

HATE SPEECH AND SOCIAL MEDIA

Jelena Vučković*, Sonja Lučić

University of Kragujevac, Faculty of Law, Kragujevac, Serbia

Abstract

In the 21st century – the time of the Internet, hate speech is present throughout social media. Hate speech existed even before the advent of the Internet. However, what is different about online hate speech is the speed at which it is transmitted on certain social media, such as Twitter and Facebook. In this sense, the question of whether and to what extent online hate speech should, or rather should not be protected was posed. This paper aims to explore whether online hate speech deserves freedom of expression protection. Public officials, civil servants, and politicians often use different social media platforms, especially Twitter, to communicate new political ideas, upcoming events, and even conspiracy theories. In this sense, the aim of the paper is to investigate whether a need to balance between constitutionally protected freedom of expression and the core values of democratic society arose from the development of Internet technology.

Key words: hate speech, social media, the social network, Internet, freedom of expression

ГОВОР МРЖЊЕ И ДРУШТВЕНИ МЕДИЈИ

Апстракт

У 21. веку, односно у времену интернета, говор мржње је у великој мери заступљен на друштвеним медијима. Говор мржње је постојао и пре појаве интернета. Међутим, оно што је другачије у вези са онлајн говором мржње је брзина којом се говор мржње преноси на појединим друштвеним медијима, попут Твитера и Фејсбука. У том смислу, отворило се питање да ли и у којој мери онлајн говор мржње треба, односно не треба штитити. Овај рад има за циљ да истражи да ли онлајн говор мржње заслужује да буде заштићен правом слободе изражавања. Јавни званичници, државни службеници и политичари често користе различите платформе друштвених медија, најчешће Твитер, за саопштавање нових политичких идеја, предстојећих догађаја, па чак и теорија завере. У том смислу, рад има за циљ да истражи да ли се са развојем интернет технологије јавила потреба балансирања између уставом заштићене слободе изражавања и заштите основних вредности демократског друштва.

* Corresponding author: Jelena Vučković, University of Kragujevac, Faculty of Law, Jovana Cvijića 1, 34000 Kragujevac, Serbia, jvuckovic@jura.kg.ac.rs

Кључне речи: говор мржње, друштвени медији, друштвене мреже, интернет, слобода говора

INTRODUCTION

Social media play a significant role in our society. They serve as a means of communication and as platforms for marketing products, and represent a modern form of self-expression. Before the advent of social media, communication between people took place in writing or by phone. Today, however, the dominant form of communication is digital (Vučković, 2021b, pp. 521-539). The originators of this development are social media platforms – a group of Internet-based applications that enable the creation and sharing of user-generated content (Kaplan, Haenlan, 2010, p. 59).

Social media differ from traditional electronic media (Vučković, 2021a, pp.197-201) because they directly support or create social networks using the information and communication technologies (Kaplan, Haenlan, 2010, p. 61). Social media has developed at a rapid pace. In the last ten years, they have grown faster and changed more than any other Internet innovation (Mihajlov Prokopović, 2016, p.39). This has been largely contributed to by the number of adults who own smartphones, which allow social media to be followed anywhere, anytime (Duggan, Smith, 2013. p. 2).

Many social media encourage emotional self-expression and invite users to regularly share their thoughts, feelings, and experiences (Derks, Fischer, Bos, 2008, p. 770). Empirical studies have shown that individuals post both positive and negative emotional expressions online, albeit with a positive bias. This ‘positivity bias’ may be due to the prevailing norms of positivity established by social media (Reinecke, Trepte, 2014, pp. 105-107). Positive emotions are more suitable for online disclosure. On the other hand, revealing negative emotions seems more intimate, and may be perceived as inappropriate behaviour.

Today, social networks are an integral part of many people’s daily lives and a form of entertainment (Vučković, 2022, p. 55). Research has shown that most social media users integrate two or more platforms into their daily activities (Davenport, Bergman, Bergman, Fearington, 2014, p. 215). Now, there are numerous social networks that address target groups. The most important social media platforms are Facebook, YouTube, WhatsApp, Instagram, Pinterest, and Twitter.

Social media greatly influence the formation of opinions. People form, among other things, an opinion about how the state and the society they live in should be designed, and which political decisions they support or reject. This free formation of individual opinion is a basic prerequisite for the functioning of democracy. Nevertheless, the media today represent a source for a wide range of information transmission

and, at the same time, a potential danger for the creation of a dominant opinion on certain topics (Veljanovski, 2010, pp. 41-55).

Social media is becoming increasingly important for the democratic process of opinion formation. This presents both opportunities and dangers for democracy and diversity of opinion. Access to information is easy and seemingly unlimited. At the same time, social and political divisions are constantly increasing in many countries. Social media certainly contribute to these divisions: social media users share information they perceive as objective, even though it is already heavily filtered, on Facebook, Twitter, YouTube, etc. This leads to the formation of so-called filter bubbles (Vučković, 2022b, p. 555). Namely, social media do not create their own content, but only enable the technical transmission of information on the basis of which people form opinions on various issues. The fact that, in this way, people only get information that supports their views and isolate themselves from controversial discussions is seen as a danger to democracy. A special danger is the appearance of hate speech that threatens the exchange of opinions which is necessary for every democratic society (Schmidt, 2018, p. 61).

Hate speech existed even before the advent of the Internet and social media (Milinković, 1996, p. 42). But in times of online communication, hate speech has reached a “new, often viral quality” (Fleischhack, 2017, pp. 23-25). There is no single definition of hate speech in science (Petrušić, 2012, p.75) - there are a number of different, sometimes conflicting, definitions of this term. The lack of a generally applicable definition of ‘hate speech’ is considered the central problem of its regulation (Kiska, 2012, p. 107). A large number of authors from different disciplines – from legal and political to linguistic sciences – have tried to define hate speech (Brown, 2017. p. 424). Stefanowitsch defines hate speech as a political expression with more or less strong legal facts (Stefanowitsch, 2018, p. 11). It is about humiliating and belittling people, especially because of their belonging to a group.

By hate speech, Sponholz means “more conscious public communication and/or intentional messages of discriminatory content” (Sponholz, 2018, p. 48). The same author emphasises that hate speech can be communicated in various ways, not only verbally, which is why the term itself is wrong. For example, hatred can be expressed non-verbally, through pictures, facial expressions, or gestures (Meibauer 2013. pp. 1-3). The characteristics of hate speech which are repeatedly emphasised are derogatory language and judgment of other people. However, hate speech should be distinguished from hate propaganda and cyberbullying.

Another group of authors believes that the central characteristic of hate speech is group connection. In this sense, hate speech is defined as “aggressive or generally derogatory statements about people who are assigned to certain groups” (Geschke, Klaßen, Quent, Richter, 2019, pp. 14-16). Addressing such groups may be based on characteristics such as

gender, nationality, profession, sexual orientation, religious affiliation, or appearance.

Hate speech is also expressed in sexist comments on social networks. Sexism is based on theories and prejudices that view people as inferior because of their gender, and manifests itself in stereotypical and discriminatory behaviour – more often online than offline. Unlike traditional thinking that is limited to a female-male gender perspective, the term sexism expands this understanding to include hostility toward lesbian, gay, bisexual, and transgender people (Nikolić, 2018).

Hate speech should be distinguished from discrimination. Discrimination refers to cases in which individuals or groups are subjected to different treatment without an objective reason. Inferior treatment can be based on various grounds such as age, gender, race, ethnic origin, sexual orientation, etc. Many of these grounds overlap with those relating to hate speech. For this reason, it may happen that hate speech includes incitement to discrimination against certain groups or individuals.

Hate speech should be distinguished from hate crimes (Zekavica, 2019, p.39). Hate crimes are crimes that are motivated by bias or prejudice against certain groups of people. The OSCE defines bias or prejudice:

as preconceived negative opinions, stereotypical assumptions, intolerance or hatred directed at a particular group that shares common characteristics, such as race, ethnicity, language, religion, nationality, sexual orientation, gender or any other basic characteristics

(OSCE, Hate Crime Reporting)

Therefore, the basic difference between hate crime and hate speech lies in the fact that hate speech does not constitute a criminal offense. However, hate speech can be evidence of a hate crime (Milenković, 2010, p.66). Hate crimes fall into the category of violent crimes (Dimovski, 2021, p. 740).

METHODOLOGY

Online hate speech differs from offline hate speech, which, among other things, has posed numerous challenges to the attempts to regulate so-called “cyberhate” (Brown, 2018, p. 297). One of the advantages of the Internet is that it gives individuals the ability to say what they want while remaining anonymous. It has been proven that the anonymity of the internet, and the fact that people are sitting behind a screen encourages people to say things they would never say in real life (Brown, 2018, p. 300). On the other hand, it is easier for victims of hate speech to express themselves and defend themselves against hate speech on the Internet. Even when individuals do not want to hide their identity, online

invisibility contributes to making statements on social media that would never be uttered in real life. On the other hand, in traditional forms of media, there is editorial supervision before the publication of certain content (Mihajlov Prokopović, 2018, p.1086).

The legal regulation of hate speech varies from country to country. The USA leads the way in terms of the legal regulation of hate speech. American courts give almost absolute protection to freedom of expression (Brugger, 2002, p. 7). The protection of freedom of expression is absent only in cases in which the statement directly calls for immediate illegal action and is suitable to trigger such behaviour. This standard, which is still valid today, was upheld by the US Supreme Court in 1969 in the *Brandenburg vs. Ohio* 395 U.S. 444 case, after which it was named the Brandenburg test (Bleich, 2011, p. 922). The occasion for this case was a speech made by Clarence Brandenburg at a rally in Ohio, in which he made “derogatory and vengeful comments” (Fagan, 2000, p. 609). Brandenburg was arrested and convicted under the Ohio Criminal Syndicalism Act for assembling and promoting illegal activities for the purpose of political reform. However, Brandenburg appealed the ruling, citing the First and Fourteenth Amendments to the US Constitution. The Supreme Court ultimately ruled that speech is protected, and that the State may not constitutionally prohibit the advocacy of violations of the law unless that advocacy has caused immediate illegal action.

The First Amendment to the US Constitution prohibits the government and public authorities from restricting free speech. There are slight exceptions for hate speech, understood as speech likely to incite immediate violence. The First Amendment, however, does not prevent private actors, such as social media platforms, from imposing their own restrictions on hate speech. Social media platforms are further protected from private litigation because they are not considered the publishers of the content published on their sites within the meaning of section 230 of the Communications Decency Act 1996 (Vučetić, Bončić & Pešić, 2016, pp. 8-9).

At the level of the European Union, no legal regulation on the prohibition of hate speech on the Internet has been adopted to date. However, in May 2016, the EU Code of Conduct on the Fight against Hate Speech on the Internet was adopted, to which four IT companies – Facebook, Twitter, YouTube, and Microsoft committed themselves. To date, other IT companies have also decided to join the Code (Instagram, Google+, Snapchat, Dailymotion, Jeuxvideo.com, TikTok, and LinkedIn).

The EU directive on audio-visual media services contains rules against hate speech. The provisions of this Directive bind both Video on Demand and video sharing platforms such as YouTube, Netflix, or Facebook. The provisions of this Directive oblige Member States to ensure that there is no hatred, violence, or calls to terrorism in audio-

visual media. The operators of such platforms must be available for cooperation and must create mechanisms through which, for example, videos glorifying violence or hate speech can be reported. Platform providers must delete such content after appropriate review.

The e-commerce directive also contains certain provisions on hate speech. Article 3(2) and (4) of the e-Commerce directive provides that:

Member States may not, for reasons falling within the coordinated field, restrict the freedom to provide information society services from another Member State... Member States may take measures to derogate from paragraph 2 in respect of a given information society service if the following conditions are fulfilled: (a) the measures shall be:

(i) necessary for one of the following reasons:

- public policy, in particular the prevention, investigation, detection, and prosecution of criminal offenses, including the protection of minors and the fight against any incitement to hatred on grounds of race, sex, religion, or nationality, and violations of human dignity concerning individual persons.

(Directive 2000/31 on Electronic Commerce)

The EU Court of Justice has so far only dealt with the definition of hate speech (e.g. in the joined cases of *Mesopotamia Broadcast and Roj TV*, C-244/10 and 245/10). However, in no case did the Court deal with the dimension of hate speech from Art. 3(4) of the Directive on electronic commerce. Unlike the EU Court of Justice, the national courts of the Member States, including the European Court of Human Rights, have on many occasions dealt with the issue of hate speech in the media and in the online context.

The European Court of Human Rights dealt with hate speech on the Internet for the first time in the *Delphi vs. Estonia* case (*Delfi vs. Estonia*). In this case, the Court held that the Estonian newspaper internet portal *Delphi*, which published an article about how the ferry company SLK destroyed the territory traditionally used for sailing from the Estonian mainland to its islands, was responsible for the offensive comments of its readers. Under the article, readers wrote allegedly offensive and threatening comments. In this case, ECHR established several important principles. First, ECHR established that the *Delphi* portal is the publisher of the comments. The website of the portal states that the authors of the comments will be responsible for their content, and that threatening or offensive comments are not allowed. However, the Court concluded that, although it knew and could have prevented the defamation, *Delphi* failed to do so and left it on the website for six weeks.

According to ECHR, defamation contained in electronic communication, compared to traditional print or electronic media, differs in that it can remain there forever and cause much greater

damage. ECHR has drawn particular attention to the uncontrolled spread of potentially defamatory and hateful rhetoric.

In the case of *OOO Flavus a. o. vs. Russia*, The ECHR established that the application of measures to block internet media or sites due to a critical attitude towards the authorities or the political system can never be considered a necessary restriction of freedom of expression (*Flavus a. o. vs. Russia*). Blocking complete access to a website is an extreme measure that should be compared to banning a newspaper or television station.

Some individual EU Member States have adopted their own regulations containing provisions against hate speech. In Germany, statements that can be characterised as ‘hate speech’ are often brought under the criminal offense of Incitement of the Masses in accordance with Art. 130 of the Criminal Code. Besides, hate speech can be classified under the criminal offenses of Public Incitement to Commit Offenses (Section 111), Insult (Section 185), and Threat (Section 241). In 2017, the Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act) was adopted in Germany in order to better enforce substantive law on the Internet. This Law obliges social network operators to delete obviously illegal content or block access to it within 24 hours of receiving a complaint. For content subject to deletion or blocking, the Law does not use the term hate speech but refers to the existing provisions of the Criminal Code (Sections 86, 86a, 89a, 91, 100a, 111, 126, 129 to 129b, 130, 131, 140, 166, 184b, in connection with 184d, 185 to 187, 241 or 269 of the Criminal Code).

The Network Enforcement Act defines social networks as:

Telemedia service providers which, for profit-making purposes, operate internet platforms which are designed to enable users to share any content with other users or to make such content available to the public (social networks).

(Network Enforcement Act, Art. 1, Sec. 1 (1))

Platforms offering journalistic or editorial content, the responsibility for which lies with the service provider itself, do not constitute social networks within the meaning of this Act. The same applies to platforms that are designed to enable individual communication or the dissemination of specific content (Network Enforcement Act, Section 1(1)). Section 2(1) of the Act states:

Providers of social networks which receive more than 100 complaints per calendar year about unlawful content shall be obliged to produce half-yearly German-language reports on the handling of complaints about unlawful content on their platforms.

(Network Enforcement Act, Sec. 2(1))

Fines for social network providers that violate reporting requirements or regulations for handling complaints about illegal content can amount to up to €50 million.

The French Law on the Freedom of the Press and the French Criminal Code contain various norms directed against hate speech. In the middle of 2019, a law was adopted in France with the aim of suppressing hate content on the Internet. The law established a specialised prosecutor's office for the fight against hateful content on the Internet. In addition, the Law obliges large operators of online platforms, whose activity consists of connecting multiple people in order to share content or refer to this content, to remove 'obviously' illegal content within 24 hours of receiving the notification. Otherwise, they risk being fined up to €1.25 million. Removal applies to content that includes incitement to hatred, violence, racist or even religious insults. For terrorist or child pornography content, the removal period is reduced to one hour. Platforms are given a week to remove less explicit content. Content reporting is up to the user, and relies on a 'report' button made visible by the platforms. However, the Constitutional Council condemned these provisions in a decision dated June 18, 2020.

As a critic of the law, the Constitutional Council condemned the obligation imposed on social networks, such as Facebook, Twitter, Snapchat, and YouTube, to delete 'hate' content reported to them within twenty-four hours under the threat of large fines. According to the opinion of the Constitutional Council, this mechanism, which is devoid of legal intervention, can cause "an attack on the exercise of freedom of expression and communication that is not necessary, appropriate and proportionate to its goal" (Constitutional Council of France, Decision no 2020-801 DC). The absence of legal intervention specifically refers to the obligation of social networks to remove any child pornography or terrorist content reported by the authorities within an hour.

In Great Britain, the Public Order Act from 1986 provides that:

A person who uses threatening, abusive or insulting words or behavior, or displays any written material which is threatening, abusive or insulting, is guilty of an offense if - (a) he intends thereby to stir up racial hatred, or (b) having regard to all the circumstances racial hatred is likely to be stirred up thereby.

(Public Order Act)

Section 17 of the same Law reads: "racial hatred means hatred against a group of persons defined by reference to color, race, nationality (including citizenship) or ethnic or national origins" (Public Order Act, Section 17).

Unlike Germany and France, Great Britain has not passed a Law on the Suppression of Hate on the Internet. However, the UK Government has repeatedly emphasised that what is illegal offline is also illegal online. In April 2019, the British government published the White Paper, which

outlined the government's proposals for regulating harmful content on the Internet. Online harm is defined in the White Paper as:

online content or activity that harms individual users, particularly children, or threatens the way of life in the UK, either by undermining national security or by undermining trust and undermining shared rights, responsibilities and opportunities to foster integration.

(The Online Harms White Paper)

In the middle of 2021, a draft of the Online Safety Bill was published in Great Britain, but it is yet to be adopted.

Legal Regulation of Hate Speech in Serbia

In Article 46, paragraph 1, the Constitution of the Republic of Serbia guarantees freedom of expression. This freedom can be limited by law "if it is necessary to protect the rights and reputation of others, preserve the authority and impartiality of the court and protect public health, morals of a democratic society and national security of the Republic of Serbia" (The Constitution of the Republic of Serbia, Article 46, paragraph 1). Article 43, paragraph 4 of the Constitution stipulates that:

freedom of expression of religion or belief can be limited by law, only if it is necessary in a democratic society, for the purpose of protecting people's lives and health, the morals of a democratic society, the freedoms and rights of citizens guaranteed by the Constitution, public security and public order, or to prevent or inciting religious, national or racial hatred.

(The Constitution of the RS, Art. 50, para. 3)

In addition, the Competent Court can prevent the dissemination of information and ideas through the means of public information, among other things, in order to prevent advocacy of racial, national or religious hatred, which incites discrimination, hostility, or violence.

Unlike the Constitution, the Law on Prohibition of Discrimination expressly prohibits hate speech. This Law defines hate speech as

the expression of ideas, information, and opinions that incite discrimination, hatred or violence against a person or a group of persons because of their personal characteristics, in public newspapers and other publications, at gatherings and places accessible to the public, by writing and displaying messages or symbols and in another way.

(Law on Prohibition of Discrimination, Art. 11)

Such behaviour is prohibited by this Law.

The Criminal Code prescribes four criminal offenses related to the prohibition of discrimination: (1) violation of equality (Section 128); (2) violation of the right to use language and script (Section 129); (3) racial discrimination (Section 387); and (4) inciting national, racial and religious hatred and intolerance (Section 317).

Hate speech in Serbia is also regulated by media regulations. Article 75 of the Law on Public Information and Media prohibits hate speech:

Ideas, opinions, or information published in the media must not incite discrimination, hatred or violence against a person or group of persons because of their belonging or not belonging to a certain race, religion, nation, gender, or because of their sexual orientation or other personal property, regardless of whether a criminal offense was committed by publishing it.

(Law on Public Information and Media, Art. 75)

However, hate speech does not exist if that information is part of an objective journalistic report and if there was an intention to critically point to discrimination, hatred, or violence against a person or group, or to phenomena that represent or can represent incitement to such behaviour (Law on Public Information and Media, Article 76).

Hate speech is also prohibited by the Law on Electronic Media. This Law prescribes that:

the regulator ensures that the program content of the media service provider does not contain information that encourages, in an open or covert manner, discrimination, hatred or violence because of race, skin color, ancestry, citizenship, national affiliation, language, religious or political beliefs, gender, gender identity, sexual orientation, property status, birth, genetic characteristics, health status, disability, marital and family status, convictions, age, appearance, membership in political, trade union and other organizations and other real or assumed personal characteristics.

(Law on Electronic Media, Art. 51)

To date, no special law on banning hate speech on the Internet has been passed in Serbia.

RESULTS AND DISCUSSION

The legal regulation of hate speech is largely related to the issue of freedom of expression as one of the pillars of a democratic society, and a basic prerequisite for ensuring the protection of individuals' other human rights. In many countries, freedom of expression enjoys constitutional protection. As a fundamental right, freedom of expression includes the right of every individual to express his opinion in any available way, or to remain silent, as well as the right to be informed about what is happening around the world.

In the EU, freedom of expression is explicitly recognised in Article 2 of the EU Treaty. On the other hand, Article 11 of the Charter of Fundamental Rights of the EU stipulates that:

everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.

(Charter of Fundamental Rights, Art. 11)

However, Article 52 of the EU Charter contains a general clause on the possibility of limiting freedoms and rights when they conflict with other rights.

Paragraph 153 of the General Data Protection Regulation, no. 2016/679, confirms that the reconciliation between the data protection framework defined by the Regulation and the rules on the protection of freedom of expression is the task of Member States, allowing special exceptions regarding the processing of personal data exclusively for journalistic purposes or for the needs of academic, artistic or literary expressions.

Freedom of expression is contained in Article 10 of the European Convention on Human Rights. However, Article 10 (2) of this Convention prescribes cases in which freedom of expression may be restricted in order to strike a balance with other conflicting rights. The protection of this freedom was also developed by the European Court of Human Rights in its rich judicial practice. In this sense, the ECHR took the position that the restriction of freedom of expression is legitimate if the following three conditions are met:

- 1) the interference must be prescribed by law;
- 2) it must pursue a legitimate aim as stated in Article 10;
- 3) it must be necessary in a democratic society, which implies verifying whether the national intervention corresponds to a 'pressing social need'.

(Handyside vs. United Kingdom, para. 48)

Freedom of expression is legally protected in almost all European countries. Most national constitutions include this freedom among the general principles related to the rights of citizens. Constitutional provisions make a clear distinction between freedom of expression and freedom of the media in only a few countries (e.g. Belgium, Greece, Serbia, and Romania). However, the need to balance freedom of expression is also recognised in national constitutions. For example, Article 46 of the Constitution of RS guarantees "freedom of thought and expression, as well as the freedom to seek, receive and disseminate information and ideas by speech, writing, image or otherwise". However, the same regulation stipulates that:

freedom of expression can be limited by law, if it is necessary to protect the rights and reputation of others, preserve the authority and impartiality of the court, and protect public health, morals of a democratic society, and national security of the Republic of Serbia.

(The Constitution of the RS, Art. 46)

Freedom of expression is one of the central human rights, and is considered necessary for the realisation and protection of all human rights, and for the functioning of a democratic constitutional state. However, freedom of expression is not absolute (Milojević, Surčulija, 2016). The difference between freedom of expression and dissemination, that is, incitement to discrimination, hatred, and threats to minorities, which must be prevented and punished, requires careful consideration. Freedom of expression certainly includes indecent and even offensive speech that insults the honour and reputation of those affected but still does not fall under hate speech. Conversely, statements from which no explicit messages of hate can be inferred may also constitute hate speech. According to the Committee on the Elimination of Racial Discrimination (CERD), the following facts must be taken into account when assessing whether a statement constitutes hate speech:

- Content and form of speech;
- The existing social, economic, and political climate in which the speech took place, as well as the existing patterns of discrimination against the minority in question (for example in relation to asylum seekers, groups of foreigners, sexual minorities, etc.);
- The position of the speaker in society or in the relevant media (e.g. politicians, leaders); and
- The reach of speech (internet or mainstream media etc.).

Another thing to consider is the purpose of the speech. Here the Committee points out that, for example, speaking in defence of human rights or certain groups should not be criminalised.

CONCLUSION

Social media is a big part of many people's lives today. Sharing information and communicating through social media platforms has made human life much easier. Today, information is not only obtained from traditional media, but also through social media. Namely, media companies publish information on social networks and share that information with users. It is no longer necessary to wait for news on television or radio. Today's breaking news is mostly spread through social media. In addition, users can quickly and easily react to information, and become broadcasters themselves by sharing information. Additionally, users can communicate with the world without restrictions.

Spreading and sharing information on social media are associated with certain challenges. Namely, social media greatly influence the formation of individuals' opinions. However, the problem is that fake news and hate speech are often spread and shared on social media. Regulating harmful speech in online spaces requires drawing the line between legitimate free speech and hate speech. Freedom of speech is protected by major international human rights treaties and by the constitutions of most countries around the world.

However, in today's digitised and increasingly polarised world, the current question is how to legally regulate hate speech while simultaneously protecting freedom of expression. The right to free speech also includes speech that someone might find deeply offensive. However, the right to freedom of expression is not an absolute right. States may limit this right in certain exceptional circumstances. However, all restrictions must meet the following criteria: (a) restrictions must be provided by law; (b) restrictions must pursue a legitimate aim as detailed in Article 19 (3) of the International Covenant on Civil and Political Rights; and (c) restrictions must be necessary and proportionate to the goal pursued. In addition, states must restrict speech that amounts to advocacy of hatred that incites discrimination, hostility, or violence based on protected personal characteristics.

Meeting the aforementioned criteria under which freedom of speech can be restricted faces numerous challenges in the online space. Social media offers a unique space for the expression of opinion, encouraging public debate and information exchange, and strengthening the space for civic engagement. However, hate speech and abuse in the online space have led to increased pressure on social media companies to control the content posted and shared by their users. Social media platforms make decisions about what is acceptable to express online, which calls into question the protection of freedom of expression in the online space. In addition, it is problematic to hold social media platforms responsible for the content shared by their users. Furthermore, often the content moderation policies of social media platforms are not clear and accessible enough for their users to assess what is and is not allowed on the platform.

The legal regulation of hate speech and content moderation in the online space is highly complex. Certain European countries, such as Germany and France, have adopted regulations on suppressing hate speech on the Internet which are, in our opinion, too broad and unclear. On the one hand, these regulations impose serious obligations on social media in the fight against hate speech on the Internet, while on the other hand, they have serious implications for freedom of expression. In our opinion, before passing regulations on suppressing hate speech on the Internet, it is necessary to take measures aimed at solving the basic causes of discrimination, as well as social problems that contribute to the spread of hatred on the Internet. In this sense, state institutions must actively promote tolerance and equality in cooperation with other actors such as civil society organisations, media, religious leaders, and other social actors. The fight against hate speech requires the protection of freedom of expression. Existing international human rights norms, including the Rabat Action Plan, provide guidance on how to address hate speech in a manner consistent with freedom of expression. This includes calling on political and religious leaders, officials, and the media not only to refrain from hate speech but to actively reject it and speak out against it.

REFERENCES

- Bleich, E. (2011). The Rise of Hate Speech and Hate Crime Laws in Liberal Democracies, *Journal of Ethnic and Migration Studies*, Vol. 37, No. 6, 917-934.
- Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), 297-326.
- Brown, A. (2017). What is hate speech? Part 1: The Myth of Hate, *Law and Philosophy*, 419-468.
- Brugger, W. (2002). Ban on or protection of hate speech-some observations based on German and American law. *Tulane European and Civil Law Forum*, 17, 1-21.
- Davenport, S. W., Bergman, S. M., Bergman, J. Z., Fearrington, M. E. (2014). Twitter versus Facebook: Exploring the role of narcissism in the motives and usage of different social media platforms. *Computers in Human Behavior*, 32, 212-220. <https://doi.org/10.1016/j.chb.2013.12.011>
- Fagan, B. (2000). Rice v. Paladin Enterprises: Why Hit Man is beyond the pale, *Chicago-Kent Law Review*, 76(1), 603-636.
- Fleischhack, J., (2017). *Der „Hass“ der vielen Formen*. In Kaspar, K., Gräber, L., Riffi A. Online Hate Speech. Perspektiven auf eine neue Form des Hasses. Schriftenreihe zur digitalen Gesellschaft NRW, Bd. 4. Düsseldorf, München: Kopaed, 23-29.
- Dimovski, D. (2021). Case law of the European Court of Human Rights of hate crimes. *Teme*, 2/2021, 739-755, <https://doi.org/10.22190/TEME201021042D>.
- Duggan, M., Smith, A. (2013). *Cell Internet Use 2013*. Washington, DC: Pew Research Center, https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_CellInternetUse2013.pdf
- Geschke, D., Klaßen, A., Quent, M., Christoph R., (2019): *#Hass im Netz: Der schleichende Angriff auf unsere Demokratie. Eine bundesweite repräsentative Untersuchung*. Institut für Demokratie und Zivilgesellschaft.
- Kaplan A.M., Haenlein M. (2010). Users of the world, unite! The challenges and opportunities of Social Media, *Business Horizons*, 59-68, doi:10.1016/j.bushor.2009.09.003
- Kiska, R. (2012). Hate speech: A comparison between the European Court of Human Rights and the United States Supreme Court jurisprudence, *Regent University Law Review*, Vol. 25, 107-151.
- Meibauer, J. (2013). *Hassrede – von der Sprache zur Politik*. In: Ders. (Hrsg.): *Hassrede/HateSpeech. Interdisziplinäre Beiträge zu einer aktuellen Diskussion*. Gießen: Gießener Elektronische Bibliothek.
- Milenković, D. (2010), *Ljudska i manjinska prava, sloboda izražavanja i uloga medija u uslovima sukoba i napetosti* [Human and minority rights, freedom of expression and the role of the media in conditions of conflict and tension], Priručnik za rad lokalnih medija, Niš, 80.
- Milinković, B. (1996), *Medijske slobode - prava i ograničenja* [Media freedom - rights and restrictions], Medija centar, Beograd, 84.
- Milojević, Surčulija, J. (2016), *Dozvoljenost ograničenja slobode izražavanja u skladu sa međunarodnim evropskim instrumentima i medijskim zakonodavstvom Republike Srbije*, doktorska disertacija, [Permissible restrictions on freedom of expression in conformity with European law and regulation and media legislation of the Republic of Serbia, PhD dissertation], Pravni fakultet u Beogradu, www.nardus.mpn.gov.rs
- Mihajlov Prokopović, A. (2018). Internet social networks and traditional media. *Teme*, 4/2018, 1081-1105, DOI Number 10.22190/TEME1804081M.

- Nikolić, M.P. (2018). *Govor mržnje u internet komunikaciji u Srbiji*, doktorska disertacija [Internet hate speech in Serbia, PhD dissertation], Fakultet političkih nauka. Beograd, www.fpn.bg.ac.rs.
- N. Petrušić, N. (2012). Građanskopravni instrumenti za suzbijanje govora mržnje koje se šire posredstvom medija, [Civil rights instruments for suppressing hate speech spread through the media], *Zbornik radova Pravnog fakulteta u Nišu*, LXI, 75-97.
- Prokopović, Mihajlov, A. (2016). *Transformacija tradicionalnih medija u novom tehnološkom okruženju*, doktorska disertacija [Transformation of traditional media in a new technological environment, PhD dissertation], Beograd, Fakultet političkih nauka. www.nardus.mpn.gov.rs
- Reinecke L., Trepte S. (2014). Authenticity and well-being on social network sites: A two-wave longitudinal study on the effects of online authenticity and the positivity bias in SNS communication, *Computers in Human Behavior*, 30, 95-102, doi:10.1016/j.chb.2013.07.030.
- Schmidt, J.H. (2017). *Meinungsbildung in und mit sozialen Medien*. In Social Media, Springer VS. Wiesbaden, 61-74.
- Sponholz, L. (2018). *Hate Speech*. In Hate Speech in den Massenmedien. Theoretische Grundlagen und empirische Umsetzung, Springer Fachmedien, Wiesbaden, 31-92.
- Stefanowitsch, A. (2015). *Was ist überhaupt Hate Speech?* In Amadeu Antonio Stiftung, <https://www.amadeu-antonio-stiftung.de/w/files/pdfs/hatespeech.pdf>
- Veljanovski, R. (2010). Izazovi i mogućnosti online novinarstva [Challenges and possibilities of online journalism], *Sloboda izražavanja na internetu*. Fond za otvoreno društvo, 41-45.
- Vučetić, D., Bončić, G., & Pešić, P. (2016). *The Liability of Social Networks for Extreme Forms of Speech*, Facta Universitatis, Law and Politics, Vol.14 , No 1, 1-14
- Vučković, J. (2022a). Uloga interneta u pružanju medijskih usluga [The Role of the Internet in providing Media Services], *Sadašnjosti budućnost uslužnog prava*. Pravni fakultet Kragujevac, Institut za pravne i društvene nauke, 2022, 555-572. doi: 10.46793/XVIIIIMajsko.555V.
- Vučković, J. (2022b). *Medijsko pravo* [Media law], Medivest, Niš, 368.
- Vučković, J. (2021a). Digitalni mediji i medijsko zakonodavstvo Republike Srbije [Digital media and media legislation of the Republic of Serbia] *Zbornik radova Pravnog fakulteta u Nišu*. 2021, god. 60, br. 92, 195-217, doi: 10.5937/zrpfno-33993.
- Vučković, J. (2021b). Neka pitanja digitalizacije medijskih usluga [Some issues of digitisation of media services], *Usluge i vladavina prava*, Zbornik referata sa Međunarodnog naučnog skupa, Pravni fakultet u Kragujevcu, Institut za pravne i društvene nauke, 2021. [521]-537. doi:10.46793/UV21.521V.
- Zekavica, R. (2019). Normativni i institucionalni okvir zabrane govora mržnje u Republici Srbiji [Normative and institutional framework for prohibition of hate speech in the Republic of Serbia], *Fondacija Centar za javno pravo*. No 37. 39-48, http://www.fcjp.ba/analize/Sveske_zajavno_pravo_broj-37.pdf

Regulations

- Audiovisual Media Services Directive 2010/13/EU: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0013>.
- Directive on electronic commerce 2000/31/EC: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32000L0031>.

- Zakon o zabrani diskriminacije [Law on Prohibition of Discrimination], "Sl.glasnik RS", br. 22/2009 i 52/2021. https://www.paragraf.rs/propisi/zakon_o_zabrani_diskriminacije.html.
- Zakon o javnom informisanju i medijima [Law on Public Information and Media], "Sl.glasnik RS", br. 83/2014, 58/2015, 12/2016-autentično tumačenje. https://www.paragraf.rs/propisi/zakon_o_javnom_informisanju_i_medijima.html.
- Zakon o elektronskim medijima [Law on Electronic Media] "Sl.glasnik RS", br.83/2014, 6/2016-dr.zakon i 129/2021. https://www.paragraf.rs/propisi/zakon_o_elektronskim_medijima.html.
- Network Enforcement Act: https://www.bmj.de/DE/Themen/FokusThemen/NetzDG/NetzDG_EN_node.html
- Online Harms White Paper: <https://www.gov.uk/government/consultations/online-harms-white-paper>. OSCE, Hate crime reporting: <https://hatecrime.osce.org/index.php/>.
- Public Order Act 1986: <https://www.legislation.gov.uk/ukpga/1986/64/contents>.
- Ustav Republike Srbije [The Constitution of the Republic of Serbia] "Sl.glasnik RS", br. 98/2006, 16/2022. https://www.paragraf.rs/propisi/ustav_republike_srbije.html.

Case Law

- European court of human right, Case Delfi AS v. Estonia - 64569/09: [https://hudoc.echr.coe.int/fre#{%22itemid%22:\[%22002-8960%22\]}](https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22002-8960%22]}).
- European court of human right, Case OOO Flavus a. o. v. Russia: [https://hudoc.echr.coe.int/fre#{%22itemid%22:\[%22001-203178%22\]}](https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22001-203178%22]}).
- European court of human right, Case Handyside v. The United Kingdom: [https://hudoc.echr.coe.int/eng#{%22dmdocnumber%22:\[%22695376%22\],%22itemid%22:\[%22001-57499%22\]}](https://hudoc.echr.coe.int/eng#{%22dmdocnumber%22:[%22695376%22],%22itemid%22:[%22001-57499%22]}).

ГОВОР МРЖЊЕ И ДРУШТВЕНИ МЕДИЈИ

Јелена Вучковић, Соња Лучић

Универзитет у Крагујевцу, Правни факултет, Крагујевац, Србија

Резиме

Регулисање штетног говора у онлајн простору захтева повлачење границе између легитимне слободе говора и говора мржње. Слобода говора је заштићена уставима већине земаља широм света, као и главним међународним споразумима о људским правима. Међутим, у данашњем дигитализованом и све више поларизованом свету, актуелно је питање како правно регулисати говор мржње уз истовремену заштиту слободе изражавања. Испуњавање наведених критеријума под којима се може ограничити слобода говора наилази на бројне изазове у онлајн простору. Друштвени медији нуде јединствен простор за изражавање мишљења, подстичући јавну дебату и размену информација и јачајући простор за грађанско ангажовање. Међутим, говор мржње и злостављање у онлајн простору довели су до повећаног притиска на компаније друштвених медија да контролишу садржај који објављују и деле њихови корисници. Платформе друштвених медија доносе одлуке о томе шта је прихватљиво изразити на онлајн мрежи, што доводи у питање заштиту слободе изражавања у онлајн простору. Поред тога, проблематично је сматрати платформе друштвених медија одговорним за садржај који деле њихови корисници. Приде, политике модерирања садржаја платформи друштвених меди-

ја често нису довољно јасне и доступне својим корисницима, те је корисницима тешко да процене шта јесте, а шта није дозвољено на платформи.

Правно регулисање говора мржње и модерирања садржаја у онлајн простору је у великој мери сложено. Поједине европске државе попут Немачке и Француске су донеле, по нашем мишљењу, прешироке и нејасне прописе о сузбијању говора мржње на интернету. Ови прописи са једне стране намећу озбиљне обавезе друштвеним медијима у борби против говора мржње на интернету, док са друге стране имају озбиљне импликације по слободу изражавања. По нашем мишљењу, пре доношења прописа о сузбијању говора мржње на интернету неопходно је предузети мере у циљу решавања основних узрока дискриминације, као и друштвених проблема који доприносе ширењу мржње на интернету. У том смислу, државне институције морају активно промовисати толеранцију и једнакост у сарадњи са другим актерима као што су организације цивилног друштва, медији, верски лидери и други друштвени актери. Борба против говора мржње захтева заштиту слободе изражавања. Постојеће међународне норме о људским правима, укључујући Рабатски акциони план, дају смернице о томе како се позабавити говором мржње на начин који је у складу са слободом изражавања. То укључује позивање политичких и верских лидера, званичника и медија да се не само уздрже од говора мржње, већ и да га активно одбацују и говоре против говора мржње.