

İFADE ÖZGÜRLÜĞÜ TEORİSİ İŞİĞİNDA SOSYAL MEDYADA NEFRET SÖYLEMİNİ DÜZENLEMENİN ZORLUKLARI*

Biruk PAULOS**

Dr. Öğretim Üyesi Seydi ÇELİK***

ÖZ

Uluslararası insan hakları belgeleri herkesin düşünce ve ifade özgürlüğüne 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.

* Yayın Kuruluna Ulaştığı Tarih: 20/04/2021

Kabul Edildiği Tarih: 07/05/2021

** Dilla University, School of Law, Doktora Öğrencisi (PhD)

E-Mail: brookcanada2017@gmail.com

ORCID: 0000-0001-6059-972X

*** Kocaeli Üniversitesi, Hukuk Fakültesi, Genel Kamu Hukuku Anabilim Dalı

E-Mail: seydi.celik@kocaeli.edu.tr

ORCID: 0000-0003-1224-9877

THE CHALLENGES OF REGULATING HATE SPEECH ON SOCIAL MEDIA IN LIGHT OF THE THEORY OF FREEDOM OF EXPRESSION

ABSTRACT

Despite the fact that international human rights documents stipulate that everyone has the right to freedom of thought and expression, many nations nowadays have laws that restrict or ban some forms of social media expression, such as speech that provokes 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. Other scholars, on the other hand, contend that regulating hate speech on social media is crucial for protecting minority populations from the damages that such speech can inflict. Hate speech poses a difficult question that blurs the lines of free expression. 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.

INTRODUCTION

Hate speech on social media is a legal issue that has sparked debate on both an international and national scale.¹ Some scholars claim that hate speech is a form of free speech that deserves legal protection, while others say that hate speech should not be included in the definition of free speech.² For governments attempting to regulate hate speech, striking a balance between hate speeches on the one hand and freedom of expression on the other has proven difficult.³ Countries all over the world struggle to strike a balance between the freedom of expression required for a democratic society and the prevention of harm to individuals or minority groups as a result of hate speech.⁴ Survivors of massive massacres and other human rights crimes, particularly those belonging to minority groups and indigenous peoples, continue to be threatened in many parts of the world.⁵ The rising incidence of attacks against immigrants and other minorities has sparked new concerns about the link between online hate speech and actual violence.⁶ Hate speech and hate crimes are on the rise all around the world, owing to shifting political climates and rising social media violence.⁷ While hate speech on social media is on the rise in many nations, the hazards of violence, political repression, and mass murder directed against certain groups of people have appeared, or the possibility of such events occurring is

¹ J. Walker, *Hate Speech and Freedom of Expression: Legal Boundaries in Canada*, Legal and Social Affairs Division, 2018, June 29. Retrieved from Library of Parliament: https://lop.parl.ca/sites/PublicWebsite/default/en_CA/ResearchPublications/201825E#ftn2 (26.11.2020).

² See, F. Schauer, *İfade Özgürlüğü: Felsefi bir inceleme*, Çev. B. Seçilmişoğlu, Liberal Düşünce Topluluğu Yayınları, Ankara, 2002; See, O. Uygun, "Avrupa İnsan Hakları Sözleşmesi ve Türk Hukukunda İfade Özgürlüğünün Sınırlandırılması", *Türkiye'de İfade Özgürlüğü*, (Ed. T. Koçak, T. Doğan, Z. Kutluata) bgst yayınları, İstanbul, 2009, ss.15-67; See, K. Şahin, *İfade Özgürlüğü: Gerekçeleri ve Sınırları*, XII. Levha yayınları, İstanbul, 2009.

³ R. Apalara, "Striking a Balance: Freedom of Expression and the Prohibition of Hate Speech and Offensive Remarks", *Social Science Research Network (SSRN)*, 1-76, 2017, s.2. Retrieved June 7, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035410, (7.10.2020).

⁴ M. Y. Dalkılıç, "Nefret Söylemi ve İfade Özgürlüğü Tartışmaları Çerçevesinde İslamofobyası: L'Express Dergisi Örneği", *Medya ve Din Araştırmaları Dergisi*, 1(1), 89-109, 2018, s.90. Retrieved February 20, 2021, from <https://dergipark.org.tr/tr/download/article-file/496944>, (9.12.2020); Silvana Neshkovska, Z. T. *The Essential of Hate Speech*. IJET, 2018.

⁵ Cease Fire Centre for Civilian Rights, *Peoples under Threat 2019*, Minority Rights Group International, 2019.

⁶ V. Pejchal, "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), 2015, 58-74; Ayşe Kalav- Aliye Bilge Certel Fırat, "Amerikan Sosyal Medyasında Göçmen Karşıtlığı ve Dijital Nefret Söylemi: Twitter Özelinde Bir İnceleme", *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22, 2017, pp.2209-2222. Retrieved February 20, 2021, from <https://iibfdergi.sdu.edu.tr/assets/uploads/sites/352/files/yil-2017-cilt-22-sayi-kayfor15-yazi42-30122017.pdf> (9.12.2020)

⁷ C. George, *Ethical Journalism Network*, 2020, June 3. Retrieved from *Hate Speech: A Dilemma for Journalists the World Over*: <https://ethicaljournalismnetwork.org/resources/publications/ethics-in-the-news/hate-speech> (5.11.2020); Timur Demirbaş, "Nefret Söylemi ve Nefret Suçları", *D.E.Ü. Hukuk Fakültesi Dergisi*, pp.2693-2701, 2017. Retrieved February 20, 2021, from <https://hukuk.deu.edu.tr/wp-content/uploads/2017/11/3-TIMUR-DEMIRBAS.pdf> (19.12.2020).

increasing.⁸ While social media has benefited the global community in many ways, it has also been used to spread hate speech. Dehumanizing language and open incitement to mass murder have been used on social media platforms such as Facebook and Twitter in Myanmar, adding to the widespread targeting of the Muslim Rohingya minority.⁹ Following Ethiopia's sweeping democratic change in 2018, the spread of hate speech and disinformation on social media aimed at ethnic and religious minorities has resulted in widespread sectarian violence, mass murder, and millions of people being internally displaced.¹⁰ Online hate speech and provocation have fueled cycles of atrocities between Christians and Muslims in the Central African Republic (CAR) in recent years.¹¹ Hate speech against religious and ethnic minorities has penetrated Turkey's mainstream media, posing a threat to the country's social peace.¹² Hate speeches, which are the dominating news discourse in Turkish news media, have reinforced anti-Semitism, homophobia, and racism in the country, according to a research article.¹³ In Turkey, hate crimes targeting dissident groups are on the rise, and political hate speech could be a catalyst for this trend.¹⁴ Hate speech on social media has also contributed to xenophobia, discrimination, and marginalization of Syrian refugees in Turkey.¹⁵ Rumors shared over WhatsApp and hate speech has led to a series of gang attacks against Muslims in India, where social media

-
- ⁸ European Parliament. The Role Of Social Media In Ethnic Conflicts, 2018, November 6, Retrieved from Parliamentary questions: https://www.europarl.europa.eu/doceo/document/E-8-2018-005636_EN.html (17.8.2020).
- ⁹ E. Irving, "Suppressing Atrocity Speech on Social Media", Symposium on Non-State Actors and New Technologies in Atrocity Prevention. *Ajil Unbound*, 113, pp.256-261, 2018, p.257.
- ¹⁰ C. Maas, Ethiopia Parliament Passes Law Punishing Hate Speech, 2020, February 15. Retrieved from Jurist.org: <https://www.jurist.org/news/2020/02/ethiopia-parliament-passes-law-punishing-hate-speech/> (9.12.2020).
- ¹¹ OHCHR, Preventing Incitement to Hatred and Violence In the Central African Republic. 2019, May 29, Retrieved from United Nations Human Rights Office of the High Commissioner: <https://www.ohchr.org/EN/NewsEvents/Pages/PeacekeepersDay2019.aspx> (9.2.2021).
- ¹² H. E. Arcan, "Interrupted Social Peace: Hate Speech in Turkish Media", *The IAFOR Journal of Media, Communication and Film*, 1(1), 2013, 43-56; B. G. Dural, Othering Through Hate Speech: The Turkish-Islamist (V)AKIT Newspaper as a Case Study", *Turkish Studies*, 13(3), 2012, pp. 489-507. doi:<http://dx.doi.org/10.1080/14683849.2012.715482>
- ¹³ Ş. Balcı- A. Gölcü, "Hate Speech In Turkish And Israel Press: A Comparison Analysis Of Mavi Marmara Raid", In 9 th International symposium communication in the millennium, pp. 272-290, San Diego, San Diego State University, 2011 May 22-25. https://www.researchgate.net/publication/289506459_Hate_Speech_in_Turkish_and_Israel_Press_A_Comparison_Analysis_of_Mavi_Marmara_Event (9.2.2021); Hülya Öztekin, "Yeni Medyada Nefret Söylemi: Ekşi Sözlük Örneği", *Uluslararası Sosyal Araştırmalar Dergisi*, 8(38), pp. 925-936, 2015, p.926 et al. Retrieved February 20, 2021, from https://www.sosyalarastirmalar.com/cilt8/sayi38pdf/6iksisat_kamu_isletme/oztekinhulya.pdf (9.9.2020)
- ¹⁴ B. Perry- D. Akça- F. Karakuş- M. F. Baştuğ, "Planting Hate Speech to Harvest Hatred: How Does Political Hate Speech Fuel Hate Crimes in Turkey?" *International Journal for Crime Justice and Social Democracy*. 9(2), Advance online publication, pp. 1-17, June 2020.
- ¹⁵ T. Yazıcı, "Yeni Medyanın Nefret Dili: Suriyeli Mültecilerle İlgili Ekşi Sözlük Örneği", *Global Media Journal TR Edition*, 7(13), 2016, pp. 115-136. Retrieved February 19, 2021, from <https://globalmediajournaltr.yeditepe.edu.tr/sites/default/files/T%C3%BClay%20YAZICI.pdf> (16.2.2021).

has depicted an aggressive and privileged Hindu nationalism.¹⁶ Similarly, rumors on social media spurred anti-Muslim gang violence in Sri Lanka last year, which were reignited following ISIS's Easter Sunday church and hotel bombings in April 2019.¹⁷ Al-Shabaab has utilized Twitter and Facebook to spread hate speech propaganda and recruit supporters in Somalia, where mobile phone usage has steadily expanded in recent years.¹⁸ In Sudan, social media, which has been outlawed but is nonetheless available through virtual private networks (VPNs), has played a key role in recruiting and organizing demonstrators around an iconic video of a young female activist that has gone viral around the world.¹⁹ The perpetrators of recent white nationalist crimes in the United States have frequented online racist communities and used social media to publicize their crimes.²⁰ In June 2015, the Charleston church shooter, who killed nine black clergy and congregants, began a process of self-discovery that led him to conclude that achieving white supremacy required murder.²¹ Hate speech on social media in Germany has resulted in significant violence against some minorities, particularly immigrants and refugees.²² In March 2019, a far-right terrorist attack in Christchurch, New Zealand, was carried out using social media.²³ Radio, print media, SMS messaging, and even lyrics were once used to propagate nasty, threatening, and biased messages.²⁴ Today, social media platforms such as Facebook, Twitter, YouTube, Instagram, and a variety of internet news sites are being utilized to promote hate speech in novel ways.²⁵ Many legal issues have arisen as a result of hate speech on social media, including what constitutes hate speech, who should control social media, and to what degree hate speech should be regulated. While some countries ban all forms of hate speech and have laws governing social media, others allow social media

¹⁶ P. Thompson- R. Itaoui- H. Bazian, "Islamophobia in India: Stoking Bigotry", Berkeley: Islamophobia Studies Center, 2019. Retrieved March 21, 2020, from <https://irdproject.com/wp-content/uploads/2019/04/Islamophobia-in-India-Web-Spread.pdf> (9.12.2020).

¹⁷ AFP, Hate speech in focus again: Facebook apologises for role in 2018 Sri Lanka unrest, 2020, May 13. 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> (2.10.2020).

¹⁸ HRW, Human Rights Watch. 2018, Retrieved from Somalia: Events of 2018: <https://www.hrw.org/world-report/2019/country-chapters/somalia> (9.12.2020)

¹⁹ US Department of State, Sudan 2018 Human Rights Report, 2019, March. Retrieved from <https://www.state.gov/wp-content/uploads/2019/03/Sudan-2018.pdf> (15.10.2020).

²⁰ ALA, American Library Association. Retrieved from Hate Speech and Hate Crime, 2017, December. <http://www.ala.org/advocacy/intfreedom/hate> (9.12.2020).

²¹ Z. Laub, "Council on Foreign Relations. Retrieved from Hate Speech on Social Media", Global Comparisons, 2019, June 7: <https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons> (12.9.2020).

²² K. M. Schwarz, "Fanning the Flames of Hate: Social Media and Hate Crime", SSRN Electronic Journal, pp.1-85, 2019, November 3.

²³ L. Matthew- P. B. Williams, "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), pp. 93-117, 2019, p.94.

²⁴ A. Tsesis, "Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements", 2002. Researchgate. Retrieved June 7, 2020, from https://www.researchgate.net/publication/228321836_Destructive_Messages_How_Hate_Speech_Paves_the_Way_for_Harmful_Social_Movements (9.9.2020).

²⁵ S. Neshkovska- Z. Trajkova, "The Essential of Hate Speech", *IJET*, pp.71-80, 2018. https://www.researchgate.net/publication/329874520_The_Essential_of_Hate_Speech, (9.12.2020).

corporations to establish their own regulatory rules to control hate speech on their platforms.²⁶ However, in many nations, the question of what constitutes hate speech and to what degree it should be restricted remains a perplexing legal matter. Excessive regulation of hate speech on social media jeopardizes freedom of expression and silences political dissent, whereas lax control leads to violence and criminality.²⁷ When hate speech on social media is unchecked, it can lead to actual crimes on the ground, putting some people in danger. On the other hand, in those countries where the government severely censors online hate speech, some individuals are outraged, claiming that such laws are intended to stifle political dissidents and oppress ethnic or religious minorities. Thus, in light of the theory of freedom of expression, this article aims to investigate the issues of regulating hate speech on social media.

I. THEORETICAL ANALYSIS OF 'FREEDOM OF EXPRESSION' AND 'HATE SPEECH'

A- The Theory of Freedom of Expression

1. *The Truth Theory*

Public debate, according to John Stuart Mill, can lead to the discovery of truth.²⁸ As a result, society should not impose any restrictions on the exercise of free public debate in order to avoid impeding the pursuit of truth. In other words, the government cannot be the sole arbiter of all facts when it comes to determining the best policies for its citizens. People can only come up with such sound policies through vigorous public debate. In line with this view, the spread of false views should also be tolerated. It is important to remember that, as Daniel Jacobson²⁹ points out, Mill does not intend to restrict freedom of expression simply because someone is harmed by the statements of others. Mill contrasts between legitimate and illegitimate damages, claiming that hate speech can only be controlled or regulated if it results in a direct and clear infringement.³⁰ As a result, Mill argues that speech cannot be restricted just to avert injury. Only when there is a blatant and evident violation of rights does it make sense to restrict speech. Conversely, if the speech does not cause a direct and obvious violation of others' rights, it should not be banned or regulated. As a result, in

²⁶ S. Solmone, "Regulate social media? It's a bit more complicated than that", 2018, May 3. Retrieved from The Conversation <https://theconversation.com/regulate-social-media-its-a-bit-more-complicated-than-that-103797> (18.10.2020).

²⁷ See, Tthesis, 2002.

²⁸ K. Appiagyei-Atua, "A Review of Theories of Expression in the Context of the Development Argument", University of Ghana Law Journal, XXIV, pp. 197-227, January 2005. 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 (9.12.2020)

²⁹ D. Jacobson, "Freedom of Speech under Assault on Campus", Policy Analysis No. 796, 2016, August 30. Retrieved from CATO Institute: <https://www.cato.org/publications/policy-analysis/freedom-speech-under-assault-campus> (9.2.2021).

³⁰ K. N. Hylton, "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), 1996, pp. 35-57. Retrieved June 8, 2020, from <https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1377&context=roundtable> (23.2.2021).

light of this argument, hate speech can be considered free speech as long as it does not infringe on others' rights.

2. The Marketplace of Ideas

The marketplace of ideas is an analogy to the economic concept of the free market that justifies freedom of expression.³¹ This rationale claims that freedom of expression is crucial because better ideas will eventually win out in a marketplace of ideas due to competition. Referring to the public realm as a “marketplace of ideas” is a common metaphor used to defend a legal right to freedom of expression.³² The idea market raises social value by making truth or reality more accessible. However, arguments based on this logic are more effective in protecting statements that can be true or false than statements that cannot.³³ This is significant since many types of argumentation, such as art, are unrelated to any truth value, and much less so to obscenity and pornography.³⁴ As a result, according to this theory, hate speech must pass a social acceptability or rejection test in the marketplace of ideas. Furthermore, some proponents of this theory advocate the possibility of regulating hate speech by comparing it to market flaws regulation.³⁵

3. The Democracy Theory

Alexander Meiklejohn was the first political theorist to find a genuinely democratic framework to defend freedom of expression.³⁶ He believes that freedom of expression and press are the most inclusive and thus the best forms of public discourse. As a result, Meiklejohn considers freedom of expression to be the foundation or most essential requirement of the democratic process, placing it above practically all other rights. As a result, self-government and participation in the democratic process provide a strong rationale for freedom of expression. Political speech, according to this argument, is at the top of the hierarchy of freedom of expression. However, the logic of democratic self-government only protects some manifestations, while others (such as pornography) are excluded from these parameters. The self-government reasoning plainly implies that democratic theories and justifications justify freedom of expression. One argument against hate speech

³¹ D. Schultz, “Marketplace of Ideas”, *The First Amendment Encyclopedia*, June 2017. Retrieved from <https://www.mtsu.edu/first-amendment/article/999/marketplace-of-ideas> (9.11.2020).

³² S. Lee, “Hate Speech in the Marketplace of Ideas”, *The Philosophical Foundations of Law and Justice*, 13-26, May 2010, s. 14. doi:10.1007/978-90-481-8999-1_2

³³ C. Lombardi, “The Illusion of a “Marketplace of Ideas” and the Right to Truth”, *American Affairs*, III(1), Spring 2019. Retrieved from *American Affairs*: <https://americanaffairsjournal.org/2019/02/the-illusion-of-a-marketplace-of-ideas-and-the-right-to-truth/> (2.2.2021).

³⁴ R. P. Macedo, “Freedom of Expression: What Lessons Should We Learn From US Experience?”, *Revista Direito GV*, 13(1), pp. 274-302, April 2017. doi:<http://dx.doi.org/10.1590/2317-6172201711>

³⁵ Lee, 2010, p. 17.

³⁶ JMC, Jack Miller Center. Retrieved June 4, 2020, from Alexander Meiklejohn: Free Speech and its Relation to Self-government, 2020, June 4. <https://jackmillercenter.org/cd-resources/alexander-meiklejohn-free-speech-relation-self-government/> (15.2.2021).

prohibitions is that they damage the state's democratic legitimacy.³⁷ A counter-argument contends that hate speech undermines the democratic process and, as a result, should be limited to the defense of democracy.³⁸ Hate speech can undermine the legitimacy of a process when equal political standing is a requirement for legitimate policymaking. As a result, if hate speech is compatible with logic, rationality, and democratic norms, it can be considered freedom of expression, according to proponents of the notion of self-government (democracy). Freedom of expression, according to this view, makes the government responsible and responsive. As a result, the theory of democracy indicates that, in a democratic society, measures other than legislative prohibitions can be used to mitigate the detrimental repercussions of hate speech.³⁹ It is preferable to compensate for the harm caused by hate speech by other means than than controlling it.

4. *Autonomy Theory*

Freedom of expression, according to this view, is necessary for self-realization and development.⁴⁰ The key notion is that the government should not impose any limits on speech on the basis of the presumption that speech only leads to the creation of negative views or causes negative acts as a result of those ideas.⁴¹ This theory promotes freedom of expression on the basis of the notion that it has intrinsic value. Being autonomous includes the ability to freely express one's opinions. Individual autonomy, in this view, is important regardless of whether or not it leads to positive outcomes, i.e. whether or not individual autonomy has external value. Part of what it means to be autonomous is the ability to express oneself freely. This idea contends that, like autonomy, freedom of expression is essentially necessary. There appear to be two sorts of harmful speech that should not be criminalized. Speech cannot be prohibited only because it has one of the following negative consequences: First, if it leads to incorrect beliefs in the audience; and second, if it has negative effects by leading individuals to believe (or increase their predisposition to believe) that they should engage in destructive behavior.⁴² There are two reasons why freedom of expression is important in the pursuit of autonomy: self-reflection and talent.⁴³ First, when people's freedom of expression is restricted, they lose valuable information about the decisions they make as well as the tools to question their beliefs and desires. Second, because many actions are carried out through verbal, written, or symbolic communication, restrictions on freedom of expression prevent people from exercising

³⁷ J. Weinstein, "Hate Speech Bans, Democracy, and Political Legitimacy", *Constitutional Commentary*, 32(527), 527-587, 2017, p. 528. Retrieved June 8, 2020. from <https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1475&context=concomm> (28.8.2020).

³⁸ A. Reid, "Does Regulating Hate Speech Undermine Democratic Legitimacy? A Cautious 'No'", *Res Publica*, 26, 181-199, 2020, p. 182 et al. doi:<https://doi.org/10.1007/s11158-019-09431-6>

³⁹ Weinstein, 2017, p. 530.

⁴⁰ S. M. Oltmann, "Intellectual Freedom and Freedom of Speech: Three Theoretical Perspectives", *Information Science Faculty Publications*, 153-171, 2016. Retrieved June 3, 2020, from https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1032&context=slis_facpub (19.9.2020).

⁴¹ T. Scanlon, "A theory of free expression", *Philosophy and Public Affairs*, 1 (2): pp. 204-226, 1972.

⁴² Scanlon, 1972, p. 204 et al.

⁴³ F. Peonidis, "Freedom of Expression, Autonomy, and Defamation", *Law and Philosophy*, 17(1), pp. 1-17, 1998. doi: 10.1023/A:1005850325859

their autonomy. The government is responsible for creating conditions in which people can freely make their own decisions about their lives and realize their own potential.⁴⁴ As John Stuart Mill and Scanlon argued, the notion of autonomy explains hate speech as freedom of expression unless it defies the ethic of injury. Thus, the self-actualization argument for free expression provides a strong foundation for the preservation of all types of speech, including political speech, literature and art, as well as extreme, non-intellectual forms of expression like nudism, pornography or vulgarity, and commercial speech.⁴⁵ While the autonomy theory says that freedom of expression would be fully protected by the law in order to maintain individual autonomy, some scholars think that hate speech is destructive and should be restricted. If the harm produced by hate speech surpasses the restriction on hate speech, which significantly contributes to the occurrence of violence, the logic of autonomy will not be convincing.⁴⁶

B- The Definition Of 'Hate Speech'

Many scholars have attempted to define hate speech in recent years, particularly in relation to freedom of expression on social media.⁴⁷ Hate speech has various nomenclatures: 'incitement to hatred';⁴⁸ 'antagonistic speech';⁴⁹ 'cyber hate';⁵⁰ and 'inflammatory rhetoric'.⁵¹ Despite the many nomenclatures, for the purposes of this article, the term "hate speech" will be used. One of the reasons why hate speech is difficult to define is that it encompasses a wide range of expressions. For example, distinguishing between hate speech and discriminatory language that humiliates people based on a shared identity; cliché generalization based on a common identification of people, whatever the intent; dangerous speech that provokes violence (hate crime); and symbolic expression, such as Swastikas, redesigned emojis, and memes, is extremely difficult.⁵² As a result, there is no agreement on what constitutes

⁴⁴ Macedo, 2017, p. 274 et al.

⁴⁵ M. S. Weiner, "Self-Governance and Free Speech", Retrieved from Civil Liberties in the United States, 2012, September 3. <https://uscivil liberties.org/themes/4466-self-governance-and-free-speech.html> (9.11.2020).

⁴⁶ See, I. Hare- J. Weinstein, *Extreme Speech and Democracy*, Oxford Scholarship Online, 2009. doi:10.1093/acprof:oso/9780199548781.001.0001

⁴⁷ See, for example, T. Gelashvili, "Hate Speech on Social Media: Implications of Private Regulation and Governance Gaps", JAMM07 Master Thesis. Lund University, Faculty of Law, 2018.

⁴⁸ See, OSCE-ODIHR, *Incitement to Hatred vs. Freedom of Expression: Challenges of combating hate crimes motivated by hate on the Internet*, 2010. Warsaw: Report of the OSCE-ODIHR Expert Meeting, p. 2. Retrieved March 18, 2020, from <https://www.osce.org/odihr/68750?download=true> (9.9.2020).

⁴⁹ See, R. Meza- H. O. Vincze- A. Mogos, "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), pp. 26-50, 2019.

⁵⁰ M. Roiha- O. Jubany, "Backgrounds, Experiences and Responses to Online Hate Speech: A Comparative Cross-Country Analysis", *Words are Weapons. Preventing Redressing Inhibiting Hate Speech in New Media*, 2016, p.5 et al.

⁵¹ Neshkovska- Trajkova, 2018.

⁵² J. Donovan, "Navigating the Tech Stack: When, Where and How Should We Moderate Content?" 2019, October 28. Retrieved from Internet Governance, Platform Governance: https://www.cigionline.org/articles/navigating-tech-stack-when-where-and-how-should-we-moderate-content?gclid=CjwKCAjw8pH3BRAXEiwA1pvMsYYWtF-G2eGsXdOFFdGvSfJt1zQjOjIoybYETAYXfZLihmXNkjAGRoC-SsQAvD_BwE (11.11.2020).

“hate speech”. The focus of this article does not include the search for the perfect definition of the word “hate speech”. However, in order to provide a proper analysis, it is necessary to establish a working definition of hate speech in this article. Hate speech is typically defined as a statement that disparages a person or group of people based on a variety of characteristics, including but not limited to race, ethnicity, sexual orientation, gender, religion, sexual orientation, caste, gender identity, serious disease or disability, and nationality.⁵³ Hate speech can occur in any kind of media, whether online or offline. Hate speech on the internet is described as any technology-mediated speeches or digital communications that harass, discriminate, or degrade a person based on a group-defining trait such as race, ethnicity, gender, nationality, sexual orientation, religion, age, handicap, or others.⁵⁴ Anti-Semitic, racist, prejudiced, extremist, or terrorist messages or information can also be propagated through online hate speech.⁵⁵ On social media, hate speech is spread through harsh discourses that include racism, xenophobia, discrimination, misogyny, and homosexuality.⁵⁶ Speech that is humiliating, insulting, discriminatory, inflammatory, or even incites or encourages violence or provokes violent response is considered hate speech.⁵⁷ 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”.

C- Current Academic Debates On ‘Hate Speech’

There are numerous arguments for and against regulating hate speech on social media. There are three types of arguments: liberalist (post structuralist) arguments, progressivist arguments, and multiculturalist debates. In reality, these arguments do not contradict each other. The heart of these debates is striking a balance between protecting free expression and regulating hate speech.⁵⁸ Liberals argue that freedom

⁵³ See, A. Al-Hassan- H. Al-Dossari, “Detection of Hate Speech in Social Networks: A Survey On Multilingual Corpus”, *Computer Science & Information Technology (CS & IT)*, 83-100, 2019. doi: 10.5121/csit.2019.90208; Meza, et al., 2019; See, T. Quinn, “Monitoring Hate Speech”, *Challenges and Strategies. HateBase 2018*; See, L. Silva- M. Mondal- D. Correa- et al, “Analyzing the Targets of Hate in Online Social Media”, *AAAI ICWSM*, 1-4: 687-690, 2016.

⁵⁴ E. P. Melhuish, “Online Hate Speech: A Survey on Personal Experiences and Exposure among Adult New Zealanders”, *Netsafe*, 2018. Retrieved March 21, 2020, from <https://www.netsafe.org.nz/wp-content/uploads/2019/11/onlinehatespeechsurvey-survey-2018.pdf> (9.10.2020).

⁵⁵ C. Blaya- C. Audrin, “Toward an Understanding of the Characteristics of Secondary School Cyberhate Perpetrators”, *Frontiers in Education*, 4(48), June 2019. doi:<https://doi.org/10.3389/feduc.2019.00046>

⁵⁶ See, A. Yanık, “Fundamentals of Rising Hate Speech on Social Media” (*Sosyal Medyada Yükselen Nefret Söyleminin Temelleri*), *Global Media Journal TR Edition*, 8(15), pp. 364-383, 2017. Retrieved June 26, 2020; See, Hünler, O. S. (2019). *Popülist Sağ ve Sosyal Medyada Nefret Söylemi. Eleştirel Psikoloji Bülteni*, 8, pp. 6-12. Retrieved February 20, 2021, p.7 et al.

⁵⁷ A. Yadav, “Countering Hate Speech in India: Looking for answers beyond the Law”, *ILI Law Review*, II (Winter Issue), pp. 1-17, 2018.

⁵⁸ D. Jacobson, 2016, August 30.

of expression can never be an absolute right due to censorship.⁵⁹ Although they believe that absolute freedom of expression cannot be achieved in all circumstances, they believe that both liberal and non-liberal ideas should be tolerated. They advocate for the adoption of a free and neutral doctrine to promote the right to free expression, even if it means extending legal protection to non-liberal ideas and expressions.⁶⁰ Thus, in accordance with this argument, it is sometimes appropriate to outlaw and punish certain expressions for good reasons within the confines of the law.⁶¹ If freedom of expression means the freedom to express any opinion or idea, then speech acts such as incitement to violence and other derogatory words should be prohibited.⁶² As a result, academics 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,⁶³ others see such restrictions as necessary to counter the harm, discrimination and subjugation that may result from hate speech.⁶⁴ On the other end of the spectrum, progressivists argue that equality or social justice should take precedence over freedom of expression, and that equality should be achieved even at the expense of freedom of expression.⁶⁵ As a result, collective rights should take precedence over individual liberty, and the government should take steps to control hate speech in order to protect society's collective rights.⁶⁶ However, most

⁵⁹ N. Alkiviadou, "Hate speech on social media networks: towards a regulatory framework?" *Information & Communications Technology Law*, 28(1): pp. 1-17, 2018. doi:10.1080/13600834.2018.1494417

⁶⁰ A. Romero, "The Line Between Free Speech and Hate Speech", Researchgate, 2016. (www.theintelligencer.com, Monday, November 28, 2016), Retrieved June 13, 2020. https://www.researchgate.net/publication/311064494_The_line_between_free_speech_and_hate_speech (9.12.2020).

⁶¹ N. Yaraghi, "Regulating Free Speech On Social Media is Dangerous and Futile", Techtank, 2018, September 21. Retrieved from Brookings: <https://www.brookings.edu/blog/techtank/2018/09/21/regulating-free-speech-on-social-media-is-dangerous-and-futile/> (6.6.2020); See, Ö. Ağtaş- A. Özkazanç, "Judith Butler'in Nefret Söylemi Eleştirisi: Dildeki Performatif ve Yaralayıcı Dil", *Fe Dergi: Feminist Eleştiri*, 10(1), 1-12, 2018. Retrieved February 18, 2021, from https://www.researchgate.net/publication/323448139_Judith_Butler'in_Nefret_Soylemi_Elestirisi_Dildeki_Performatif_ve_Yaralayici_Dil (3.10.2020).

⁶² A. G. Juanatey, "Hate Speech in Social Media: A State-of-the-Art Review", *Self-Controll on Networks!*, 2016, December 12. Retrieved from <https://www.uclg-cisd.org/sites/default/files/Hate%20Speech%20A%20State%20of%20the%20Art.pdf> (4.1.2021).

⁶³ J. Samples, "Why the Government Should Not Regulate Content Moderation of Social Media", *Cato Institute Policy Analysis*, No.865, 2019, April 9. Retrieved from Cato Institute: <https://www.cato.org/publications/policy-analysis/why-government-should-not-regulate-content-moderation-social-media> (2.12.2020).

⁶⁴ A. Altman, "Liberalism and Campus Hate Speech: A Philosophical Examination", *An International Journal of Social, Political, and Legal Philosophy*, 103(2), pp 302-317, 1993, p. 303 et al. Retrieved June 3, 2020, from <https://www.journals.uchicago.edu/doi/abs/10.1086/293497?journalCode=et> (9.12.2020).

⁶⁵ C. Demaske, "Social Justice, Recognition Theory and the First Amendment: A New Approach to Hate Speech Restriction". *Journal of Communication Law and Policy*, 24(3), pp.347-401, 2019. doi:<https://doi.org/10.1080/10811680.2019.1627800>

⁶⁶ C. Kaniklidis, "Free Speech, Hate Speech and Principles of Community: The Case Against Free Speech Absolutism", *No Surrender Breast Cancer Foundation (NSBCF)*, 2015. Researchgate. Retrieved June 12, 2020, from

progressivists oppose violence, obfuscation, and authoritarianism, despite the fact that such methods could be used to achieve social equality.⁶⁷ Multiculturalists, on the other hand, argue that any viewpoint or expression that incites violence against minorities or marginalized groups of people should not be legally protected as free expression.⁶⁸ They argue that the restriction of hate speech is necessary to safeguard minorities. They agree with postmodernists that censorship is a necessary tool for avoiding harm to society. As a result, only minorities or marginalized persons are deemed vulnerable to hate speech, according to this viewpoint.⁶⁹ The literature on hate speech is replete with social justice considerations. As a result, the vast majority of people are wary of freedom of expression and intolerance for differences and disputes.⁷⁰ A cursory study of the literature on hate speech and its relationship to freedom of expression indicates a wide range of conflicting viewpoints. At one extreme, some scholars consider hate speech as freedom of expression,⁷¹ while others at the other end consider hate speech as a highly harmful phenomenon that should be avoided or punished by law.⁷² 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.⁷³ According to this spectrum, freedom of expression is distinguished from words or statements that promote violence or have adverse consequences.⁷⁴ The right to free expression does not include the right to express anything that is harmful to others.⁷⁵ Hate speech of any kind, while potentially harmful, has no immediate negative consequences. As a

https://www.researchgate.net/publication/274835999_Free_Speech_Hate_Speech_and_Principles_of_Community_The_Case_Against_Free_Speech_Absolutism (30.10.2020).

- ⁶⁷ L. M. Seidman, "Can Free Speech Be Progressive?", Georgetown University Law Center, 2018. Retrieved June 3, 2020, from <https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=3056&context=facpub> (9.11.2020).
- ⁶⁸ D. Chong, "Free Speech and Multiculturalism In and Out of the Academy", *Political Psychology*, 27(2), pp. 29-54, 2006, p. 31 et al. doi: <https://doi.org/10.1111/j.1467-9221.2006.00448.x> (2.12.2020).
- ⁶⁹ S. Benesch, "Defining and Diminishing Hate Speech", July 2014. 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> (9.12.2020); J. Bell, "Restraining the Heartless: Racist Speech and Minority Rights", *Indiana Law Journal*, 84, 963-979, 2009. Retrieved June 13, 2020, from http://ilj.law.indiana.edu/articles/84/84_3_Bell.pdf (9.12.2020).
- ⁷⁰ A. Traum, "Contextualising the Hate Speech Debate: the United States and South Africa", *The Comparative and International Law Journal of Southern Africa*, 47(1), 64-88, 2014. Retrieved June 12, 2020, from <https://www.jstor.org/stable/24585817?seq=1> (9.12.2019).
- ⁷¹ M. Quent, "Opinion: The Clear Line Between Hate Speech And Free Speech", 2019, November 24. Retrieved from DW Made for Minds: <https://www.dw.com/en/opinion-the-clear-line-between-hate-speech-and-free-speech/a-51393218> (12.9.2020).
- ⁷² A. Scheffler, "The Inherent Danger of Hate Speech Legislation A Case Study from Rwanda and Kenya on the Failure of a Preventative Measure", Friedrich-Ebert-Stiftung, Namibia, 2015. Retrieved from Friedrich-Ebert-Stiftung (FES): <https://library.fes.de/pdf-files/bueros/africa-media/12462.pdf> (1.12.2020).
- ⁷³ Neshkovska, 2018.
- ⁷⁴ K. R. Holmes, "The Origins of "Hate Speech"", 2018, October 22. Retrieved from The Heritage Foundation,: <https://www.heritage.org/civil-society/commentary/the-origins-hate-speech> (13.2.2021).
- ⁷⁵ See, Romero, 2016.

result, when individuals' actions cause harm to or affect the rights or interests of others, their absolute free speech should be limited.⁷⁶

II. LEGAL REGIMES ON “FREEDOM OF EXPRESSION” AND “HATE SPEECH”

A- International Human Rights Instruments

Even though the idea of freedom of expression appears in major international human rights instruments⁷⁷, It is specifically stated as a fundamental right in the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR). The preface to the UDHR emphasizes freedom of expression.⁷⁸ 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.⁷⁹ In a similar approach to the UDHR, the ICCPR guarantees freedom of expression. According to the ICCPR, freedom of speech encompasses the right to seek, receive, and disseminate ideas or information in any form or medium.⁸⁰ Freedom of expression can be exercised by any media, both offline and online, under this law. Despite the fact that freedom of expression is clearly stated as the most essential aspect of human rights and fundamental freedoms, it is subject to various limitations. Accordingly, freedom of expression can be restricted by law if deemed necessary.⁸¹ Respect for others' rights or reputations; national security, public order, public health, or morals are all reasons for such limits. The ICCPR appears to allow little room for legal restrictions on freedom of expression. The question is whether hate speech can be used as a justification for curtailing freedom of expression. Only the ICCPR incorporates rules against hate speech and incitement on the subject at the international level. Any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence is prohibited by law under this covenant.⁸² This article clearly states that the advocacy of hate against national, racial

⁷⁶ Q. Wang, “The Understanding of Absolute Right to Freedom of Expression in the Case of Hate Speech. Graduate Theses and Dissertations”, pp. 1-100, March 2018, p.10 et al. Retrieved from <http://scholarcommons.usf.edu/etd/7240> (4.11.2020).

⁷⁷ Major international human rights documents express the principle of freedom of expression. Universal Declaration of Human Rights; International Covenant on Civil and Political Rights; International Covenant on Economic, Social and Cultural Rights; International Convention on the Elimination of All Forms of Racial Discrimination; Convention on the Elimination of All Forms of Discrimination against Women; ILO Convention No. 135, Workers' Representatives Convention; General Comment 11 (Article 20) of the Human Rights Committee (CCPR/C/21/Rev.1 of 19 May 1989); General Comment 10 (Article 19) 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 all encompass freedom of expression.

⁷⁸ 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”.

⁷⁹ Refer article 19 of UDHR.

⁸⁰ 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”.

⁸¹ Refer Article 19 (3) (a) and (b) of the ICCPR.

⁸² Refer Article 20(2) of ICCPR.

or religious identities with the aim of inciting discrimination, hostility or violence will be the basis for prohibiting such acts by law. As a result, states have the option or obligation of enacting domestic legislation that prohibits hate speech in accordance with international standards. Any expression that constitutes discrimination, hatred, or violence against national, racial, or religious groups is considered hate speech under the ICCPR. 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 loopholes in the definition of hate speech, the ICCPR appears to have given a legal foundation for states to prohibit hate speech through legislation. However, the question of who should regulate hate speech on social media remains unanswered: the government, social media firms, or any other international organization? This issue has yet to be adequately addressed. If hate speech is deemed an exception to freedom of expression, it is self-evident that the government can act to lawfully restrict such damaging online and offline discourse.⁸³ The obligation to regulate hate speech is aimed at states that are required by domestic law to prohibit such behavior, rather than at social media companies, which are not directly bound by human rights treaties as local institutional actors.⁸⁴ Despite the fact that businesses do not have international legal standing, efforts have been made in recent years to subject businesses to human rights law standards.⁸⁵ For instance, the UN Guiding Principles on Business and Human Rights (UN GPBHR)⁸⁶ 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 activities⁸⁷ and the responsibility to diligently detect potential harms that can ensue from the activities of the company.⁸⁸ The normative working environment of social media companies has changed significantly after the adoption of the UN Guidelines. The UN Guideline utilizes the term “responsibility” rather than “duty” to express the notion that human rights are not a set of legal obligations for businesses, but rather a standard of conduct. Furthermore, there is no agreement in place that requires social media corporations to respect human rights as part of their legal obligations under international law.⁸⁹ As a result, under international human rights law, social media companies have no legal obligation to suppress hate speech and incitement to violence on their platforms. However, the UN Guiding Principles state that all internet intermediaries (including social media companies) share the responsibility to respect human rights. Therefore, governments can pass legislation in

⁸³ See, Gelashvili, 2018.

⁸⁴ V. C. Brannon, “Free Speech and the Regulation of Social Media Content”, Congressional Research Service, 2019. Retrieved from <https://crsreports.congress.gov> (9.12.2020).

⁸⁵ Irving, 2018, p. 257.

⁸⁶ 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).

⁸⁷ Id. Principle 13.

⁸⁸ Id. Principle 17.

⁸⁹ See the work of the Open-ended intergovernmental working group on multinational corporations and other business enterprises with respect to human rights for more information on this endeavor.

their territories to ensure that social media corporations follow human rights rules when it comes to hate speech.

B- Regional Human Rights Instruments

A number of regional developments are pertinent to the current debate. The right to freedom of expression is guaranteed by the 'African Charter on Human and Peoples' Rights (the Banjul Charter)'.⁹⁰ The Resolution on the Adoption of the Declaration of Principles on Freedom of Expression in Africa was released by the African Commission on Human and Peoples' Rights (ACHPR) in 2002.⁹¹ The Declaration of Principles reaffirms Article 9 of the Banjul Charter, which recognizes freedom of expression as two rights: "the right to receive information" and "the right to express and disseminate one's opinions within the law," in response to the ACHPR's concerns. The Declaration of Principles was largely based on the UDHR, ICCPR, and Banjul Charter's provisions on freedom of expression. The Declaration of Principles was unique in that it emphasized the value of oral traditions and freedom of expression in African culture. According to the Declaration of Principles, freedom of expression is "a fundamental and inalienable human right and an indispensable component of democracy."⁹² It further stated that any restrictions on freedom of expression must be enacted by legislation, serve a legitimate purpose, and be essential in a democratic society.⁹³ Although the Declaration states that restrictions on freedom of expression may be imposed, it does not state why such limits may be applied. In 2014, the African Union adopted the Malabo Protocol⁹⁴, which will create an international criminal law section within the African Court of Justice and Human Rights⁹⁵. This Court has the authority to prosecute corporations for a variety of international crimes, including crimes against humanity and genocide, both of which, in certain situations, include hate speech and incitement.⁹⁶ According to the Malabo Protocol, a company's intent to commit a crime can be proved by demonstrating that it is the company's policy to commit the crime in question. It is theoretically possible to prosecute social media corporations for illegal materials posted on their platforms.⁹⁷ However, because the senders of illicit materials are third parties rather than social media firms, proving that the company's policy is "to perform the criminal act" will be difficult. It is enough not to remove the problematic content even with the knowledge that the content is illegal. Therefore, the Malabo Protocol is ineffective in persuading social media corporations to remove hate speech and incitement from their platforms.

⁹⁰ See article 9 of the Charter.

⁹¹ Refer to the African Commission on Human and Peoples' Rights' Declaration of Principles on Freedom of Expression in Africa, 32nd Session, Banjul, The Gambia, 17 - 23 October 2002.

⁹² See article I (1) of the Declaration of Principles.

⁹³ Article II (2) of the Declaration of Principles.

⁹⁴ June 27, 2014 Protocol on Amendments to the Protocol on the Statute of the African Court of Justice and Human Rights.

⁹⁵ Protocol on the Statute of the African Court of Justice and Human Rights, July 1, 2008.

⁹⁶ Article 46 C of the Malabo Protocol provides for corporate jurisdiction.

⁹⁷ To enter into force, the Malabo Protocol requires fifteen ratifications. More than fifteen states had signed it as of March 2020, but none had ratified it.

In Europe, the European Human Rights Convention,⁹⁸ like the ICCPR, guarantees the right to freedom of expression. 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 and for the protection of judicial independence are additional grounds for limiting the right under the European Convention⁹⁹. Hate speech is not expressly mentioned as a reason for restricting freedom of expression in the terms of the Convention. The Amsterdam Recommendations on Freedom of the Media and the Internet were endorsed or published by the Organization for Security and Co-operation in Europe (OSCE) in 2003. It suggests a range of policies addressing a variety of issues, including freedom of expression.¹⁰⁰ Importantly, the first recommendation under freedom of speech stressed that there is a balance between free flow of information and social media misuse, and that illicit materials should be prosecuted in the country of origin.¹⁰¹ Furthermore, it stated that the right to transmit and receive information is a fundamental human right, and that new kinds of censorship should not be developed. Besides that, in 2016, the European Commission presented a Code of Conduct on Countering Illegal Hate Speech Online to Twitter, Facebook, YouTube, and Microsoft.¹⁰² The methods for discovering, reporting, and eliminating hate speech from social media platforms are the emphasis of the provisions in this code. As a result, they commit to establishing clear and effective protocols for reviewing reported content and removing unlawful hate speech within 24 hours of notification. Hate speech is not defined in the Code of Conduct, although it is recognized as incitement to violence against specific groups. The Commission emphasized the need of member states complying with EU law to prohibit hate speech in local legal systems in a press release launching this code in 2016.¹⁰³ In the case of social media firms, however, the Code of Conduct strives to use the word “commitments” rather than “duties”. The code is meant to create a normative compliance environment, similar to the UN Business and Human Rights Guidelines, although it is not meant to bind social media firms legally. While subsequent EU initiatives have been more explicit, precise, and powerful in their rhetoric, they effectively maintain the voluntary method to controlling hate speech on social media.¹⁰⁴ Therefore, while it is emphasized that

⁹⁸ Refer to Article 10 of the European Convention on Human Rights.

⁹⁹ Refer article 10 (2) of the European Convention on Human Rights.

¹⁰⁰ See, A. Wehbé, “Increasing International Legal Protections for Freedom of Expression”, *Notre Dame Journal of International & Comparative Law*, 8(2), 1-20, 2018. Retrieved June 4, 2020, from <https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=1085&context=ndjicl> (26.11.2020).

¹⁰¹ Recommendations from the conference on freedom of the media and the Internet, Amsterdam, 13-14 June 2003.

¹⁰² Press Release IP/16/1937 of the European Commission, European Commission and IT Companies Announce Code of Conduct on Illegal Online Hate Speech (May 31, 2016). In 2018, more businesses joined. Also see Countering Illegal Hate Speech Online #NoPlace4Hate, European Commission (Oct. 18, 2018).

¹⁰³ Irving, E., 2018, p. 259.

¹⁰⁴ Tackling Illegal Content Online: Towards Enhanced Responsibility of Online Platforms, European Commission, COM (2017) 555 final (Sept. 28, 2017). The European Union has taken moves to put compulsory measures on service providers to remove terrorist internet content, which may include hate speech and incitement but is a considerably broader category of content than hate speech and

freedom of expression must not be censored, recent international and regional legal developments have imposed an obligation on States to regulate hate speech in any medium, whether online or offline. Furthermore, social media platforms must pledge to remove hate speech information from their sites in accordance with human rights norms.

C- National Human Rights Legislation

In Ethiopia, the right to free expression is guaranteed by the constitution. According to the Ethiopian Constitution, the right to free expression includes the freedom to seek, receive, and impart information and ideas of any kind, regardless of borders, orally, in writing or in print, in the form of art, or through any media of his choice.¹⁰⁵ In accordance with international human rights standards, the Constitution imposes some limitations on this right. As a result, restrictions can be imposed in order to protect the well-being of youth as well as individuals' honor and reputation. Furthermore, any propaganda for war, as well as public expressions of opinion intended to harm human dignity, are prohibited by law.¹⁰⁶ It is unclear whether hate speech is one of these restrictions on freedom of expression. However, it appears that the phrase "public expression of opinion intended to injure human dignity" includes the concept of hate speech as a justification for restricting freedom of speech and expression. Hate speech, as defined above, is any expression intended to inflict harm on another's dignity. The Ethiopian Constitution imposes restrictions on freedom of expression in a significantly different way than other international human rights instruments. However, this does not imply that the restrictions are insufficient by international standards. Ethiopia has passed a law that affects freedom of expression in addition to the Constitutional guarantee. It's worth talking about the current rule on disinformation and hate speech on social media. Following Ethiopia's 2018 political reform, hate speech and disinformation have proliferated on social media, resulting in widespread ethnic tensions and bloodshed across the country.¹⁰⁷ Thus, Ethiopia has adopted a new proclamation¹⁰⁸ to prevent and suppress hate speech and disinformation. Hate speech is seen as a danger to social harmony, political stability, national unity, human dignity, diversity, and equality, according to the proclamation. The Ethiopian government's banning of hate speech has sparked a raging debate in academics and the media, notably concerning its impact on freedom of expression. Despite the fact that the law was enacted after lengthy consideration and with the participation of academia and civil society in the drafting process, there are still significant legal issues and reservations about it.¹⁰⁹ To begin with, the proclamation defined hate speech as "speech that deliberately promotes hatred, discrimination or

incitement. Terrorist Content Online: Council Adopts Negotiating Position on New Rules to Prevent Dissemination, Press Release from the Council of the European Union (Dec. 6, 2018).

¹⁰⁵ Refer to Article 29 (2) of the FDRE Constitution.

¹⁰⁶ See Constitution of the Federal Democratic Republic of Ethiopia, Article 29(6).

¹⁰⁷ JMC, 2020, June 4.

¹⁰⁸ See Hate Speech and Disinformation Prevention and Suppression Proclamation No. 1185/2020.

¹⁰⁹ B. Taye, "Ethiopia's Hate Speech And Disinformation Law: The Pros, The Cons, And A Mystery", 2020, May 19. Retrieved from accessnow.org: <https://www.accessnow.org/ethiopias-hate-speech-and-disinformation-law-the-pros-the-cons-and-a-mystery/> (9.12.2020).

attack against a person or a discernable group of identity, based on ethnicity, religion, race, gender or disability”¹¹⁰. In failing to define what constitutes hostility, prejudice, and aggression, this definition lacks objective clarity. As a result, any effective execution of the law without harming persons’ right to freedom of expression is impossible. Second, the legislation criminalizes hate speech disproportionately and imposes sanctions that are not appropriate to the crime, despite the fact that there are several other techniques for suppressing hate speech, such as computer literacy promotion, education, and public messaging. Punishing or imprisoning people only for posting hate speech on social media, for example, is a violation of international norms. Third, the proclamation includes rules that augment the punishment for people and online groups with over five thousand followers.¹¹¹ Because most politicians and activists with more than five thousand followers share their political ideas on social media, this provision has the potential to stifle political dissent. Finally, the government is expected to draft directives to aid in the implementation of the law. This directive can spot some keywords that are not in the proclamation, reducing their detrimental influence on free speech. However, institutionalizing mass surveillance of social media users in order to control the information that people publish online can increase the risk of endangering people’s rights even more. As a result, Ethiopia’s approach to policing hate speech on social media is extremely problematic, as it puts freedom of expression in grave jeopardy.¹¹²

III. THE REGULATORY ISSUES ON ‘HATE SPEECH’ ON SOCIAL MEDIA

A- The Rationales for Legal Control On ‘Hate Speech’

Most democracies across the world nowadays restrict hate speech since such inflammatory statements not only injure but also disrupt public order by generating violent consequences such as hate crimes, among other things. The criterion of “incitement to violence” appears to be a more tangible basis for legislating speech restrictions. In comparison to other forms of hate speech, incitement to violence necessitates a higher amount of injury to be shown, and hence is appropriately subject to criminal punishment.¹¹³ When it comes to the criminalization of speech, legal scholars disagree on what types of speech should ideally be criminalized, whether only certain types of hate speech should be prohibited, and whether all hate speech should be punished by criminal law or dealt with 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.¹¹⁴ Some argue that criminal sanctions can be used most effectively to reduce hatred. Sometimes hate speech is directed at a specific individual, but other times it is directed at a group of people, or even an entire community. As a result, the gravity of the consequences of hate speech is directly

¹¹⁰ Article 2 (2) of the Proclamation.

¹¹¹ See article 7(4) of the Proclamation.

¹¹² A. Brown, “What is hate speech? Part 1: The Myth of Hate”, *Law and Philosophy*, 36(4), 2017. from https://www.researchgate.net/publication/315473313_What_is_hate_speech_Part_1_The_Myth_of_Hate. doi:10.1007/s10982-017-9297-1

¹¹³ Yadav, 2018, p. 5 et al.

¹¹⁴ See, Brannon, 2019.

related to and heavily dependent on its target. On a personal level, given that humans have an intrinsic need to be respected and treated with dignity, when they are the subject of hate speech, the harm done is almost always in the form of acute psychological injury, while physical assault and bodily hurt are not fully ruled out. In the best-case scenario, those who have been victims of hate speech may be inspired to speak up against those who have hurt them and defend their dignity, integrity, and well-being. Injury, in the worst case, can lead to a variety of psychological conditions and disorders, one of which is severe depression, which can completely devastate not only the victim's life but also the lives of those closest to them. When hate speech is directed at a group of people or a whole community, it constitutes a far more significant threat than it can result in an immediate public disorder, escalating violence, or fatalities.¹¹⁵ Unfortunately, the world has seen many deadly and catastrophic events in the past, many of which were sparked, among other things, by hate speech. Using the Anti-Defamation League's "Pyramid of Hate Tool," Roiha demonstrated how hate speech can contribute to stigmatization, denigrating, and threatening members of various vulnerable groups in a society.¹¹⁶ From the bottom to the top, the Pyramid depicts biased behaviors that increase in complexity - generic bias, individual acts of prejudice, discrimination, bias-motivated violence, and genocide. Although the behaviors at each level have negative repercussions for people, groups, and society, the actions at higher levels have more grave consequence; therefore the lower levels sustain the upper levels. Lower-level behavior becomes more acceptable if people or organizations treat it as acceptable or normal. However, little is known about how hostile digital media influences and encourages unfavorable attitudes, prejudices, and illegal or hostile action toward specific persons and groups.¹¹⁷ It is often assumed that the spread of hate speech on the internet, especially by politicians, can lead to public unrest, violence, and ethnic massacres.¹¹⁸ The two main arguments for regulating hate speech on social media are incitement to violence and social harm. Many countries are currently enacting hate speech legislation, with the underlying arguments being incitement to violence and the protection of certain groups of people, particularly minorities. Although other countries' hate speech laws prioritize the protection of vulnerable groups and minorities, it appears that the Turkish approach prioritizes the protection of state unity over the protection of vulnerable groups and minorities.¹¹⁹

¹¹⁵ See, Neshkovska, 2018.

¹¹⁶ Roiha- Jubany, 2016.

¹¹⁷ . Rohlfig, "The Role of Social Networking in Shaping Hatred: An Exploration into User Responses to and Influence and Permissibility of Online Hatred", The thesis is submitted in partial fulfilment of the requirements for the award of the degree, 2017, January 17. Portsmouth, UK.

¹¹⁸ See, Z. B. Varda, Nefret Söylemi ve Yeni Medya. Maltepe Üniversitesi İletişim Fakültesi Dergisi, 2(1), ss.132-156, 2015. Retrieved February 21, 2021, from <https://dergipark.org.tr/en/download/article-file/172390> (8.8.2020); See, S. Mishra- A. K. Shukla, "Balancing Freedom of Expression and Hate Speech: Case of India", Pramana Research Journal, 9(6), pp.1409-1419, 2019. doi:10.6084/m9.figshare.11345261

¹¹⁹ See, C. Deveci- B. N. B. Kınık, "Nationalist Bias in Turkish Official Discourse on Hate Speech: A Rawlsian Criticism", Turkish Studies, 20(""), ss.1-24, June 2018. doi:10.1080/14683849.2018.1479961

B- The Challenges of Regulating 'Hate Speech'

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.¹²⁰ 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, Yadav¹²¹ proposes that instead of issuing excessive legal restrictions and criminalization of hate speeches, 'Alternative Dispute Resolution (ADR)¹²² and 'counter-speech'¹²³ could be successful answers to the country's hate speech problem. Many countries are currently debating hate speech laws and banning of hate speech on social media.¹²⁴ Excessive prohibitions on hate speech on social media, according to some human rights defenders, can drastically impair an individual's right to freedom of expression.¹²⁵ Besides, some politicians and think tanks argue that excessive speech restriction and media censorship can be manipulated to silence political opposition in some countries.¹²⁶ For example, the German government adopted a law against online hate speech that threatens online platform providers, such as Facebook, with fines of up to fifty million euros if they fail to erase unlawful content that is obviously illegal.¹²⁷ It is obvious that excessive regulation and criminalization of online hate speech will invite policy makers to employ social media censorship. As shown in Turkey, social media restriction can stifle the

¹²⁰ Yadav, 2018.

¹²¹ Yadav, 2018.

¹²² ADR shifts the attention away from formal legal proceedings in the courtroom and toward the solution of disputes between parties through negotiation, mediation, arbitration, or conciliation.

¹²³ Simply described, counter-speech is a reaction to harsh speech that may call for violence, encourage hatred, or use incendiary phrases to incite or defame others. Counter-speech is a firm and precise response to hate speech that is completely focused on correcting the harm caused by the hate speech in the first place. It could be a direct response 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, cartoons, memes, and portrayals to counter and defuse the tension that a hate message aims to promote, depending on the medium used to spread hateful speech.

¹²⁴ D. L. Hudson, "Free Speech Or Censorship? Social Media Litigation Is A Hot Legal Battleground", April 12019. Retrieved from Abajournal: <https://www.abajournal.com/magazine/article/social-clashes-digital-free-speech> (19.2.2021).

¹²⁵ See, V. C. Brannon, 2019; E. Çelik, "Nefret Söylemi İfade Özgürlüğünün Neresinde?" İnönü Üniversitesi Hukuk Fakültesi Dergisi, 4(2), pp. 205-240, 2013, p. 235. Retrieved February 21, 2021, from <https://dergipark.org.tr/tr/download/article-file/208301> (9.12.2020).

¹²⁶ S. Repucci, "Media Freedom: A Downward Spiral", Freedom House, 2019. Retrieved from Freedom House: <https://freedomhouse.org/report/freedom-and-media/2019/media-freedom-downward-spiral> (9.12.2020); See, D. Ç. Ulusoy, "A Comparative Study of the Freedom of Expression in Turkey and EU", The Turkish Year Book of International Relations, C.44, pp. 51-148, 2013. doi:10.1501/Intrel_0000000287

¹²⁷ Schwarz, 2019, November 3, p. 11.

right to freedom of expression and political dissent.¹²⁸ Other methods, such as creating an ethical code of behavior and instituting disciplinary actions, can be used to mitigate the detrimental consequences of hate speech on social media instead of undue criminalization.¹²⁹

2. Issues of Jurisdiction

Some countries are passing laws to address hate speech on social media platforms. Among the measures are the criminalization of hate speech and the imposition of civil liabilities in the form of compensation paid to the victim by the offender.¹³⁰ However, when it comes to the criminal investigation and prosecution of hate crimes, there are challenges connected to the place of crime, which concerns the location of the crime and the legal status of the social media company in the country in which the service is delivered. But while a derogatory social media post may be considered an online hate crime in the country where it was produced, it may not be in the country where the service was provided. Any attempt to govern material on the Internet, particularly while protecting freedom of expression, is extremely challenging due to the Internet's unique nature. In contrast to traditional media, identifying the authors of internet content is frequently difficult. Material developed in one country is copied, modified, and distributed across geographical boundaries, and it can be hosted in numerous countries, each with their own set of rules.¹³¹ There are other practical obstacles to contend with, in addition to the obstacles provided by jurisdiction and regulation. For example, once a piece of content is published on the Internet, it is frequently replicated across multiple servers or online archives, making it difficult, if not impossible, to totally erase all copies. Because of their global nature, indecision over the right regulation and jurisdiction is a major hurdle to addressing hate speech on social media platforms.¹³² Legality is a crucial requisite for restricting freedom of expression, according to international human rights law. One of the most important requirements of Article 19(3) of the ICCPR is legal certainty. This rule states that any restriction on free expression must be justified by law. However, because social media sites do not always operate inside national borders, the demand for legality in the digital sphere becomes more problematic. The legal reasons for limitation in one jurisdiction may not constitute unlawful hate speech in some other.

¹²⁸ Y. Akdeniz- K. Altıparmak, "Silencing Effect on Dissent and Freedom of Expression in Turkey in Journalism at Risk: Threats, Challenges and Perspectives", Council of Europe Publishing, pp. 145-173, 2015, p. 147 et al. 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 (14.9.2020).

¹²⁹ E. Taş, "Yeni Medyada Nefret Söylemi", e-Journal of New Media, 1(1), 60-71, 2017. doi:10.17932/IAU.EJNM.m.25480200.2017.1/1.60-71

¹³⁰ C. O'Regan, "Hate Speech Online: an (Intractable) Contemporary Challenge?", *Current Legal Problems*, 71(1), pp. 403-429, 2018

¹³¹ OSCE-ODIHR, *Incitement to Hatred vs. Freedom of Expression: Challenges of combating hate crimes motivated by hate on the Internet*, 2010. Warsaw: Report of the OSCE-ODIHR Expert Meeting. Retrieved March 18, 2020, from <https://www.osce.org/odihr/68750?download=true> (9.9.2020).

¹³² See, Gelashvili, 2018.

3. *The Conceptual Fluidity of 'Online Hate Speech'*

Regulating online hate speech has been particularly difficult due to the lack of an existing international consensus on what defines online hate speech and what falls in the sphere of freedom of expression.¹³³ While the EU and major social media companies have issued a code of conduct to combat and prohibit online hate speech, the First Amendment in the United States has always protected hate speech.¹³⁴ In the United States, for example, there is no clear separation between hate speech and hate crime. Only if hate speech is regarded a clear and immediate threat to other individuals or groups will it be labeled a hate crime.¹³⁵ As a result, law enforcement has been unable to prosecute the majority of cases. While there can be prosecutions in both online and offline environments where there is a clear and immediate threat to individuals or groups, proving what exactly constitutes a danger in online environments is more difficult. Determining what constitutes hate speech on social media is a difficult task. The challenge of detecting and filtering hate speech from online discussions is extremely challenging due to the lack of agreement on what constitutes hate speech.¹³⁶ As a result, the line between freedom of expression and hate speech should be more clearly defined in the literature and in the framework of criminal law for the benefit of legal regulation.¹³⁷ It is also crucial to distinguish hate speech that takes place offline from hate speech that takes place online. Hate speech on social media has created new problems, demanding the creation of new definitions of the term in order to properly govern it.¹³⁸ Online hate speech has unique qualities such as tenacity, itinerancy, anonymity, and transnationality, despite the fact that there is no fundamental difference in effect. First, hate speech can linger online for a long time, resurfacing in many ways across multiple platforms.¹³⁹ As a result, online hate speech has the potential to be more destructive and long-lasting than offline hate speech. Second, even if content is removed, it can be resurrected elsewhere, such as under a different identity on the same platform or in other virtual platforms.¹⁴⁰ Third, when people have the option of remaining anonymous, they are more likely to express

¹³³ A. Sellars, "Defining Hate Speech", Berkman Klein Center Research Publication No. 2016-20; Boston Univ. School of Law, Public Law Research Paper No. 16-48, 2016. Retrieved June 12, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2882244 (9.12.2020).

¹³⁴ See, Rohlfing, 2017, January 17.

¹³⁵ L. E. Beausoleil, "Free, Hateful, and Posted: Rethinking First Amendment Protection of Hate Speech in a Social Media World", Boston College Law Review, 60(7), 2019. Retrieved June 12, 2020, from <https://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3797&context=bclr> (5.5.2020).

¹³⁶ L. Silva- M. Mondal, et al, "Analyzing the Targets of Hate in Online Social Media". AAAI ICWSM, 1-(4), pp. 687-690, 2016, p. 689-690.

¹³⁷ Perry, and et al, p. 3-10.

¹³⁸ F. G. Onbaşı, "Social Media and the Kurdish Issue in Turkey: Hate Speech, Free Speech and Human Security", Turkish Studies, 16(1), pp. 115-130, 2015. doi:10.1080/14683849.2015.1021248

¹³⁹ M. Roiha- O. Jubany, 2016; See, B. Levin, "Cyberhate: A Legal and Historical Analysis of Extremists' Use of Computer Networks in America", American Behavioural Scientist, 45(6), pp.958-988, 2002. doi:<https://doi.org/10.1177/0002764202045006004>

¹⁴⁰ G. Udoh-Oshin, "Hate Speech on the Internet: Crime or Free Speech?", Undergraduate Honors College Theses, 2017. Retrieved June 13, 2020, from https://digitalcommons.liu.edu/cgi/viewcontent.cgi?article=1009&context=post_honors_theses (5.10.2020); See, Roiha- Jubany, 2016.

their hatred because they will not be identified or face any penalties.¹⁴¹ Encryption can make it difficult to prosecute crimes, especially hate crimes committed online. Fourth, global internet access magnifies the effects of hate speech while simultaneously making legal mechanisms for addressing online hate speech more problematic.¹⁴²

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.¹⁴³ Based on the principle of universality of human rights, the right to freedom of expression should be afforded equal protection in all countries around the world. Similarly, for a better result, national laws covering hate speech should be harmonized.¹⁴⁴ For example, the Council of Europe's attempts to promote greater uniformity in member states' hate crime law have largely proved successful in the EU.¹⁴⁵ However, when European national laws are compared to those of the United States of America, numerous significant discrepancies come into view.¹⁴⁶ The United States' foreign policy is heavily impacted by the federal Constitution's First Amendment. Most of the time, hate speech, which is often regarded as political expression, is protected by it.¹⁴⁷ Many European countries, on the other hand, are not affected by this issue. In Denmark, Germany, the United Kingdom, and France, people have been charged with offences including such hate speech on the internet.¹⁴⁸ Other parts of the world have more extreme examples of national laws, particularly when it comes to hate speech directed at religious communities. In Bangladesh, for example, defaming a religion can result in up to ten years in prison.¹⁴⁹

¹⁴¹ Roiha- Jubany, 2016; See, R. Č. Meznar, "Online Hate-Speech And Anonymous Internet Comments: How To Fight The Legal Battle In Slovenia?" *Annales Series historia et Sociologica*, 24(3), pp.477-488, 2014. Retrieved June 12, 2020, from <https://zdp.si/wp-content/uploads/2015/08/ceferin-meznar.pdf> (9.2.2021).

¹⁴² See, Roiha, et al, 2016; Matthew and Williams, 2015.

¹⁴³ D. J. Richards, "Rethinking Free Speech and Civil Liability", *Columbia Law Review*, 109(7), pp. 1650-1707, 2009. Retrieved June 7, 2020, from <https://www.jstor.org/stable/40380373> (9.12.2020).

¹⁴⁴ Alkiviadou, 2018.

¹⁴⁵ V. Jourová, "Code of Conduct—Illegal Online Hate Speech Questions and Answers", Fact Sheet, June 2016. Retrieved from https://ec.europa.eu/info/sites/info/files/code_of_conduct_hate_speech_en.pdf (7.9.2020).

¹⁴⁶ Neshkovska, 2018

¹⁴⁷ Beausoleil, 2019.

¹⁴⁸ OBC Transeuropa, "Hate Speech: What it is And How To Contrast it", 2018, January 31. Retrieved from Resource Centre for Media Freedom in Europe: <https://www.rcmediafreedom.eu/Dossiers/Hate-speech-what-it-is-and-how-to-contrast-it> (2.12.2020).

¹⁴⁹ S. Rabbee, "Bangladesh: The Latest Assault On Free Speech", *The Interpreter*, 2018, October 2. Retrieved from The Lowy Institute: <https://www.loyyinstitute.org/the-interpreter/bangladesh-latest-assault-free-speech> (9.10.2020).

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.¹⁵⁰ 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. There are at least three different frameworks for examining regulatory constraints on social media platforms' abilities to regulate user material based on lessons learned in the United States.¹⁵¹ First, adopting the company town analogy, social media platforms could be viewed as state actors that are obligated to observe the First Amendment while regulating protected speech. Even in the absence of legislative restriction, if social media companies are viewed as state actors under the First Amendment, the Constitution will restrain their behavior. In the second main framework, social networking platforms can be compared to special enterprises such as common carriers or broadcast media. In this case, if specific aspects of social media platforms pose a threat to the medium's use for communicative or expressive purposes, courts may introduce neutral legislation to solve the problem. The final analogy compares social media platforms to news reporters, who are generally afforded complete First Amendment protection when making editorial decisions. If social media platforms were held to the same standards as newspaper editors when deciding whether and how to present user content, those editorial decisions would be afforded the broadest First Amendment protections. Which of these three frameworks will be used is primarily determined by the precise action being regulated. In the United States, for example, social media platforms may be more likely to be protected under the Constitution if they exercise more content discretion in providing user-generated content rather than broadcasting all of it uniformly.¹⁵² Furthermore, the constitution provides less protection for certain sorts of expression. As a result, courts are more inclined to back rules that target specific types of dejecting speech, such as vulgar or violence-inducing speech. Lastly, if legislation focuses on the behavior of a social media site rather than speech, the constitutional guarantee of free speech may never be invoked.

¹⁵⁰ See, C. George, "Hate Speech Law and Policy", The International Encyclopedia of Digital Communication and Society, First Edition, 2015. Retrieved from Wiley Online Library: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118767771.wbiedcs139> (6.10.2020).

¹⁵¹ See, Brannon, 2019.

¹⁵² See, S. Hawkins, "How Free Speech and Social Media Fit Together", 2012, March 8. Retrieved from Social Media Examiner: <https://www.socialmediaexaminer.com/how-free-speech-and-social-media-fit-together/> (8.4.2020).

C- Approaches to Regulating Hate Speech on Social Media

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.¹⁵³ However, such laws and restrictions continue to be the exception rather than the rule. The prevalent approach at the international, regional, and national levels is to rely on voluntary commitments and self-regulation to filter internet content.¹⁵⁴ Germany has passed a law to combat illegal online content, including hate speech, that is possibly the most highly contested and contentious measure enacted by a liberal democracy.¹⁵⁵ Social media networks with more than two million registered members are subject to intermediary liability under the legislation. These platforms must remove illegal materials such as hate speech and individual blacklisting of religions. Within a 24-hour period, all blatantly illegal materials should be removed. All other unauthorized content has a seven-day deadline. Failure to delete unauthorized materials may result in a punishment of up to 50 million euros. The law has been criticized for being imprecise and over-inclusive, privatizing Internet censoring with no transparency or due process, and encouraging over-implementation by offering an incentive to err on the side of caution rather than free speech. Furthermore, malicious remarks and insults, as well as defamation that violates international human rights norms, are examples of prohibited content categories. Many countries have been influenced by Germany's national efforts to strengthen its intermediary obligation, either directly or indirectly. Since Germany's ground-breaking law, at least 13 nations¹⁵⁶, including the EU, have implemented or proposed intermediary liability regimes that are generally similar to the German model.

2. Self-Regulation by Social Media Companies

Because social media platforms have properly realized the importance of their involvement in the spread of hate messages, they have agreed to write out terms of agreement in the hopes of regulating and restricting such potentially dangerous behaviors. For example, Facebook's terms of service state that content that insults someone on the basis of their actual or perceived race, ethnicity, national origin, religion, sex, gender, sexual orientation, handicap, or disease is not permitted on the

¹⁵³ O'Regan, 2018, p. 411.

¹⁵⁴ Irving, 2018, p. 260.

¹⁵⁵ J. Fissand- J. Mchangama, *The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship*, Justitia, Copenhagen, 2019. https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2019/11/Analyse_The-Digital-Berlin-Wall-How-Germany-Accidentally-Created-a-Prototype-for-Global-Online-Censorship.pdf (2.7.2020).

¹⁵⁶ According to Freedom House's 2019 assessment of Internet freedom, five of those countries are classified as "not free" (Belarus, Honduras, Vietnam, Venezuela, and Russia), five are classified as "partly free" (Kenya, India, Malaysia, Philippines, and Singapore), and only three are classified as "free" (Australia, France, and UK). With the exception of India, Kenya, Vietnam, and Australia, all of these countries, as well as the EU, have explicitly cited the German model law as inspiration or justification for their intermediary liability models.

platform.¹⁵⁷ YouTube’s policy also prohibits the sharing of any speech that disparages or denigrates a group based on race or ethnic origin, religion, disability, gender, age, veteran status, or sexual orientation or gender identity. Direct, specific threats of violence against others are likewise prohibited on Twitter. Overall, it appears that social media users are in agreement when it comes to banning hate speech. In this regard, major social media companies such as Facebook, Google, Microsoft, and Twitter have committed to follow a European Union code of conduct that compels them to review the majority of legitimate notifications within twenty four hours for the removal of illegal hate speech posted on their platforms.¹⁵⁸ 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.¹⁵⁹ 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 platforms. Offensive messages created by corrupt users are shared and transmitted blindly. As the harms of social media grow more obvious, states and academics are questioning how this problem can and should be addressed. It notably calls attention to what social media companies should do or do to prevent hate speech and incitement on their platforms. In terms of regulation, social media companies have been left to their own ways up until now. This strategy has come under scrutiny in the wake of data breaches, election intervention, hate speech, and provoking scandals.¹⁶⁰ Artificial intelligence and computational technologies are also being used by major social media platforms to detect and delete online hate speech from their sites.¹⁶¹ There are concerns, however, about the bias of software-based judgments of what is appropriate online content, as well as how the learning of an artificial intelligence system might be used in unanticipated and unforeseen ways.¹⁶² 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. To put it another way, social media written text is rarely well-formed and frequently contains extralinguistic components like URLs, emojis, and other special characters.¹⁶³ Another source of difficulty is hate speech detection's intrinsically

¹⁵⁷ T. Çiftçi- L. Gashi- R. Hoffmann- D. Bahr- et al, “Hate speech on Facebook”, 4th European Conference on Social Media Research, At Mykolas Romeris University Vilnius, Lithuania, 2018. Retrieved June 12, 2020, from https://www.researchgate.net/publication/323512148_Hate_speech_on_Facebook (13.10.2020).

¹⁵⁸ 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>.

¹⁵⁹ Mishra, et al., 2019, p. 1411 et al.

¹⁶⁰ See, Irving, 2018

¹⁶¹ K. Budek, “How Artificial Intelligence Can Fight Hate Speech In Social Media”, in Data science, Machine learning, 2019, January 24. Retrieved from deepsense.ai: <https://deepsense.ai/artificial-intelligence-hate-speech/> (29.10.2020).

¹⁶² Melhuish, 2018

¹⁶³ T. Simonite, “Facebook’s AI for Hate Speech Improves. How Much Is Unclear”, 2020, May 12. Retrieved from <https://www.wired.com/story/facebook-ai-hate-speech-improves-unclear/> (9.12.2020).

unstable nature, which is exacerbated by the fact that the vast majority of social media posts do not contain hate speech or vulgarity.¹⁶⁴ Finally, there is a third level of difficulty due to the subjective nature of hate speech categorization. When the term “fuck”, for example, is used as part of a hash-tag, people react very differently than when it is used without a tagline, demonstrating the subjective nature of hate speech and offensive content.¹⁶⁵

CONCLUSION

The problem of hate speech laws has piqued the interest of many legal and philosophical scholars. The vast majority of scholarly works concentrate on presenting and critiquing arguments for and against hate speech restrictions. Despite the fact that hate speech has been recognized as an increasing threat to human rights on both a global and local level, enacting legislative prohibitions on it has proven difficult. Some of the major issues of regulating hate speech have been identified in this article. Primarily, the lack of a global legal consensus on what constitutes online hate speech and what falls within the realm of freedom of speech or expression has made regulating online hate speech extremely problematic. Second, excessive media control and over-criminalization of hate speech have impeded the right to freedom of expression and political dissidence in some countries. Third, the indeterminacy of applicable legislation and jurisdiction is a significant hurdle to combating hate speech because of the global nature of social media platforms. 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.

¹⁶⁴ R. Gorwa- R. Binns- C. Katzenbach, “Algorithmic Content Moderation: Technical And Political Challenges In The Automation of Platform Governance”, *Big Data & Society*, 1-25, 2020. doi:<https://doi.org/10.1177/2053951719897945>

¹⁶⁵ P. Alonso- R. Saini- G. Kovacs, “The North at HASOC 2019: Hate Speech Detection in Social Media Data”, *FIRE* 2019, pp. 293-299, 2019. Kolkata: Creative Commons License Attribution 4.0 International (CC BY 4.0).

REFERENCES

AFP, Hate speech in focus again: Facebook apologises for role in 2018 Sri Lanka unrest, 2020, May 13. 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> (2.10.2020).

Ağtaş, Ö.- Özkazanç, A., "Judith Butler'in Nefret Söylemi Eleştirisi: Dildeki Performatif ve Yaralayıcı Dil. Fe Dergi: Feminist Eleştiri," 10(1), 1-12, 2018. Retrieved February 18, 2021, from https://www.researchgate.net/publication/323448139_Judith_Butler'in_Nefret_Soylemi_Elestirisi_Dildeki_Performatif_ve_Yaralayici_Dil (3.10.2020).

ALA, American Library Association, 2017, December. Retrieved from Hate Speech and Hate Crime: <http://www.ala.org/advocacy/intfreedom/hate> (9.12.2020).

Al-Hassan, A. and Al-Dossari, H., "Detection of Hate Speech in Social Networks: A Survey On Multilingual Corpus", Computer Science & Information Technology (CS & IT), 83-100, 2019. DOI: 10.5121/csit.2019.90208.

Alkiviadou, N., Hate speech On Social Media Networks: Towards A Regulatory Framework? Information & Communications Technology Law, 28(1): 1-17, 2018. doi:10.1080/13600834.2018.1494417

Alonso, P. and Saini, R. and Kovacs, G., The North at HASOC 2019: Hate Speech Detection in Social Media Data. FIRE 2019 (pp. 293-299). Kolkata: Creative Commons License Attribution 4.0 International (CC BY 4.0).

Altıparmak, K.- Akdeniz Y., "Silencing Effect on Dissent and Freedom of Expression in Turkey in Journalism at Risk: Threats, Challenges and Perspectives", Council of Europe Publishing," 145-173, 2015. 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 (14.9.2020).

Altman, A., "Liberalism and Campus Hate Speech: A Philosophical Examination", An International Journal of Social, Political, and Legal Philosophy, 103(2), 302-317, 1993. Retrieved June 3, 2020, from <https://www.journals.uchicago.edu/doi/abs/10.1086/293497?journalCode=et> (9.12.2020).

Apalara, R., "Striking a Balance: Freedom of Expression and the Prohibition of Hate Speech and Offensive Remarks", Social Science Research Network (SSRN), 1-76, 2017. Retrieved June 7, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035410 (7.10.2020).

Appiagyei-Atua, K., "A Review of Theories of Expression in the Context of the Development Argument", University of Ghana Law Journal, XXIV, 197-227, 2005. 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 (9.12.2020).

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

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

Balcı, Ş.- Gölcü, A., "Hate Speech In Turkish And Israel Press: A Comparison Analysis Of Mavi Marmara Raid", *Conference: 9th International Symposium Communication in the Millennium*, 2011, San Diego/ USA, Volume: 9, 272-290, 2012. https://www.researchgate.net/publication/289506459_Hate_Speech_in_Turkish_and_Israel_Press_A_Comparison_Analysis_of_Mavi_Marmara_Event (9.2.2021).

Beausoleil, L. E., "Free, Hateful, and Posted: Rethinking First Amendment Protection of Hate Speech in a Social Media World", *Boston College Law Review*, 60(7), 2019. Retrieved June 12, 2020, from <https://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3797&context=bclr> (5.5.2020).

Bell, J., "Restraining the Heartless: Racist Speech and Minority Rights", *Indiana Law Journal*, 84, 963-979, 2009. Retrieved June 13, 2020, from http://ilj.law.indiana.edu/articles/84/84_3_Bell.pdf (9.12.2020).

Benesch, S., "Defining and Diminishing Hate Speech. Retrieved from *State of the World's Minorities and Indigenous Peoples 2014*", 2014. <https://minorityrights.org/wp-content/uploads/old-site-downloads/mrg-state-of-the-worlds-minorities-2014-chapter02.pdf> (9.12.2020).

Brannon, V. C., "Free Speech and the Regulation of Social Media Content", *Congressional Research Service*, 2019. Retrieved from <https://crsreports.congress.gov> (9.12.2020).

Brown, A., "What is hate speech? Part 1: The Myth of Hate", *Law and Philosophy*, 36(4). 2017. doi:10.1007/s10982-017-9297-1 https://www.researchgate.net/publication/315473313_What_is_hate_speech_Part_1_The_Myth_of_Hate

Budek, K., "How Artificial Intelligence Can Fight Hate Speech In Social Media", 2019, January 24. Retrieved from [deepsense.ai: https://deepsense.ai/artificial-intelligence-hate-speech/](https://deepsense.ai/artificial-intelligence-hate-speech/) (29.10.2020).

Cease Fire Centre for Civilian Rights, *Peoples under Threat 2019*. Minority Rights Group International. from: <https://minorityrights.org/publications/peoples-under-threat-2019/>

Chong, D., "Free Speech and Multiculturalism In and Out of the Academy", *Political Psychology*, 27(2), 29-54, 2006. doi: <https://doi.org/10.1111/j.1467-9221.2006.00448.x> (2.12.2020).

Çelik, E., "Nefret Söylemi İfade Özgürlüğünün Neresinde?", *İnönü Üniversitesi Hukuk Fakültesi Dergisi*, 4(2), 205-240, 2013. Retrieved February 21, 2021, from <https://dergipark.org.tr/tr/download/article-file/208301> (9.12.2020).

Çiftçi T. vd., (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 (13.10.2020).

Dalkılıç, M. Y., "Nefret Söylemi ve İfade Özgürlüğü Tartışmaları Çerçevesinde İslamofobia: L'Express Dergisi Örneği", *Medya ve Din Araştırmaları Dergisi*, 1(1), 89-109, 2018. Retrieved February 20, 2021, from [https://dergipark.org.tr/tr/download/article-file/496944\(9.12.2020\)](https://dergipark.org.tr/tr/download/article-file/496944(9.12.2020)).

Demaske, C., "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, 2019. doi:<https://doi.org/10.1080/10811680.2019.1627800>

Demirbaş, T., "Nefret Söylemi ve Nefret Suçları", *D.E.Ü. Hukuk Fakültesi Dergisi*, 2693-2701, 2017. Retrieved February 20, 2021, from <https://hukuk.deu.edu.tr/wp-content/uploads/2017/11/3-TIMUR-DEMIRBAS.pdf> (19.12.2020).

Donovan, J., "Navigating the Tech Stack: When, Where and How Should We Moderate Content?" 2019, October 28. Retrieved from INTERNET GOVERNANCE, PLATFORM GOVERNANCE: https://www.cigionline.org/articles/navigating-tech-stack-when-where-and-how-should-we-moderate-content?gclid=CjwKCAjw8pH3BRAXEiwA1pvMsYYWtF-G2eGsXdOFFdGvSfJt1zQjOjOjybYETAYxfZLihmXNkjAGRoC-SsQAvD_BwE (11.11.2020).

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

European Parliament, "The role of social media in ethnic conflicts", 2018, November 6. Retrieved from Parliamentary questions: https://www.europarl.europa.eu/doceo/document/E-8-2018-005636_EN.html (17.8.2020).

Fiss J. -Mchangama, J. "The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship", *Justitia*, Copenhagen, 2019. https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2019/11/Analyse_The-Digital-Berlin-Wall-How-Germany-Accidentally-Created-a-Prototype-for-Global-Online-Censorship.pdf (2.7.2020).

Gelashvili, T., "Hate Speech on Social Media: Implications of Private Regulation and Governance Gaps" 2018. JAMM07 Master Thesis. Lund University, Faculty of Law. <https://lup.lub.lu.se/luur/download?func=downloadFile&recordId=8952399&fileId=8952403> (6.10.2020).

George, C., "Hate Speech Law and Policy", 2015. Retrieved from Wiley Online Library: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118767771.wbiedcs139> (6.10.2020).

George, C., "Ethical Journalism Network", 2020, June 3. Retrieved from Hate Speech: A Dilemma for Journalists the World Over: <https://ethicaljournalismnetwork.org/resources/publications/ethics-in-the-news/hate-speech> (5.11.2020).

Gorwa, R. - Binns, R. - Katzenbach C., "Algorithmic content moderation: Technical and political challenges in the automation of platform governance", *Big Data & Society*, 1-25, 2020. doi:<https://doi.org/10.1177/2053951719897945>

Hare, I. - Weinstein, J., "Extreme Speech and Democracy", *Oxford Scholarship Online*, 2009. doi:10.1093/acprof:oso/9780199548781.001.0001

Hawkins, S., "How Free Speech and Social Media Fit Together", 2012, March 8. Retrieved from *Social Media Examiner*: <https://www.socialmediaexaminer.com/how-free-speech-and-social-media-fit-together/> (8.4.2020).

Holmes, K. R., "The Origins of "Hate Speech"", 2018, October 22. Retrieved from *The Heritage Foundation*: <https://www.heritage.org/civil-society/commentary/the-origins-hate-speech> (13.2.2021).

HRW, Human Rights Watch. Retrieved from *Somalia: Events of 2018*: <https://www.hrw.org/world-report/2019/country-chapters/somalia> (9.12.2020).

Hünler, O. S., "Popülist Sağ ve Sosyal Medyada Nefret Söylemi", *Eleştirel Psikoloji Bülteni*, 8, 6-12, 2019. Retrieved February 20, 2021, from https://www.researchgate.net/publication/336837817_Populist_Sag_ve_Sosyal_Medyada_Nefret_Soylemi (11.1.2021).

Hylton, K. N., *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), pp.35-57, 1996. Retrieved June 8, 2020, from <https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1377&context=roundtable> (23.2.2021).

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

Jacobson, D., "Freedom of Speech under Assault on Campus", 2016, August 30. Retrieved from *CATO Institute*: <https://www.cato.org/publications/policy-analysis/freedom-speech-under-assault-campus> (9.2.2021).

JMC., Jack Miller Center, 2020, June 4. 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/> (15.2.2021).

Jourová, V., "Code of Conduct—Illegal Online Hate Speech Questions And Answers", 2016, June. Retrieved from https://ec.europa.eu/info/sites/info/files/code_of_conduct_hate_speech_en.pdf (7.9.2020).

Juanatey, A. G. (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> (4.1.2021).

Kalav, A. - Certel Fırat, A. B., "Amerikan Sosyal Medyasında Göçmen Karşıtlığı ve Dijital Nefret Söylemi: Twitter Üzerinde Bir İnceleme". *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22, 2209-2222, 2017. Retrieved February 20, 2021,

from <https://iibfdergi.sdu.edu.tr/assets/uploads/sites/352/files/yil-2017-cilt-22-sayi-kayfor15-yazi42-30122017.pdf> (9.12.2020).

Kaniklidis, C., "Free Speech, Hate Speech and Principles of Community: The Case Against Free Speech Absolutism", 2015. 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 (30.10.2020).

Deveci C.- Kınık, B. N. B., "Nationalist Bias In Turkish Official Discourse On Hate Speech: A Rawlsian Criticism", *Turkish Studies*, 20(1), 1-23, June 2018. doi:10.1080/14683849.2018.1479961

Hudson, D. L., "Free Speech Or Censorship? Social media litigation is a hot legal battleground", 2019, April 1. Retrieved from ABAJOURNAL: <https://www.abajournal.com/magazine/article/social-clashes-digital-free-speech> (19.2.2021).

Laub, Z., "Council on Foreign Relations", 2019, June 7. Retrieved from Hate Speech on Social Media: Global Comparisons: <https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons> (12.9.2020).

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

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

Lombardi, C., "The Illusion of a "Marketplace of Ideas" and the Right to Truth", 2019. Retrieved from *American Affairs*: <https://americanaffairsjournal.org/2019/02/the-illusion-of-a-marketplace-of-ideas-and-the-right-to-truth/> (2.2.2021).

Maas, C., "Ethiopia parliament passes law punishing hate speech", 2020, February 15. Retrieved from *Jurist.org*: <https://www.jurist.org/news/2020/02/ethiopia-parliament-passes-law-punishing-hate-speech/> (9.12.2020).

Macedo, R. P., "Freedom of Expression: what lessons should we learn from US experience?", *Rev. Direito GV*, 13(1), pp. 274-302, 2017. doi:<http://dx.doi.org/10.1590/2317-6172201711>

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

Matthew L. - Williams, P. B., "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, 2019. doi:10.1093/bjc/azz049

Melhuish, E. P., "Online Hate Speech: A Survey on Personal Experiences and Exposure among Adult New Zealanders", *Netsafe*, 2018. Retrieved March 21, 2020, from

<https://www.netsafe.org.nz/wp-content/uploads/2019/11/onlinehatespeechsurvey-survey-2018.pdf> (9.10.2020).

Meza, R. - Vincze, H. O. - Mogos, A., "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), pp 26-50, 2019. doi:10.17356/ieejsp.v4i4.503

Meznar, R. C., "Online Hate-Speech And Anonymous Internet Comments: How To Fight The Legal Battle In Slovenia?" *Annales, Ser. Historia Sociology*, 2014. Retrieved June 12, 2020, from <https://zdjp.si/wp-content/uploads/2015/08/ceferin-meznar.pdf> (9.2.2021).

Mishra, S. - Shukla A. K., "Balancing Freedom of Expression and Hate Speech: Case of India", *Pramana Research Journal*, 9(6), 1409-1419, 2019. doi:10.6084/m9.figshare.11345261

Neshkovska, S. - Trajkova, Z., "The Essential of Hate Speech", *IJET*, 71-80, 2018. https://www.researchgate.net/publication/329874520_The_Essential_of_Hate_Speech, (9.12.2020).

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

OBC Transeuropa, "Hate Speech: What It Is and How to Contrast It", 2018, January 31. Retrieved from Resource Centre for Media Freedom in Europe: <https://www.rcmediafreedom.eu/Dossiers/Hate-speech-what-it-is-and-how-to-contrast-it> (2.12.2020).

OHCHR, Preventing incitement to hatred and violence in the Central African Republic, 2019, May 29. Retrieved from United Nations Human Rights Office of the High Commissioner:

<https://www.ohchr.org/EN/NewsEvents/Pages/PeacekeepersDay2019.aspx> (9.2.2021).

Oltmann, S. M., "Intellectual Freedom and Freedom of Speech: Three Theoretical Perspectives", *Information Science Faculty Publications*, 153-171, 2016. Retrieved June 3, 2020, from https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1032&context=slis_facpub (19.9.2020).

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

OSCE-ODIHR, Incitement to Hatred vs. Freedom of Expression: Challenges of combating Hate Crimes Motivated by Hate on the Internet, 2010. Warsaw: Report of the OSCE-ODIHR Expert Meeting. Retrieved March 18, 2020, from <https://www.osce.org/odihr/68750?download=true> (9.9.2020).

Öztekin, H., "Yeni Medyada Nefret Söylemi: Ekşi Sözlük Örneği". *Uluslararası Sosyal Araştırmalar Dergisi*, 8(38), 925-936, 2015. Retrieved February 20, 2021, from https://www.sosyalarastirmalar.com/cilt8/sayi38pdf/6iksisat_kamu_isletme/oztekinhulya.pdf (9.9.2020).

Pejchal, V., "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, 2015. doi: 10.5204/ijcjsd.v7i2.500.

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

Perry, B. - Akça, D. - Karakuş F. - Baştuğ M. F., "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, 2020. doi:<https://doi.org/10.5204/ijcjsd.v9i4.1514>

Quent, M., "Opinion: The Clear Line Between Hate Speech and Free Speech", 2019, November 24. Retrieved from DW Made for Minds: <https://www.dw.com/en/opinion-the-clear-line-between-hate-speech-and-free-speech/a-51393218> (12.9.2020).

Quinn, T., "Monitoring Hate Speech: Challenges and Strategies", 2018, HateBase.

Rabbee, S., "Bangladesh: The Latest Assault On Free Speech", 2018, October 2. Retrieved from The Lowy Institute: <https://www.loyyinstitute.org/the-interpreter/bangladesh-latest-assault-free-speech> (9.10.2020).

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

Repucci, S., "Media Freedom: A Downward Spiral", 2019. Retrieved from Freedom House: <https://freedomhouse.org/report/freedom-and-media/2019/media-freedom-downward-spiral> (9.12.2020).

Richards, D. J., "Rethinking Free Speech and Civil Liability", *Columbia Law Review*, 109(7), 1650-1707, 2009. Retrieved June 7, 2020, from <https://www.jstor.org/stable/40380373> (9.12.2020).

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

Roiha, M. - Jubany, O., "Backgrounds, Experiences and Responses to Online Hate Speech: A Comparative Cross-Country Analysis", *Words are Weapons*, 2016. Preventing Redressing Inhibiting Hate Speech in New Media. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.734.2403&rep=rep1&type=pdf>

Romero, A., "The line between free speech and hate speech", 2016. ResearchGate. Retrieved June 13, 2020, from https://www.researchgate.net/publication/311064494_The_line_between_free_speech_and_hate_speech (9.12.2020).

Samples, J., *Why the Government Should Not Regulate Content Moderation of Social Media*, 2019, April 9. Retrieved from CATO Institute: <https://www.cato.org/>

publications/policy-analysis/why-government-should-not-regulate-content-moderation-social-media (2.12.2020)

Scanlon, T., "A theory of free expression", *Philosophy and Public Affairs*, 1 (2): pp. 204-226, 1972.

Schauer, F., *İfade Özgürlüğü: Felsefi Bir İnceleme*, Çev. B. Seçilmişoğlu, Liberal Düşünce Topluluğu Yayınları, Ankara, 2002.

Scheffler, A., *The Inherent Danger of Hate Speech Legislation A Case Study from Rwanda and Kenya on the Failure of a Preventative Measure*, 2015. Retrieved from Friedrich-Ebert-Stiftung (FES): <https://library.fes.de/pdf-files/bueros/africa-media/12462.pdf> (1.12.2020).

Schultz, D., "The First Amendment Encyclopedia", 2017, June. Retrieved from <https://www.mtsu.edu/first-amendment/article/999/marketplace-of-ideas> (9.11.2020).

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

Seidman, L. M., "Can Free Speech Be Pr ee Speech Be Progressive?" *Georgetown University Law Center*, 2018. Retrieved June 3, 2020, from <https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=3056&context=facpub> (9.11.2020).

Sellers, A., "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 (9.12.2020).

Silva, L. - Mondal, M. - Correa D. - Benevenuto, F. et al, "Analyzing the Targets of Hate in Online Social Media", *AAAI ICWSM*, 1-4: 687-690, 2016.

Simonite, T., "Facebook's AI for Hate Speech Improves. How Much Is Unclear", 2020, May 12. Retrieved from <https://www.wired.com/story/facebook-ai-hate-speech-improves-unclear/> (9.12.2020).

Solmone, S., "Regulate social media? It's a bit more complicated than that", 2018, May 3. Retrieved from *The Conversation*: <https://theconversation.com/regulate-social-media-its-a-bit-more-complicated-than-that-103797> (18.10.2020).

Şahin, K., *İfade Özgürlüğü: Gerekçeleri ve Sınırları*, XII. Levha yayınları, İstanbul, 2009.

Taş, E., "Yeni Medyada Nefret Söylemi", *e-Journal of New Media*, 1(1), 60-71, 2017. doi:10.17932/IAU.EJNM.m.25480200.2017.1/1.60-71

Taye, B., "Ethiopia's Hate Speech And Disinformation Law: The Pros, The Cons, And A Mystery", 2020, May 19. Retrieved from [accessnow.org: https://www.accessnow.org/ethiopias-hate-speech-and-disinformation-law-the-pros-the-cons-and-a-mystery/](https://www.accessnow.org/ethiopias-hate-speech-and-disinformation-law-the-pros-the-cons-and-a-mystery/) (9.12.2020).

Thompson, P. - Itaoui, R. - Bazian H., "Islamophobia in India: Stoking Bigotry", *Berkeley: Islamophobia Studies Center*, 2019. Retrieved March 21, 2020, from

<https://irdproject.com/wp-content/uploads/2019/04/Islamophobia-in-India-Web-Spread.pdf> (9.12.2020).

Traum, A., "Contextualising The Hate Speech Debate: The United States and South Africa", *The Comparative and International Law Journal of Southern Africa*, 47(1), 64-88, 2014. Retrieved June 12, 2020, from <https://www.jstor.org/stable/24585817?seq=1> (9.12.2019).

Tsesis, A., "Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements", 2002. Researchgate. Retrieved June 7, 2020, from https://www.researchgate.net/publication/228321836_Destructive_Messages_How_Hate_Speech_Paves_the_Way_for_Harmful_Social_Movements (9.9.2020).

Udoh-Oshin, G., "Hate Speech on the Internet: Crime or Free Speech?" Undergraduate Honors College Theses, 2017. Retrieved June 13, 2020, from https://digitalcommons.liu.edu/cgi/viewcontent.cgi?article=1009&context=post_honors_theses (5.10.2020).

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

US Department of State, Sudan 2018 Human Rights Report, 2019, March. Retrieved from <https://www.state.gov/wp-content/uploads/2019/03/Sudan-2018.pdf> (9.10.2020).

Uygun, O., "Avrupa İnsan Hakları Sözleşmesi ve Türk Hukukunda İfade Özgürlüğünün Sınırlandırılması", *Türkiye'de İfade Özgürlüğü*, (Ed. T. Koçak, T. Doğan, Z. Kutluata) bgst yayınları, İstanbul, 2009, ss.15-67

Varda, Z. B., "Nefret Söylemi ve Yeni Medya", *Maltepe Üniversitesi İletişim Fakültesi Dergisi*, 132-156, 2015. Retrieved February 21, 2021, from <https://dergipark.org.tr/en/download/article-file/172390> (8.8.2020).

Walker, J., "Hate Speech and Freedom of Expression: Legal Boundaries in Canada", 2018, June 29. Retrieved from Library of Parliament: https://lop.parl.ca/sites/PublicWebsite/default/en_CA/ResearchPublications/201825E#ftn2 (26.11.2020).

Wang, Q., "The Understanding of Absolute Right to Freedom of Expression in the Case of Hate Speech", *Graduate Theses and Dissertations*, pp. 1-100, 2018, March. Retrieved from <http://scholarcommons.usf.edu/etd/7240> (4.11.2020).

Wehbé, A., "Increasing International Legal Protections for Freedom of Expression", *Notre Dame Journal of International & Comparative Law*, 8(2), 1-20, 2018. Retrieved June 4, 2020, from <https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=1085&context=ndjicl> (26.11.2020).

Weiner, M. S., "Self-Governance and Free Speech", 2012, September 3. Retrieved from *Civil Liberties in the United States*: <https://uscivil liberties.org/themes/4466-self-governance-and-free-speech.html> (9.11.2020).

Weinstein, J., "Hate Speech Bans, Democracy, and Political Legitimacy", *Constitutional Commentary*, 32(527), 527-587, 2017. Retrieved June 8, 2020, from <https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1475&context=concom> (28.8.2020).

Yadav, A., "Countering Hate Speech in India: Looking for Answers Beyond The Law", *II Law Review*, II (Winter Issue), 1-17, 2018.

Yanık, A., *Sosyal Medyada Yükselen Nefret Söyleminin Temelleri (Fundamentals of Rising Hate Speech on Social Media)*. *Global Media Journal TR Edition*, 8(15), 364-383, 2017. Retrieved June 26, 2020, from https://www.researchgate.net/publication/320893358_Sosyal_Medyada_Yukselen_Nefret_Soylemine_Temelleri (9.12.2020).

Yaraghi, N., "Regulating Free Speech on Social Media Is Dangerous and Futile", 2018, September 21. Retrieved from Brookings: <https://www.brookings.edu/blog/techtank/2018/09/21/regulating-free-speech-on-social-media-is-dangerous-and-futile/> (6.6.2020).

Yazıcı, T., "Yeni Medyanın Nefret Dili: Suriyeli Mültecilerle İlgili Ekşi Sözlük Örneği", *Global Media Journal TR Edition*, 7(13), 115-136, 2016. Retrieved February 19, 2021, from <https://globalmediajournaltr.yeditepe.edu.tr/sites/default/files/T%C3%BClay%20YAZICI.pdf> (16.2.2021).