84*, 963-979. Retrieved June 13, 2020, from http://ilj.law.indiana.edu/articles/84/84_3_Bell.pdf

Benesch, S. (2014). *Defining and Diminishing Hate Speech*. Retrieved from State of the World's Minorities and Indigenous Peoples 2014: https://minorityrights.org/wp-content/uploads/old-site-downloads/mrg-state-of-the-worlds-minorities-2014-chapter02.pdf

Brannon, V. C. (2019). *Free Speech and the Regulation of Social Media Content.* Congressional Research Service. Retrieved from https://crsreports.congress.gov

Brown, A. (2017). What is hate speech? Part 1: The Myth of Hate. *Law and Philosophy, 36*(4). doi:10.1007/s10982-017-9297-1

Budek, K. (2019, January 24). *How artificial intelligence can fight hate speech in social media*. Retrieved from deepsense.ai: https://deepsense.ai/artificial-intelligence-hate-speech/

CeaseFire Centre for Civilian Rights. (2019). *Peoples under Threat 2019* . Minority Rights Group International.

Chong, D. (2006). Free Speech and Multiculturalism In and Out of the Academy. *Political Psychology, 27*(2), 29-54. doi: https://doi.org/10.1111/j.1467-9221.2006.00448.x

Demaske, C. (2019). Social Justice, Recognition Theory and the First Amendment: A New Approach to Hate Speech Restriction. *Journal of Communication Law and Policy, 24*(3), 347-401. doi:https://doi.org/10.1080/10811680.2019.1627800

Donovan, J. (2019, October 28). *Navigating the Tech Stack: When, Where and How Should We Moderate Content?* Retrieved from INTERNET GOVERNANCE, PLATFORM GOVERNANCE:

https://www.cigionline.org/articles/navigating-tech-stack-when-where-and-how-should-we-moderate-content?gclid=CjwKCAjw8pH3BRAXEiwA1pvMsYYWtF-G2eGsXdOFFdGvSfJt1zQjOJlOjybYETAYXfZLihmXNkjAGRoC-SsQAvD_BwE

Dural, B. G. (2012). Othering Through Hate Speech: The Turkish-Islamist (V)AKIT Newspaper as a Case Study. *Turkish Studies, 13*(3), 489-507. doi:http://dx.doi.org/10.1080/14683849.2012.715482

European Parliament. (2018, November 6). *The role of social media in ethnic conflicts*. Retrieved from Parliamentary questions: https://www.europarl.europa.eu/doceo/document/E-8-2018-005636_EN.html

Fiss, J. M. (2019). *The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship.* Copenhagen: Justitia.

Gelashvili, T. (2018). Hate Speech on Social Media: Implications of Private Regulation and Governance Gaps. *JAMM07 Master Thesis*. Lund University, Faculty of Law.

George, C. (2015). *Hate Speech Law and Policy*. Retrieved from Wiley Online Library:

https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118767771.wbiedcs139

George, C. (2020, June 3). *Ethical Journalism Network*. Retrieved from Hate Speech: A Dilemma for Journalists the World Over: https://ethicaljournalismnetwork.org/resources/publications/ethics-in-the-news/hate-speech

GÖLCÜ, Ş. B. (2012). Hate Speech In Turkish And Israel Press: A Comparison Analysis Of Mavi Marmara Raid. Turkey: Selçuk University.

Hawkins, S. (2012, March 8). *How Free Speech and Social Media Fit Together*. Retrieved from Social Media Examiner: https://www.socialmediaexaminer.com/how-free-speech-and-social-media-fit-together/

Holmes, K. R. (2018, October 22). *The Origins of "Hate Speech"*. Retrieved from The Heritage Foundation: https://www.heritage.org/civil-society/commentary/the-origins-hate-speech

HRW. (2018). *Human Rights Watch*. Retrieved from Somalia: Events of 2018: https://www.hrw.org/world-report/2019/country-chapters/somalia

Hylton, K. N. (1996). Implications of Mill's Theory of Liberty for the Regulation of Hate Speech and Hate Crimes. *The University of Chicago Law School Roundtable, 3*(1).

Retrieved June 8, 2020, from https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1377&context=roundtable

Irving, E. (2018). Suppressing Atrocity Speech on Social Media: Symposium on Non-State Actors and New Technologies in Atrocity Prevention. *AJIL UNBOUND, 113*, 256-261. doi:10.1017/aju.2019.46

Jacobson, D. (2016, August 30). *Freedom of Speech under Assault on Campus*. Retrieved from CATO Institute: https://www.cato.org/publications/policy-analysis/freedom-speech-under-assault-campus

JMC. (2020, June 4). *Jack Miller Center*. Retrieved June 4, 2020, from Alexander Meiklejohn: Free Speech and its Relation to Self-government: https://jackmillercenter.org/cd-resources/alexander-meiklejohn-free-speech-relation-self-government/

Jourová, V. (2016, June). *Code of Conduct–Illegal online hate speech Questions and answers.* Retrieved from https://ec.europa.eu/info/sites/info/files/code_of_conduct_hate_speech_en.pdf

Juanatey, A. C. (2016, December 12). *Hate speech in social media: a state-of-the-art review*. Retrieved from https://www.uclg-

cisdp.org/sites/default/files/Hate%20Speech%20A%20State%20of%20the%20Art.pdf

Kaniklidis, C. (2015). Free Speech, Hate Speech and Principles of Community: The Case Against Free Speech Absolutism. *Researchgate*. Retrieved June 12, 2020, from https://www.researchgate.net/publication/274835999_Free_Speech_Hate_Speech_and_Principles_of_Community_The_Case_Against_Free_Speech_Absolutism

Kınık, C. D. (2018). Nationalist bias in Turkish official discourse on hate speech: a Rawlsian criticism. *Turkish Studies*, 1-24. doi:10.1080/14683849.2018.1479961

L.Hundson, D. (2019, April 1). *Free speech or censorship? Social media litigation is a hot legal battleground*. Retrieved from ABAJOURNAL: https://www.abajournal.com/magazine/article/social-clashes-digital-free-speech

Laub, Z. (2019, June 7). *Council on Foreign Relations*. Retrieved from Hate Speech on Social Media: Global Comparisons: https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons

Leandro Silva, M. M. (2016). Analyzing the Targets of Hate in Online Social Media. *AAAI ICWSM*, 1-4.

Lee, S. P. (2010). Hate Speech in the Marketplace of Ideas. *The Philosophical Foundations of Law and Justice*, 13-26. doi:10.1007/978-90-481-8999-1_2

Lee, S. P. (2010). Hate Speech in the Marketplace of Ideas. *AMINTAPHIL: The Philosophical Foundations of Law and Justice, 3*. doi:https://doi.org/10.1007/978-90-481-8999-1_2

Levin, B. (2002). Cyberhate: A Legal and Historical Analysis of Extremists' Use of Computer Networks in America. *American Behavioural Scientist*. doi:https://doi.org/10.1177/0002764202045006004

Lombardi, C. (2019). *The Illusion of a "Marketplace of Ideas" and the Right to Truth*. Retrieved from American Affairs: https://americanaffairsjournal.org/2019/02/the-illusion-of-a-marketplace-of-ideas-and-the-right-to-truth/

Maas, C. (2020, February 15). *Ethiopia parliament passes law punishing hate speech*. Retrieved from Jurist.org: https://www.jurist.org/news/2020/02/ethiopia-parliament-passes-law-punishing-hate-speech/

Macedo, R. P. (2017). Freedom of Expression: what lessons should we learn from US experience? *Rev. direito GV, 13*(1). doi:http://dx.doi.org/10.1590/2317-6172201711

Macedo, R. P. (n.d). Freedom of Expression: What lessons should we learn from US experience? *Direito USP/FGV Direito*. São Paulo.

Matthew L. Williams, P. B. (2015). Cyberhate on Social Media in the aftermath of Woolwich: A Case Study in Computational Criminology and Big Data. *The British Journal of Criminology, 56*(2), 211–238. doi:https://doi.org/10.1093/bjc/azv059

Matthew L. Williams, P. B. (2019). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *BRIT. J. CRIMINOL., 2020*(60), 93-117. doi:doi:10.1093/bjc/azz049

Melhuish, E. P. (2018). Online Hate Speech: A Survey on Personal Experiences and Exposure among Adult New Zealanders. Netsafe. Retrieved March 21, 2020, from https://www.netsafe.org.nz/wp-content/uploads/2019/11/onlinehatespeechsurvey-survey-2018.pdf

MEŽNAR, R. Č. (2014). ONLINE HATE-SPEECH AND ANONYMOUS INTERNET COMMENTS: HOW TO FIGHT THE LEGAL BATTLE IN SLOVENIA? *ANNALES · Ser. hist. sociol.* Retrieved June 12, 2020, from https://zdjp.si/wp-content/uploads/2015/08/ceferin-meznar.pdf

O'Regan, C. (2018). Hate Speech Online: an (Intractable) Contemporary Challenge? *Current Legal Problems, 71*(1), 403–429. doi:10.1093/clp/cuy012

OBC Transeuropa. (2018, January 31). *Hate speech: what it is and how to contrast it*. Retrieved from Resource Centre for Media Freedom in Europe: https://www.rcmediafreedom.eu/Dossiers/Hate-speech-what-it-is-and-how-to-contrast-it

OHCHR. (2019, May 29). *Preventing incitement to hatred and violence in the Central African Republic*. Retrieved from United Nations Human Rights Office of the High Commissioner: https://www.ohchr.org/EN/NewsEvents/Pages/PeacekeepersDay2019.aspx

Oltmann, S. M. (2016). Intellectual Freedom and Freedom of Speech: Three Theoretical Perspectives. *Information Science Faculty Publications*, 153-171. Retrieved June 3, 2020, from https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1032&context=slis_facpub

Onbaşi, F. G. (2015). Social Media and the Kurdish Issue in Turkey: Hate Speech, Free Speech and Human Security. *Turkish Studies, 16*(1), 115-130. doi:10.1080/14683849.2015.1021248

OSCE-ODIHR. (2010). *Incitement to Hatred vs. Freedom of Expression: Challenges of combating hate crimes motivated by hate on the Internet.* Warsaw: Report of the OSCE-ODIHR Expert Meeting. Retrieved March 18, 2020, from https://www.osce.org/odihr/68750?download=true

Paula Thompson, R. I. (2019). *Islamophobia in India: Stoking Bigotry.* Berkeley: Islamophobia Studies Center. Retrieved March 21, 2020, from https://irdproject.com/wp-content/uploads/2019/04/Islamophobia-in-India-Web-Spread.pdf

Pedro Alonso, R. S. (2019). The North at HASOC 2019: Hate Speech Detection in Social Media Data. *FIRE 2019* (pp. 12-15). Kolkata: Creative Commons License Attribution 4.0 International (CC BY 4.0).

Pejchal, V. (2015). Hate Speech Regulation in Post-Communist Countries: Migrant Crises in the Czech and Slovak Republics. *International Journal for Crime, Justice and Social Democracy, 7*(2), 58-74. doi: 10.5204/ijcjsd.v7i2.500.

Peonidis, F. (1998). Freedom of Expression, Autonomy, and Defamation. *Law and Philosophy*, 1-17. doi: 10.1023/A:1005850325859

Quent, M. (2019, November 24). *Opinion: The clear line between hate speech and free speech*. Retrieved from DW Made for Minds: https://www.dw.com/en/opinion-the-clear-line-between-hate-speech-and-free-speech/a-51393218

Quinn, T. (2018). *Monitoring Hate Speech: Challenges and Strategies.* HateBase.

Rabbee, S. (2018, October 2). *Bangladesh: the latest assault on free speech*. Retrieved from The Lowy Institute: https://www.lowyinstitute.org/the-interpreter/bangladesh-latest-assault-free-speech

Radu Meza, A. M. (2019). Targets of Online Hate Speech in Context. A Comparative Digital Social Science Analysis of Comments on Public Facebook Pages from Romania and Hungary. *Intersections. East European Journal of Society and Politics, 4*(4), 26-50. doi:10.17356/ieejsp.v4i4.503

Reid, A. (2019). Does Regulating Hate Speech Undermine Democratic Legitimacy? A Cautious 'No'. *Res Publica, 26*, 181–199. doi:https://doi.org/10.1007/s11158-019-09431-6

Repucci, S. (2019). *Media Freedom: A Downward Spiral*. Retrieved from Freedom House:

https://freedomhouse.org/report/freedom-and-media/2019/media-freedom-downward-spiral

Richards, D. J. (2009). Rethinking Free Speech and Civil Liability. *Columbia Law Review, 109*(7), 1650-1707. Retrieved June 7, 2020, from https://www.jstor.org/stable/40380373

Robert Gorwa, R. B. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 1-25. doi:https://doi.org/10.1177/2053951719897945

Rohlfing, S. (2017, January 17). The Role of Social Networking in Shaping Hatred: An Exploration into User Responses to and Influence andPermissibility of Online Hatred . *The thesis is submitted in partial fulfilment of the requirements for the award of the degree*. Portsmouth, UK.

Roiha, O. J. (2016). Backgrounds, Experiences and Responses to Online Hate Speech: A Comparative Cross-Country Analysis. *Words are Weapons. Preventing Redressing Inhibiting Hate Speech in New Media*.

Romero, A. (2016). The line between free speech and hate speech. *ResearchGate*. Retrieved June 13, 2020, from https://www.researchgate.net/publication/311064494_The_line_between_free_speech_and_hate_speech

Samarth Mishra, A. K. (2019). Balancing Freedom of Expression and Hate Speech: Case of India. *Pramana Research Journal, 9*(6). doi:10.6084/m9.figshare.11345261

Samples, J. (2019, April 9). *Why the Government Should Not Regulate Content Moderation of Social Media*. Retrieved from CATO Institute: https://www.cato.org/publications/policy-analysis/why-government-should-not-regulate-content-moderation-social-media

Scanlon, T. (1972). *A theory of free expression. Philosophy and Public Affairs.*

Scheffler, A. (2015). *The Inherent Danger of Hate Speech Legislation A Case Study from Rwanda and Kenya on the Failure of a Preventative Measure*. Retrieved from Friedrich-Ebert-Stiftung (FES): https://library.fes.de/pdf-files/bueros/africa-media/12462.pdf

Schultz, D. (2017, June). *The First Amendment Encyclopedia*. Retrieved from https://www.mtsu.edu/first-amendment/article/999/marketplace-of-ideas

Schwarz, K. M. (2019, November 3). Fanning the Flames of Hate: Social Media and Hate Crime. *SSRN Electronic Journal*, 1-85. doi:10.2139/ssrn.3082972

Seidman, L. M. (2018). Can Free Speech Be Pr ee Speech Be Progressive? *Georgetown University Law Center*. Retrieved June 3, 2020, from https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=3056&context=facpub

Sellars, A. (2016). Defining Hate Speech. *Berkman Klein Center Research Publication No. 2016-20; Boston Univ. School of Law, Public Law Research Paper No. 16-48*. Retrieved June 12, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2882244

Silvana Neshkovska, Z. T. (2018). The Essential of Hate Speech. *IJET*.

Simonite, T. (2020, May 12). *Facebook's AI for Hate Speech Improves. How Much Is Unclear*. Retrieved from https://www.wired.com/story/facebook-ai-hate-speech-improves-unclear/

Solmone, S. (2018, May 3). *Regulate social media? It's a bit more complicated than that*. Retrieved from The Conversation: https://theconversation.com/regulate-social-media-its-a-bit-more-complicated-than-that-103797

TAŞ, E. (2017). YENİ MEDYADA NEFRET SÖYLEMİ. *e-Journal of New Media, 1*(1), 60-71. doi:10.17932/IAU.EJNM.m.25480200.2017.1/1.60-71

Taye, B. (2020, May 19). *Ethiopia's hate speech and disinformation law: the pros, the cons, and a mystery*. Retrieved from accessnow.org: https://www.accessnow.org/ethiopias-hate-speech-and-disinformation-law-the-pros-the-cons-and-a-mystery/

Traum, A. (2014). Contextualising the hate speech debate: the United States and South Africa. *The Comparative and International Law Journal of Southern Africa, 47*(1), 64-88. Retrieved June 12, 2020, from https://www.jstor.org/stable/24585817?seq=1

Tsesis, A. (2002). Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements. *Researchgate*. Retrieved June 7, 2020, from https://www.researchgate.net/publication/228321836_Destructive_Messages_How_Hate_Speech_Paves_the_Way_for_Harmful_Social_Movements

Tuba Ciftci, R. H. (2018). Hate speech on Facebook. *4th European Conference on Social Media Research.* At Mykolas Romeris University Vilnius, Lithuania. Retrieved June 12, 2020, from https://www.researchgate.net/publication/323512148_Hate_speech_on_Facebook

Udoh-Oshin, G. (2017). Hate Speech on the Internet: Crime or Free Speech? *Undergraduate Honors College Theses*. Retrieved June 13, 2020, from https://digitalcommons.liu.edu/cgi/viewcontent.cgi?article=1009&context=post_honors_theses

Ulusoy, D. Ç. (2013). A Comparative Study of the Freedom of Expression in Turkey and EU. *The Turkish Year Book of International Relations, 44*, 51 - 148 . doi:10.1501/Intrel_0000000287

US Department of State. (2019, March). *Sudan 2018 Human Rights Report.* Retrieved from https://www.state.gov/wp-content/uploads/2019/03/Sudan-2018.pdf

Walker, J. (2018, June 29). *Hate Speech and Freedom of Expression: Legal Boundaries in Canada*. Retrieved from Library of Parliament: https://lop.parl.ca/sites/PublicWebsite/default/en_CA/ResearchPublications/201825E#ftn2

Wang, Q. (2018, March). The Understanding of Absolute Right to Freedom of Expression in the Case of Hate Speech. *Graduate Theses and Dissertations*, pp. 1-100. Retrieved from http://scholarcommons.usf.edu/etd/7240

Wehbé, A. (2018). Increasing International Legal Protections for Freedom of Expression. *Notre Dame Journal of International & Comparative Law, 8*(2), 1-20. Retrieved

June 4, 2020, from https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=1085&context=ndjicl

Weiner, M. S. (2012, September 3). *Self-Governance and Free Speech*. Retrieved from Civil Liberties in the United States: https://uscivilliberties.org/themes/4466-self-governance-and-free-speech.html

Weinstein, I. H. (2009). *Extreme Speech and Democracy.* Oxford Scholarship Online. doi:10.1093/acprof:oso/9780199548781.001.0001

Weinstein, J. (2017). Hate Speech Bans, Democracy, and Political Legitimacy. *Constitutional Commentary, 32*(527), 527-587. Retrieved June 8, 2020, from https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1475&context=concomm

Yadav, A. (2018). Countering Hate Speech in India: Looking for answers beyond the Law. *ILI Law Review, II*(Winter Issue), 1-17.

Yanik, A. (2017). Fundamentals of Rising Hate Speech on Social Media . *Global Media Journal TR Edition, 8*(15), 364-383. Retrieved June 26, 2020, from https://www.researchgate.net/publication/320893358_Sosyal_Medyada_Yukselen_Nefret_Soyleminin_Temelleri

Yaraghi, N. (2018, September 21). *Regulating free speech on social media is dangerous and futile*. Retrieved from Brookings: https://www.brookings.edu/blog/techtank/2018/09/21/regulating-free-speech-on-social-media-is-dangerous-and-futile/