

Petar Antić\*

## SANCTIONING HATE SPEECH ON THE INTERNET: IN SEARCH OF THE BEST APPROACH

**Abstract:** *The borderless nature of the Internet, different national approaches to the understanding of what constitutes hate speech, as well as the danger of restrictions on the freedom of speech, make it difficult to develop appropriate mechanisms against this phenomenon. The limitations of international law in providing a universal definition of hate speech, due to the different national approaches to freedom of expression, have thwarted attempts to produce an effective international treaty in order to deal with this issue. Imposing obligations on Internet portals to establish self-regulatory mechanisms for removing hate speech content has raised concerns of non-competent censorship and potential limitations of the freedom of expression. This paper focuses on the challenges encountered in the struggle against hate speech online and possible mechanisms for combating this phenomenon.*

**Key words:** hate speech, freedom of expression, Internet, online, media, dissemination.

### 1. INTRODUCTION

The Internet plays an important role, surpassing all popular media that had previously promoted violent extremism to millions of people. Public forums such as blogs, message boards, chat rooms, and websites have led to an increase in controversial hate speech, which flourishes in the insufficiently regulated Internet space.<sup>1</sup>

According to research on the presence of hate speech among teenagers on social networks in the USA in 2018, 52 percent of those surveyed said that they often or sometimes encounter racist hate speech on social media.<sup>2</sup> In Germany and United States of America, a correlation was

---

\* Doctoral Candidate, Union University Law School Belgrade; e-mail: petar.antic@live.com

1 Keum, T. B., Miller, J. M., 2018, Racism on the Internet: Conceptualization and Recommendations for Research, *Psychology of Violence*, Vol. 8, No. 6, pp. 783–791.

2 Statista Research Department, 2022, *U.S. teens encountering hate speech on social media 2018, by type*, (<https://www.statista.com/statistics/945392/teenagers-who-encounter-hate-speech-online-social-media-usa/>, 13. 09. 2022).

found between anti-immigrant and anti-Muslim tweets and attacks on them.<sup>3</sup> Similarly, a correlation was found between the Facebook posts by the far-right party Alternative for Germany and attacks on refugees.<sup>4</sup> On the other hand, the Internet was an important tool for the rise of democracy and the development of freedom of expression.<sup>5</sup> In countries where traditional media are controlled, the Internet proves to be of great importance for exercising the freedom of expression.<sup>6</sup> However it has also been noted that the lack of mechanisms against hate speech and incitement to violence can strengthen authoritarian regimes.<sup>7</sup>

The specific nature of the Internet renders all attempts to regulate online content extremely challenging, especially if freedom of expression is to be respected.<sup>8</sup> Due to different national approaches to content that incites hatred on the Internet, as well as varying legal frameworks that regulate the issues of hate speech and freedom of expression, the criminal legislation of a single country has limited impact and thus is unable to resolve the problem of hate speech online on its own.

In addition to legal mechanisms aimed against hate speech online, self-regulation mechanisms are increasingly being used, according to which social network companies are responsible for the content and remove objectionable messages. However, the self-regulation is limited and carries the risk of insufficient expertise. The weakness of self-regulation is further highlighted by the emergence of legal initiatives introducing it as an obligation of social media companies. In 2017, Germany passed a special law on online hate speech, the Network Enforcement Act, known as NetzDG, which requires social media platforms such as Facebook to quickly remove inflammatory material or face heavy fines.<sup>9</sup> Critics of the

---

3 Wilson, A. R., Land, M., 2021, Hate Speech on Social Media: Content Moderation in Context, *Connecticut Law Review*, Vol. 52, No. 3, p. 1043, pp. 1029–1242.

4 Laub, Z., 2019, *Hate Speech on Social Media: Global Comparisons*, Council on Foreign Relations, (<https://www.cfr.org/backgroundunder/hate-speech-social-media-global-comparisons>, 13. 09. 2022).

5 Cooke, L., 2007, Controlling the Net: European Approaches to Content and Access Regulation, *Journal of Information Science*, Vol. 33, No. 3, p. 362, pp. 360–376.

6 Chen, G., 2022, How Equalitarian Regulation of Online Hate Speech Turns Authoritarian: A Chinese Perspective, *Journal of Media Law*, Vol. 14, No. 2, p. 168, pp. 159–179.

7 Goldbert, A., 2015, Hate Speech and Identity Politics in Germany, 1848–1914, *Central European History*, Vol. 48, No. 4, p. 26, pp. 480–497.

8 Ubanga, C., 2016, *Hate Speech in Cyberspace: Why Education is Better than Regulation*, Social Science Research Network, p. 1.

9 Heldt, A., 2019, Reading Between the Lines and the Numbers: An Analysis of the First NetzDG Reports, *Internet Policy Review*, Vol. 8, No. 2, p. 1.

law claim its measures could stifle political speech or be used as a model for authoritarian governments to suppress freedom of speech.<sup>10</sup>

Following the introduction of restrictions against hate speech online through self-regulatory mechanism, in January 2021 Donald Trump's Twitter account was permanently deleted to prevent further incitement of violence, yet the platform's new owner, Elon Musk, restored the account on 19 November 2022, after asking the platform's users whether Donald Trump's account should be reinstated. Such examples show that it is necessary to find an adequate mechanism against hate speech that would adequately define who and in what manner decides the fine line between freedom of expression and hate speech.<sup>11</sup>

The focus of the paper is on the ways of approaching the fight against hate speech online in the context of the absence of a universal definition and the borderless nature of the Internet. It will not deal with the analysis of restrictions on the freedom of expression and the search for the best approach to defining hate speech. The intention, instead, is to analyze and understand the challenges of fighting hate speech online in order to assess success of legal and other mechanisms aimed against this phenomenon. The first question deals with whether a universal definition of hate speech is possible and the explanation of the challenges of reaching a common agreement. After elaborating on the challenges of finding the universal definition, the next question deals with the ways in which the nature of the Internet affects the fight against online hate speech and limits national legal mechanisms. The fourth section explores cross-border initiatives specifically targeting online hate speech. The fifth part analyzes self-regulation and the new legal initiatives that introduce the obligation of social media to introduce a system of self-regulation. Finally, the conclusion presents the findings and the possible future direction of the development of international cooperation in the protection against hate speech online.

## 2. THE DEFINITION OF HATE SPEECH

No consensus has been achieved among states when it comes to the definition of hate speech. Divergent historical, political, cultural, moral, and constitutional values impede intentions to regulate hate speech at the

---

10 Leerssen, P., Tworek H., 2019, *An Analysis of Germany's NetzDG Law*, Transatlantic Working Group, p. 4.

11 Holmberg, E. S., 2021, Lessons from Trump's Suspension: How Twitter Should Clarify and Strengthen Its "Public Interest" Approach to Moderating Leaders' Violence-Inspiring Speech, *Harvard Journal of Law & Technology*, Vol. 35, No. 1, p. 333, pp. 310–334.

international level.<sup>12</sup> The meaning of hate speech differs depending on the ideological framework within which one operates. Some content can be considered hate speech by some and not by others, based on their respective definitions and their understanding of freedom of expression.

The most widely accepted definition of hate speech is given by Nockleby: “any communication that disparages a target group of people based on some characteristic such as race, colour, ethnicity, gender, sexual orientation, nationality, religion, or other characteristic.”<sup>13</sup> Similarly, the UN Strategy and Plan of Action on Hate Speech defines hate speech as “any kind of communication in speech, writing or behaviour that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.”<sup>14</sup> The Council of Europe recommendation on hate speech specified that “the term ‘hate speech’ shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.”<sup>15</sup> However, none of these three definitions could be accepted as a universal definition of hate speech under international human rights law, due to the different understandings of what statements constitute disparagement of a target group and different understandings of freedom of expression.

The inexistence of a universal definition of hate speech has caused numerous challenges to the development of international standards regulating this phenomenon. Article 19 para. 2 of the International Covenant on Civil and Political Rights (ICCPR) ensures everyone’s right to freedom of expression, but para. 3 defines special duties and responsibilities by providing special restrictions for respecting the rights or reputation of others, as well as protection of national security, public order, public health and morals.<sup>16</sup> Nevertheless, there is no universal approach to the definition of

- 
- 12 Brown, A., 2017a, What Is Hate Speech? Part 1: The Myth of Hate, *Law and Philosophy*, Vol. 36, Issue 4, p. 5, pp. 297–326.
  - 13 Gilbert, O. de, Perez, N., Pablos, A. G., Cuadros, M., 2018, Hate Speech Dataset from a White Supremacy Forum, *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, Association for Computational Linguistics, p. 11; Nockleby, T., Hate Speech, in: Levy, L., Karst, K. (eds.), 2000, *Encyclopedia of the American Constitution*, Macmillan Reference USA, pp. 1277–1279.
  - 14 United Nations, 2019, *United Nations Strategy and Plan of Action on Hate Speech*, p. 2.
  - 15 Council of Europe, Committee of Ministers, Recommendation No. R (97) 20 (30 October 1997).
  - 16 UN General Assembly, International Covenant on Civil and Political Rights, UN doc. Resolution 2200A (XXI) (16 December 1966).

what constitutes speech that requires restriction. Furthermore, according to Article 20 para. 2 of the ICCPR, state parties are required to prohibit by law “any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.” However, Articles 19 and 20 of the ICCPR have been subjected to a number of reservations and declarations by states.

In comparison to the ICCPR, the regional treaties, the European Convention for the Protection of Human Rights and Fundamental Freedoms, the African Charter on Human and Peoples Rights, and the American Convention on Human Rights declare hate speech to be outside the protection of freedom of expression, but there is no provision that corresponds to Article 20 of the ICCPR regarding the criminalization of such acts.<sup>17</sup> All three regional treaties prohibit discrimination on a number of grounds, including race, ethnicity, color, sex, language, religion, or any other status – yet none of them contains a definition of hate speech.

The effort to find a universal definition of hate speech has been further complicated by the fact that Article 4 of the International Convention on the Elimination of All forms of Racial Discrimination (ICERD) defines categories of activities which need be criminalized: “dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof.”<sup>18</sup> However, there is no international consensus regarding the requirements stipulated in Article 4. Many states, including the United States of America, entered their reservation.<sup>19</sup>

On the one hand, Article 10 of the European Convention on Human Rights protects freedom of expression, while on the other, Article 17 prohibits the abuse of this right, without defining hate speech. The European Court of Human Rights (ECtHR) follows this approach, admitting regulation of hate speech in accordance with the criteria for freedom of expression only if the expression promotes a certain level of violence and when the interference is necessary for democratic society.<sup>20</sup> However, in cases of

17 ECHR, Art. 10; ACHPR, Art. 9; ACHR, Art. 13.

18 UN General Assembly, International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), UN doc. Resolution 2106 (21 December 1965).

19 Watson, K. M., 2020, The United States’ Hollow Commitment to Eradicating Global Racial Discrimination, *Human Rights*, ([https://www.americanbar.org/groups/crsj/publications/human\\_rights\\_magazine\\_home/black-to-the-future-part-ii/the-united-states--hollow-commitment-to-eradicating-global-racial/](https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/black-to-the-future-part-ii/the-united-states--hollow-commitment-to-eradicating-global-racial/), 13. 09. 2022).

20 Greer, S., 2000, The Margin of Appreciation: Interpretation and Discretion Under the European Convention on Human Rights, *Human Rights Files*, No. 17, Council of Europe, p. 9, pp. 5–58.

hate speech, the Court leaves to national authorities a “margin of appreciation” to research the necessity of restriction for democratic society.<sup>21</sup> The Court will assess whether the incitement restriction is in accordance with the tripartite test under Article 10 – “whether the restriction was prescribed by law, whether it served a legitimate purpose (e.g. the reputation or the rights of others) and whether the restriction was necessary.”<sup>22</sup> The European Court of Human Rights does not make quality assessments of the incitement restriction, instead this is left to the national authorities.<sup>23</sup> Unlike the European Court of Human Rights, the CERD and the Human Rights Committee have a mandate to carry out quality assessment by checking the way that the national law is constructed, with an analysis of its functioning in practice.<sup>24</sup> However, all these attempts to fight hate speech through international treaties have remained weak due to the different understandings of what comprises this phenomenon.

The absence of a universal definition of hate speech, caused by the varying approaches to freedom of expression in different regions, has led to the adoption of different legal measures for dealing with such content. While European states have strict policies against hate speech, the US First Amendment defends most forms of speech by protecting the freedom of expression.<sup>25</sup> In the USA, it is only advocating the use of force with imminent lawless action that is prohibited.

The US government can regulate speech-based content only in a few limited areas which are “of such slight social value as a step to the truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.”<sup>26</sup> Hence, the freedom of expression provided by the First Amendment excludes obscenity (*Roth v. United States*), child pornography (*New York v. Ferber*), defamation (*Beauharnais v. Illinois*), fighting words (*Chaplinsky v. New Hampshire*), and true threats

---

21 ECtHR, *Handyside v. the United Kingdom*, No. 5493/72, Judgment of 7 December 1976, para. 48.

22 ECtHR, *Jersild v. Denmark*, No. 15890/89, Judgment of 23 September 1994 para. 37; ECtHR, *Delfi AA v. Estonia*, No. 64569/09, Judgment of 16 June 2016, para 65; Temperman, J., 2015, *Religious Hatred and International Law*, Cambridge, Cambridge University Press, p. 147.

23 ECtHR, *Pavel Ivanov v. Russia*, No. 35222/04, admissibility decision of 20 February 2007, section The Law, para. 1; ECtHR, *Lehideux and Isorni v. France*, No. 55/1997/839/1045, Judgment of 23 September 1998, para. 50.

24 Temperman, J., 2015, *Religious Hatred and International Law*, Cambridge, Cambridge University Press, p. 148.

25 U.S. Constitution, Bill of Rights, First Amendment, Religion, Speech, Press, Assembly, Petition (1791).

26 U.S. Supreme Court, *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942), Decision of 9 March, 1942, at 572.

(*Watts v. United States*).<sup>27</sup> The Supreme Court's landmark decision in *Virginia v. Black*, defined a "true threat" as "statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or a group of individuals."<sup>28</sup> In *Chaplinsky*, the Supreme Court defined that "fighting words" are words which "by their very utterance inflict injury or tend to incite an immediate breach of the peace."<sup>29</sup>

In contrast to the USA, in European countries the area of protection against hate speech is much wider and includes incitement to national or racial discrimination.<sup>30</sup> Furthermore, the protection, with certain differences among the European countries, includes criminalization of the denial of the Holocaust, as well as of the use of writings and symbols to incite hatred, considering them to be hate speech.<sup>31</sup>

On the other hand, the European Union's approach to the criminalization of the Holocaust and genocide denials is in contravention of the UN's approach to freedom of expression. In a report submitted in October 2019, the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, recalling General Comment No. 34 (2011) of the Human Rights Committee, pointed out that "laws that 'penalize the expression of opinions about historical facts are inconsistent' with Article 19 of the Covenant," questioning laws that criminalize Holocaust denial and other crimes and similar laws that are often justified by referring to hate speech.<sup>32</sup> The Human Rights Committee has noted that opinions that are "erroneous" and "a misinterpretation of past events" cannot be subject to general prohibition, and "any restrictions on the expression of such opinion 'should not go beyond what is permitted'

27 U.S. Supreme Court, *Roth v. United States*, No. 582, Decision of 24 June 1957; U.S. Supreme Court, *New York v. Ferber* No. 81–55, Decision of 2 July 1962; U.S. Supreme Court, *Beauharnais v. Illinois*, No. 118, Decision of 28 April 1952; U.S. Supreme Court, *Watts v. United States*, No. 21528, Decision of 25 September 1968.

28 Barron, J. A., 2019, Internet Access, Hate Speech and the First Amendment, *First Amendment Law Review*, Vol. 18, No. 1, p. 8, pp. 1–24; U.S. Supreme Court, *Virginia v. Black*, No. 01–1107, Decision of 7 April 2003, para. 343.

29 U.S. Supreme Court, *Virginia v. Black*, No. 01–1107, Decision of 7 April 2003, para. 343.

30 Khan, A. R., 2013, Why Do Europeans Ban Hate Speech? A Debate Between Karl Loewenstein and Robert Post, *Hofstra Law Review*, Vol. 41, No. 3, p. 32, pp. 545–585.

31 Loi du 29 juillet 1881 sur la liberté de la presse, Art. 24; Council Framework Decision 2008/913/JHA of 28 November 2008, on combating certain forms and expressions of racism and xenophobia by means of criminal law.

32 UN General Assembly, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN doc. A/74/486 (9 October 2019), para. 22.

in Article 19(3) or ‘required under Article 20’ of the Covenant on Civil and Political Rights.<sup>33</sup>

National regulations that criminalize the use of symbols for incitement of hatred could further manifest the complexity of defining hate speech universally. For example, according to a State of New York bill, the Confederate battle flag is seen as a symbol of racism, exclusion, oppression and violence towards African Americans, while in European states such, as Germany, the Confederate battle flag tends to be seen as a symbol of independence, rebellion, freedom, and anti-authoritarianism.<sup>34</sup> However, Section 86a of the German Criminal Code criminalizes the use of symbols associated with unconstitutional organizations, such as swastikas, Hitler salutes and Nazi uniforms in public meeting or in publication, including imprisonment for a maximum of three years for offences prescribed in Section 86.<sup>35</sup> Similarly, Article R645–1 of the French Penal Code recognizes as a criminal offence the wearing or public displaying of signs or emblems worn or displayed by members of organizations found guilty under Article 9 of the Statute of the International Military Tribunal, as amended by the London Agreement of 8 April 1945.<sup>36</sup> At the same time, in the USA, neo-Nazi groups wearing Nazi uniforms can freely organize rallies against illegal immigrants.<sup>37</sup>

In order to address the issue of different national approaches to limiting exercising of freedom of expression, in November 2008 the European Union adopted the Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law. The preamble of the Framework Decision acknowledges that a standalone legislative response is not enough for combating racism and xenophobia, and that a broader comprehensive framework, containing various measures including non-regulatory solutions, would be necessary.<sup>38</sup> It is emphasized that full harmonization of criminal law provisions for combating racism and xenophobia may never be possible, owing to different historical, cultural, constitutional and legal differences within the member states of the EU. For that reason, the framework decision is “limited to

---

33 *Ibid.*

34 New York State Senate Bill S8298B, Section 146; Crelling, K., 2017, *The Confederate Battle Flag: Why Is It Perceived so Differently in the US and Europe/Germany*, University of Washington, p. 56.

35 German Criminal Code, *Federal Law Gazette I*, p. 3322 and p. 844, Section 86.

36 Code pénal Version consolidé au 1 janvier 2020, Art. R645–1.

37 Cuevas, S., 2009, Neo-Nazi Rallies Provoke “Anger, Fear”, *NPR* (<https://www.npr.org/2009/11/07/120129726/neo-nazi-rallies-provoke-anger-fear>, 8 December 2022).

38 Council Framework Decision 2008/913/JHA of 28 November 2008, on combating certain forms and expressions of racism and xenophobia by means of criminal law, p. 5.



combating particularly serious forms of racism and xenophobia by means of criminal law.”<sup>39</sup> Among such crimes, the Framework Decision includes publicly inciting to violence or hatred “directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin” and the commission of these crimes by “public dissemination or distribution of tracts, pictures or other material.”<sup>40</sup> On 9 December 2021, the Commission proposed to extend the list of so-called ‘EU crimes’, as laid down in Art. 83 TFEU, by including hate crime and hate speech.<sup>41</sup>

The abovementioned shows that in the search for an appropriate method against hate speech online, the answer should not be sought in the establishment of the universal definition of hate speech. Before embarking on the analysis of the mechanism against hate speech online, bearing in mind the non-existing definitions of hate speech and a variety of national regulations, the following will explore the nature of the Internet and the implications of this nature on cross-border cases.

### 3. CHALLENGES OF THE FIGHT AGAINST HATE SPEECH IN THE INTERNET ENVIRONMENT

In potential cross-border cases, the issue of hate speech online can be easily resolved when the national systems recognize common grounds of hate speech, but problems arise when Internet users in one country have access to web content stored in a country that stipulates different grounds of hate speech. For example, a website that contains Holocaust denial content can be easily accessed by users from the states where denial of the Holocaust is prohibited.<sup>42</sup>

The word “internet” itself can lead to the conclusion that there is a network of computers. This network, however, goes beyond a single network of computers and is rather a “network of networks” or an interconnection of computer networks. The Internet can be considered as a set of service and communication tools including the World Wide Web (WWW), electronic mail (e-mail), discussion groups (mailing lists and newsgroups), chat, and IP2 technology.

39 *Ibid.*, para. 6.

40 *Ibid.*, Art. 1, (b).

41 European Commission Communication COM (2021) 777 final of 9 December 2021 on “A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime”.

42 Alkiviadou, N., 2016, Regulating Internet Hate A Flying Pig?, 7 *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, 3, pp. 216–228.

The World Wide Web (WWW) provides the broadcasting of HTML pages. The WWW stored in any part of the world it can be visible in the whole world. It is very easy to link one web page to another and not necessarily with the knowledge of the author of the latter page.

Electronic mail enables communication between users. The only requirement for its use is knowing the address of the other party. In the message it is also possible to include audio and video material, as is the case with WWW pages. In the case of the distribution of any potential illegal material, e-mail can be used with a possibility to disappear completely from the Internet. This feature of e-mails complicates the control of the distribution of any illegal material. But there are states that oblige IPS to retain data for a certain number of days.<sup>43</sup>

Newsgroups are public discussion forums for Internet users. Unlike e-mail messages, newsgroup messages can be read by anyone. Users' ability to use service providers outside of their area of jurisdiction, only as a conduit of newsgroups' information, complicates the issues of using newsgroups as a form of public discussion and their abuse.

A chat is a service that allows users to communicate by typing messages in real time and everyone who is logged on can see and reply with a message immediately. If a user in a chat room engages in racist speech, the Internet service provider can cancel their account or forbid them from using the chat room in the future. Service providers do not have general control of the communication in the chat service and the eventual cancelling of an account can take place if users report suspected violations to administrators. Certainly, it is more complicated if the service provider is based in a state jurisdiction that applies different standards to hate speech.

The appearance of Web 2.0 technology made the communication even more interactive.<sup>44</sup> With Web 2.0 technology, users can easily collaborate between themselves. Web 2.0 technology enabled the development of web-based communities, social networks and social media applications such as Instagram, Facebook, Twitter, TikTok, the YouTube video sharing platform, the Wikipedia free encyclopedia, and many interactive games. For example, by using the Web 2.0 technology, YouTube, Facebook and Twitter allow users to store and publish audio and video material.

---

43 EU-Directive 2006/24/E on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/E., *Official Journal of the European Union L 105/54*, Art. 5.

44 Yar, M., 2018, A Failure to Regulate? The Demands and Dilemmas of Tackling Illegal Content and Behavior on Social Media, *International Journal of Cybersecurity Intelligence and Cybercrime*, Vol. 1, No. 1, p. 6, pp. 5–20.

The nature of the Internet involves several important participants in establishing responsibility in the case of hate speech. These include users, content providers, service providers, telecommunication network providers and broadcasters. The problem with regulating hate speech online is that these participants are not necessarily in one state. Attempts to regulate hate speech online by national regulations face serious obstacles due to the nature of the Internet and the different national approaches to hate speech.<sup>45</sup>

The problem of different national regulation of hate speech is particularly evident in cross-border cases when attempts by the states to regulate hate speech online were disabled by the legislation of the states in which the websites were located. Two controversial cases, *Yahoo!* and *Zundelsite*, clearly show the limitations of national regulation related to hate speech online.

In *Yahoo!*, the League Against Racism and Anti-Semitism (LICRA) and the Union of French Jewish Students (UEJF) filed a complaint against Yahoo! Inc. and Yahoo! France before the Tribunal de Grande Instance de Paris in May 2000,<sup>46</sup> due to Yahoo! Inc. allegedly hosting an auction website which advertised the sale of thousands of items of Nazi paraphernalia.

The French court ordered Yahoo! Inc. to undertake all necessary measures in order to dissuade and render impossible any access to the auction services for users from France.<sup>47</sup> Upon the adoption of the French court's decision, Yahoo! Inc. introduced monitoring and filtering mechanism, but also filed a complaint with a US court, claiming the French court's lack of jurisdiction, since the website server was based in the USA, where the First Amendment did not allow such restrictions. In November 2001, the US Court ruled that the French judgment represented a violation of the First Amendment.<sup>48</sup>

In *Zundelsite*, a complaint was filed with the Canadian Human Rights Tribunal against Ernest Zündel, a prominent distributor of revisionist neo-Nazi propaganda through his website.<sup>49</sup> The Supreme Court of Canada ordered that the dissemination of hate propaganda on the Zundelsite

45 Walker, C., Akdeniz, Y., 1999, The Governance of the Internet in Europe with Special Reference to Illegal and Harmful Content, *The Criminal Law Review*, p. 2, pp. 15–19.

46 Tribunal de Grande Instance (High Court), *UEJF and Licra v. Yahoo! Inc and Yahoo France*, RG 05308, Decision of 22 May 2000.

47 *Ibid.*, at paras. 1191–1192.

48 District Court for the Northern District of California, *Yahoo!, Inc. v. La Ligue Contre Le Racisme* 169 F. Supp. 2d 1181 (N.D. Cal. 2001), Decision of 7 November 2001, at paras. 1188–1190.

49 Canadian Human Rights Tribunal, *Citron et al. v. Zündel*, T.D. 1/02 2002/01/18, Decision of 18 January 2002, para. 118.

website could not be tolerated, as it was in violation of Section 13 (1) of the Canadian Human Rights Act.<sup>50</sup> In the Tribunal's opinion, restriction of the respondent's freedom of speech was reasonable and justified in a free and democratic society. Nevertheless, the Zundelsite continued to be hosted on the server based in the United States.

The lack of capacity to confront the problem of hate speech at the national level indicates a need for cross-border Internet regulations, which will be analyzed in the following section.

#### 4. INTERNATIONAL INITIATIVES TARGETING ONLINE HATE SPEECH

The failure to formulate a universal definition of hate speech has influenced the development of cross-border initiatives dealing with the fight against hate speech online. In a 2003 report on the fight against racism, racial discrimination, xenophobia and related intolerance, UN Special Rapporteur of the Commission on Human Rights Doudou Diène commended the adoption of the Additional Protocol to the Convention on Cybercrime and expressed hope that a universal document would be adopted, as an additional protocol to the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD).<sup>51</sup> However, an agreement on this important issue has still not been reached.

In contrast to the UN, the Council of Europe developed the first international treaty, the Convention on Cybercrime, which addresses criminal law, as well as procedural aspects of different types of offences targeting computer systems and networks.<sup>52</sup> The Convention does not contain provisions on hate crime, due to the reluctance of some states involved in its development. Because of that, there is an Additional Protocol concerning the criminalization of acts of racism and xenophobia committed via computer systems.<sup>53</sup> The aim of the Protocol is to facilitate the harmonization of substantive criminal law in fighting racism and xenophobia on the Internet and to develop cooperation. The definition of "racist and

---

50 *Ibid.*, para. 156.

51 Secretary-General, *The fight against racism, racial discrimination, xenophobia and related intolerance and the comprehensive implementation of and follow-up to the Durban Declaration and Programme of Action*, UN doc. A/58/313 (22 August 2003), para. 25.

52 Council of Europe, *Convention on Cybercrime*, ETS No. 185, (8 November 2001).

53 Council of Europe, *Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems*, ETS 189 (28 January 2003).

xenophobic material” contained in Article 2 (1) of the additional Protocol, reflects a tendency to “build upon existing national and international (UN, EU) definitions and documents as far as possible.”<sup>54</sup> Article 3, regulating dissemination of racist and xenophobic material via computer systems, requires the member states to adopt legislative and other measures that may be necessary in order to establish as a criminal offence under domestic law the provision of racist material on the Internet. However, the state parties may reserve the right not to attach criminal liability to the conduct referred to in Article 3, in cases where the racist and xenophobic material advocating, promoting or inciting discrimination is not associated with hatred or violence, provided that other effective civil or administrative remedies are available. The fact that the Protocol allows signing and ratifying member states to include reservations in relation to many of its provisions diminishes its relevance. The USA, as an observer state, has signed the Convention on Cybercrime but not the accompanying Protocol. Reacting to a report by the Council of Europe committee of experts, which stated that “unlawful hosting” would be permitted under U.S. law, the USA asserted: “The United States deplores racism and xenophobia, and the violence and other harmful conduct that racist and xenophobic groups often seek to foster. The United States also supports dialogue among Internet users, providers, and others regarding racist and xenophobic content. However, as the Report suggests, there are a number of factors – legal, as well as political, ethical, and technological – that would impose significant constraints on the implementation of any provision restricting racist and xenophobic content on the Internet. Foremost among these factors for the United States is our Constitution’s protection of freedom of speech and expression.”<sup>55</sup>

Due to the limitations of the development of international treaties in this area, emphasis is placed on strengthening cooperation and dialogue. In 2004, the Secretary-General of the United Nations established the Working Group on Internet Governance (WGIG) to support “carrying out the mandate from the World Summit on the Information Society (WSIS) with regard to convening a new forum for multi-stakeholder policy dialogue – the Internet Governance Forum (IGF).”<sup>56</sup> Emphasis was placed on involving all public and private stakeholders, as well as relevant intergovernmental

54 Council of Europe, *Explanatory report of the Additional Protocol concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems*, ETS 189 (28 January 2003), para. 13.

55 Murphy, S. D. 2004, *United States Practice in International Law: Volume 2, 2002–2004*, Cambridge, Cambridge University Press, p. 444.

56 United Nations, *United Nations Establishes Working Group on Internet Governance*, PI/1620 (<https://press.un.org/en/2004/pi1620.doc.htm>, 18. 09. 2022).

and international organizations within the Internet governance model. Despite these efforts, in a 2018 report United Nations Special Rapporteur on Freedom of Expression David Kay highlighted the failure of social media companies to properly address the rise of hate speech and invited them to apply international human rights standards in their content moderation.<sup>57</sup> Bearing in mind the special responsibility of transnational corporations as largest internet providers, it would be useful to develop global instruments to establish direct corporate international responsibility for violations of constitutionally protected human rights before constitutional courts and international public law.<sup>58</sup> However, given the lack of agreement on the universal definition of hate speech, such an initiative would not be successful, and it is more realistic to develop other measures through soft law.

Due to the growing trends of xenophobia, racism and intolerance, UN Secretary-General Antonio Guterres launched the United Nations Strategy and Action Plan on Hate Speech on 18 June 2019.<sup>59</sup> The document highlighted need for a holistic approach in combating hatred and with full respect for freedom of thought and expression was emphasized and the need to develop cooperation with relevant actors, including CSOs, media houses, technology companies and social media platforms.<sup>60</sup> It was concluded that effective online hate speech interventions include education via television, radio, youth conferences, and text messaging campaigns.<sup>61</sup>

In a response to hate speech online, the European Commission agreed the EU Code of conduct on countering illegal hate speech online with four IT companies (Facebook, Microsoft, Twitter and YouTube) on 31 May 2016.<sup>62</sup> Since then, TikTok, LinkedIn, Instagram, Snapchat, Dailymotion, Jeuxvideo.com, Rakuten Viber, and Twitch have joined the Code.<sup>63</sup>

---

57 UN General Assembly, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN doc. A/74/486 (09. 09. 2019).

58 Beširević, V., 2018, Catch Me if You Can: Reflections on Legal (Un)accountability of Transnational Corporations for Human Rights Violations, *Pravni zapisi*, Vol. 9, No. 1, pp. 38, 21–42.

59 Secretary-General, *United Nations Strategy and Plan of Action on Hate Speech: Detailed guidance on implementation for United Nations field presences* ([https://digitallibrary.un.org/record/3889286?ln=zh\\_CN](https://digitallibrary.un.org/record/3889286?ln=zh_CN), 22. 08. 2022).

60 *Ibid.*, p. 9.

61 Windisch, S., Wiedlitzka, S., Olaghere, A., Jenaway, E., 2022, Online Interventions for Reducing Hate Speech and Cyberhate: A Systematic Review, *Campbell Systematic Reviews*, Vol. 17, No. e1133, p. 1.

62 Brown, A., 2017b, What Is so Special about Online (as Compared to Offline) Hate Speech, *Ethnicities*, Vol. 18, No. 3, pp. 15, 297–326.

63 European Commission, 2022, *EU Code of Conduct against online hate speech: latest evaluation shows slowdown in progress* ([https://ec.europa.eu/commission/presscorner/detail/da/ip\\_22\\_7109](https://ec.europa.eu/commission/presscorner/detail/da/ip_22_7109), 01. 12. 2022).

All the signatories of this Code meet regularly to discuss challenges and progress, and participate in monitoring rounds. This mechanism exists as an instrument to support EU policy planning against online hate speech.

The issue of hate speech online has been raised in a number of OSCE high-level conferences and meetings.<sup>64</sup> The OSCE Office for Democratic Institutions and Human Rights aggregates the information and statistics collected by participating states, fully cooperating with the ECRI, the CERD and the European Monitoring Centre on Racism and Xenophobia, as well as relevant NGOs. The Council Decision of November 2004 expressed that “participating states should investigate and, where applicable, fully prosecute violence and criminal threats of violence, motivated by racist, xenophobic, anti-Semitic or other related bias on the Internet”.<sup>65</sup> Furthermore, this decision tasked the OSCE Representative on Freedom of the Media to “continue an active role in promoting both freedom of expression and access to the Internet and [...] to observe relevant developments in participating states”.<sup>66</sup> This involves monitoring and issuing early warnings when legislative or other prohibitions of speech motivated by racist or other bias are discriminatorily or selectively enforced. Furthermore, the emphasis has been increasingly placed on involving the private sector in governance and self-regulatory mechanisms.<sup>67</sup>

This section has pointed out the limitations of international initiatives in the harmonization of substantive criminal law in the fight against racism and xenophobia on the Internet, which is why the focus remains on improving dialogue and consultation mechanisms among all the interested parties. The lack of capacity of the traditional national legal mechanisms and the difficulty of establishing a universal treaty against hate speech online has focused attention on self-regulatory mechanisms. Therefore, the following section will explore such mechanisms.

---

64 *E.g.* Incitement to Hatred vs. Freedom of Expression: Challenges of Combating Hate Crimes Motivated by Hate on the Internet, Warsaw (22 March 2010); Role of the Internet Industry in Addressing Hate on the Internet, Amsterdam (10 May, 2010); High-Level Conference on Tolerance and Non-Discrimination, Astana (29–30 June 2010).

65 OSCE, Permanent Council, *Decision No. 633. Promoting Tolerance and Media Freedom on the Internet* (11 November 2004).

66 *Ibid.*, para 4.

67 Školkay, A. 2015, *Media Policy for a New Media Environment: The Approaches of International Organizations and the EU towards the Regulation of New Online Media Services*, School of Communication and Media, n.o., p. 26.

## 5. SELF-REGULATORY INITIATIVES TO PREVENT AND COMBAT HATE SPEECH ONLINE

Self-regulatory policies have been established by various web services, such as YouTube, Google, Blogger, and Facebook, through acceptable use policies and community guidelines.<sup>68</sup> The effectiveness of self-regulatory policies has been criticized by many non-governmental organizations, state-level regulators, international organizations and the media.<sup>69</sup> According to Christopher Wolf, the Chair of Anti-Defamation League Internet Task Force, three web services – YouTube, MySpace and Facebook – often ignored reports and the content proliferates faster than conscientious users could report it.<sup>70</sup> The problem was that web services developed self-regulatory policies that were not specialized in hate speech and could not recognize illegal content. In order to overcome the lack of knowledge about hate speech, Facebook responded by establishing the Oversight Board with 20 members who with various experiences, including professors, journalists and heads of state.<sup>71</sup> The intention was to strengthen the external complaints' process and provide independent review of decisions related to hate speech online.<sup>72</sup> The Oversight Board mainly relies on international standards of human rights when passing decisions.<sup>73</sup> However, the Oversight Board Charter provides companies with a number of escape clauses, which enables Facebook to claim their response to have been “adapted to their technical and operational feasibility and consistent with a reasonable allocation of Facebook’s resources.”<sup>74</sup>

---

68 Facebook's Terms of Use and User Conduct Policy (<https://www.facebook.com/legal/terms>, 23 September 2022); Blogger Report Abuse Page, ([https://www.blogger.com/report?hl=en&visit\\_id=637995231999443134-3081178069&rd=1](https://www.blogger.com/report?hl=en&visit_id=637995231999443134-3081178069&rd=1), 12. 09. 2022).

69 Akdeniz, Y., 2009, *Racism on the Internet*, Council of Europe, p. 112.

70 Commission on Security and Cooperation in Europe, 2008, *Hate in the Information Age, Briefing of the Commission on Security and Cooperation in Europe*, (<https://www.govinfo.gov/content/pkg/CHRG-110jhr75645/html/CHRG-110jhr75645.htm>, 23. 09. 2022).

71 Facebook Oversight Board (<https://www.oversightboard.com>, 12. 09. 2022).

72 Klonick, K., 2020, The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression, *The Yale Law Journal*, Vol. 129, No. 8, pp. 2449, 2418–2499.

73 Vukčević, I., 2021, Facebook Oversight Board's Decision on the Indefinite Suspension of Donald Trump's Account, *Pravni zapisi*, Vol. XII, No. 1, pp. 308, 295–311.

74 *Ibid.*



Council of Europe Recommendation REC (2001) 8 encouraged self-regulation of cyber content.<sup>75</sup> In May 2003, the Committee of Ministers of the Council of Europe adopted the Declaration on Freedom of Communication on the Internet, encouraging self- and co-regulation.<sup>76</sup> Additionally, the European Action Plan on promoting safer use of the Internet recommends self-regulatory mechanisms along with awareness-raising measures intended for children, parents teachers, and other users, facilitating safe use of the Internet.<sup>77</sup>

On the other hand, there are initiatives that call for refraining from measures that lead to arbitrary decision-making by web service providers, without due process by an independent, impartial, authoritative oversight body (such as a court) that has the knowledge and skills to make such decisions. It is shown that such mechanisms often fail to comply with international human rights law.<sup>78</sup> The European Commission concluded that “the adoption of blocking measures necessarily implies a restriction of human rights, in particular the freedom of expression and therefore, it can only be imposed by law, subject to the principle of proportionality, with respect to the legitimate aims pursued and to their necessity in a democratic society, excluding any form of arbitrariness or discriminatory or racist treatment.”<sup>79</sup> It was also shown that filtering methods can be circumvented.<sup>80</sup> The Committee of Ministers Recommendation CM/Rec (2008) 6, from March 2008, stressed that “public authorities should not, through general blocking or filtering measures, deny access by the public to information and other communication on the Internet, regardless of frontiers.”<sup>81</sup> Furthermore, Article 15 of the EU e-Commerce Directive does not allow the imposing of a

75 Council of Europe Committee of Ministers, *Recommendation Rec (2001) 8 of the Committee of Ministers to member states on self-regulation concerning cyber content* (5 September 2001).

76 Council of Europe, Committee of Ministers, *Declaration on freedom of communication on the Internet, the Committee of Ministers* (28 May 2003), principle 2.

77 European Parliament and the Council Decision No. 854/2005/EC of 11 May 2005 establishing a Multiannual Community Programme on promoting safer use of the Internet and new online technologies.

78 Article 19, 2018, *Self-regulation and ‘hate speech’ on social media platforms*, ([https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-‘hate-speech’-on-social-media-platforms\\_March2018.pdf](https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-‘hate-speech’-on-social-media-platforms_March2018.pdf), 21. 04. 2023), p. 4.

79 Council Framework Decision Amending Framework Decision 2002/475/JHA of 6 November 2007 on combating terrorism: Impact Assessment, para. 4.2.

80 *Ibid.*, para. 5.2.

81 Council of Europe, Committee of Ministers, *Recommendation Rec (2008) 6 of the Committee of Ministers to member states on measures to promote the respect for freedom of expression and information with regard to Internet filters*, (26 March 2008).

general monitoring obligation on ISPs.<sup>82</sup> In most national systems, liability will be imposed on the ISP only if there is its “knowledge and control” over the published and stored information. Therefore, notice-based liability has been developed in several states. In accordance with this practice, the EU Directive on Electronic Commerce introduces a limited responsibility for removing content, as well as procedures (“notice and take down”) in cases of unlawful content, while it also calls for the introduction of a code of professional ethics.<sup>83</sup> The notice can be provided by individuals, self-regulatory hotline and, in some states, by law-enforcement agencies or the courts. Some studies emphasized that ISPs based in Europe have the tendency to remove or take down content without challenging the content. On the other hand, the United States provide protection to ISPs even in the cases where they are aware of the published and stored content.<sup>84</sup>

Filtering and rating systems have been developing since the 1990s, enabling users to make their own decision regarding access to the content on the Internet. However, according to the EU Kids Online 2020: Survey results from 19 countries, in most of the countries, a minority of children (on average 22% and less) report that parents use such technological controls. an additional problem is that Internet filters can potentially censor useful websites due to technical weakness.

The clash of these two different approaches to controlling hate speech online culminated in the adoption of the Law on the Improved Enforcement of Laws on Social Networks (NetzDG) in Germany in June 2017, which requires all platforms to remove illegal content within 24 hours or face punishment.<sup>85</sup> The NetzDG established compliance requirements for social network providers, regarding setting up complaint management systems and dealing with user complaints.<sup>86</sup> The weakness of this mechanism lies in the risk that networks, in case of a suspicion of hate speech, will rather delete the content than leave it online. Consequently, over-blocking occurs, as indicated by a study commissioned by the Ger-

---

82 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services.

83 Cuhna, A. V. M. de, Andrade, G. N. N. de, Lexinski, L., Feteira, T. L., 2013, *New Technologies and Human Rights: Challenges to Regulation*, Routledge, p. 64.

84 Communication Decency Act, 47 U.S.C. (1996) Section 230, para. e.3.

85 Network Enforcement Act (Netzdurchsetzungsgesetz, NetzDG), *Federal Law Gazette I*, p. 3352 ff., Article 1, Section 2.

86 Hemmert-Halswick, M., Lessons Learned from the First Years with the NetzDG, in: Bayer, J., Holznapel, B., Korpisaari, P., Woods, L. (eds.), 2021, *Perspectives of Platform Regulation*, Nomos, pp. 417, 415–432.

man Federal Ministry of Justice and Consumer Protection in 2020.<sup>87</sup> Human Rights Committee, expressed concerns that the regulation forces social media companies to decide what is hate speech without judicial oversight “thereby limiting access to redress in cases.”<sup>88</sup>

Similarly, France adopted a Law on Combating Hate on the Internet with the same provision.<sup>89</sup> After a delay, due to the European Commission’s request to review possible violations of the members, parliament adopted the law in May 2020. The law provoked controversy, so the Constitutional Court of France abolished certain provisions. It was concluded that the regulation was expected to lead to content analysis being handled by technological platforms without the involvement of a judge, within a very short time frame, with potentially excessive penalties being passed down.<sup>90</sup> The court assessed that there was a danger of indiscriminately removing flagged content, regardless of whether it was clear hate speech or not. Accordingly, the Constitutional Court assessed that the Provisions of the Law “therefore infringe upon the exercise of freedom of expression and communication in a way that is not necessary, suitable, and proportionate.”<sup>91</sup>

The most problematic issue with Germany’s Social Media Law Enforcement Act, as well as with a similar law in France, is defining the responsibility of private companies, rather than judges, to decide whether questionable content is indeed illegal. In other words, the state has “privatized” one of its main duties – law enforcement. In the case of appeals, deleting online content is effectively deleting evidence that would be required in court (unless the evidence is secured in a way that will be released to the court). In terms of the practical application of the law, once something is deleted, it cannot be used for forensic purposes.

The criticism of assigning “duties and responsibilities” to Internet portals regarding user-generated content was further strengthened by the judgment of the European Court of Human Rights in *Magyar Jeti ZRT v. Hungary*, dated 4 December 2018.<sup>92</sup> This case raised the question of

87 Liesching, M. et al., 2021, *Das NetzDG in der praktischen Anwendung: Eine Teilevaluation des Netzwerkdurchsetzungsgesetzes*, Carl Grossmann Verlag, p. 143.

88 Human Rights Committee, Concluding observations on the seventh periodic report of Germany, UN doc. CCPR/C/DEU/CO/7, (30 November 2021), para. 46.

89 Proposition de loi visant à lutter contre les contenus haineux sur internet, ([https://www.assemblee-nationale.fr/dyn/15/textes/l15t0419\\_texte-adoptee-seance.pdf](https://www.assemblee-nationale.fr/dyn/15/textes/l15t0419_texte-adoptee-seance.pdf), 23. 09. 2022), Art. 4.

90 Conseil constitutionnel, Décision n° 2020–801 DC du 18 juin 2020, para. 19.

91 *Ibid.*, para. 8.

92 ECtHR, *Magyar Jeti ZRT v. Hungary*, No. 11257/16, Judgment of 4 December 2018.

objective responsibility for placing a hyperlink leading to the disputed content. The Court decided that the posting of a hyperlink does not in itself constitute distribution of information.<sup>93</sup> The ECtHR emphasized the importance of hyperlinks for the smooth functioning of the Internet and rejected the automatic imposition of the burden of objective responsibility on journalists, which would lead to the suppression of media freedom.<sup>94</sup>

## 6. CONCLUSION

This paper first explored the issue of the lack of the universal definition of hate speech and concluded that it is not possible to reach an agreement, due to different treatments of freedom of expression and the understanding of what elements constitute hate speech. Bearing in mind that it is not possible to establish the universal definition, the third section explores the implications of the borderless nature of the Internet and the impact of national legislation targeting hate speech online. The finding is that the legislation of a single state cannot successfully handle the problem. Considering the limitation of national legal initiatives and the lack of universal definition, the fourth section explores cross-border cooperation initiatives targeting this phenomenon. The assessment of cross-border legal Internet regulations shows the lack of will on the part of some states to participate in such initiatives. The weakness of legal mechanisms against this phenomenon has led to the creation of a growing number of cross-border initiatives supporting a holistic approach to Internet-related issues, involving cooperation between governments, international organizations, law enforcement agencies, the Internet industry, and NGOs. These initiatives, which go beyond legal mechanisms, have shown potential in fighting this phenomenon. The paper also assessed the impact of self-regulatory mechanisms against hate speech and found that they can be useful tools against hate speech, but with certain limitations, due to the lack of adequate expertise by web service providers. The introduction of an obligation for platforms to remove illegal content themselves, as was the case with NetzDG, has led to over-blocking of content and the restriction of the freedom of expression. Therefore, imposing such an obligation on an internet provider, without the support of an expert state institution or a state-formed international agency dealing with hate speech, should be avoided.

---

93 *Ibid.*, para. 76.

94 *Ibid.*, para. 73.

The future development of regulations against hate speech should be explored in light of the borderless character of the Internet and the absence of a universal definition of hate speech. Given that liability for hate speech, as explained in the third section, includes users, content providers, service providers, telecommunications network providers and broadcasters, there is a need to combine various mechanisms, including legal and self-regulatory mechanisms, cross-border cooperation and education.

Hence, the creation of a new specialized international agency concerning hate speech online should be explored as a solution. Considering the universal nature of the online hate speech, the agency could be formed as part of the United Nations system. This agency should have a mandate to support the specialization of law enforcement and prosecutors dealing with hate crimes; strengthening and institutionalizing the dialogue and cooperation between law enforcement, the Internet industry and civil society, regarding issues related to online hate speech; and enhancing international law enforcement cooperation in this field. The establishment of such an agency could be accompanied by the adoption of an additional protocol to the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), but without the intention to provide a universal definition of hate speech, similarly to the solution that was found with the adoption of the Framework Convention on National Minorities.<sup>95</sup> This solution implies a dialogue in the search for the mechanism against hate speech online instead of imposing a solution that would lead to new conflicts and non-acceptance of the protocol.

## BIBLIOGRAPHY

1. Akdeniz, Y., 2009, *Racism on the Internet*, Strasbourg, Council of Europe Publishing.
2. Alkiviadou, N., 2016, Regulating Internet Hate: A Flying Pig?, 7 *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, 3, pp. 216–228.
3. Barron, J. A., 2019, Internet Access, Hate Speech and the First Amendment, *First Amendment Law Review*, Vol. 18, Symposium Issue, pp. 1–24.
4. Beširević, V., 2018, Catch Me if You Can: Reflections on Legal (un)accountability of Transnational Corporations for Human Rights Violations, *Pravni zapisi*, Vol. 9, No. 1, pp. 21–42.
5. Brown, A., 2017a, What Is Hate Speech? Part 1: The Myth of Hate, *Law and Philosophy*, Vol. 36, Issue 4, pp. 419–468.

---

95 Trier, T., Samasile, E., 2005, 'Towards Ratification' Conference on the Framework Convention for the Protection of National Minorities, European Centre for Minority Issues, p. 18.

6. Brown, A., 2017b, What Is so Special about Online (as Compared to Offline) Hate Speech, *Ethnicities*, Vol. 18, No. 3, pp. 297–326.
7. Chen, G., 2022, How Equalitarian Regulation of Online Hate Speech Turns Authoritarian: A Chinese Perspective, *Journal of Media Law*, Vol. 14, No. 2, pp. 159–179.
8. Cooke, L., 2007, Controlling the Net: European Approaches to Content and Access Regulation, *Journal of Information Science*, Vol. 33, No. 3, pp. 360–376.
9. Crelling, K., 2017, *The Confederate Battle Flag: Why Is It Perceived so Differently in the US and Europe/Germany*, University of Washington.
10. Cuhna, A. V. M. de, Andrade, G. N. N. de, Lexinski, L., Feteira, T. L., 2013, *New Technologies and Human Rights: Challenges to Regulation*, Routledge.
11. Gilbert, O. de, Perez, N., Pablos, A. G., Cuadros, M., 2018, Hate Speech Dataset from a White Supremacy Forum, *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, Association for Computational Linguistics.
12. Goldbert, A., 2015, Hate Speech and Identity Politics in Germany, 1848–1914, *Central European History*, Vol. 48, No. 4, pp. 480–497.
13. Greer, S., 2000, The Margin of Appreciation: Interpretation and Discretion Under the European Convention on Human Rights, *Human rights files*, No. 17, Council of Europe.
14. Greer, S., 2000, The Margin of Appreciation: Interpretation and Discretion Under the European Convention on Human Rights, *Human Rights Files*, No. 17, Council of Europe, pp. 5–58.
15. Heldt, A., 2019, Reading Between the Lines and the Numbers: An Analysis of the First NetzDG Reports, *Internet Policy Review*, Vol. 8, No. 2.
16. Hemmert-Halswick, M., Lessons Learned from the First Years with the NetzDG, in: Bayer, J., Holznapel, B., Korpisaari, P., Woods, L. (eds.), 2021, *Perspectives of Platform Regulation*, Nomos, pp. 415–432.
17. Holmberg, E. S., 2021, Lessons from Trump’s Suspension: How Twitter Should Clarify and Strengthen Its “Public Interest” Approach to Moderating Leaders’ Violence-Inspiring Speech, *Harvard Journal of Law & Technology*, Vol. 35, No. 1, pp. 310–334.
18. Keum, T. B., Miller, J. M., 2018, Racism on the Internet: Conceptualization and Recommendations for Research, *Psychology of Violence*, Vol. 8, No. 6, pp. 783–791.
19. Khan, A. R., 2013, Why Do Europeans Ban Hate Speech? A Debate Between Karl Loewenstein and Robert Post, *Hofstra Law Review*, Vol. 41, No. 3, pp. 545–585.
20. Klonick, K., 2020, The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression, *The Yale Law Journal*, Vol. 129, No. 8, pp. 2418–2499.
21. Leerssen, P., Tworek H., 2019, *An Analysis of Germany’s NetzDG Law*, Transatlantic Working Group.
22. Liesching, M., et al., 2021, *Das NetzDG in der praktischen Anwendung: Eine Teilvaluation des Netzwerkdurchsetzungsgesetzes*, Carl Grossmann Verlag.
23. Murphy, S. D. 2004, *United States Practice in International Law: Volume 2, 2002–2004*, Cambridge, Cambridge University Press.
24. Nockleby, T., Hate Speech, in: Levy, L., Karst, K., (eds.), 2000, *Encyclopedia of the American Constitution, Encyclopaedia of the American Constitution*, Macmillan Reference USA, pp. 1277–1279.

25. Školkey, A. 2015, *Media Policy for a New Media Environment: The Approaches of International Organizations and the EU towards the Regulation of New Online Media Services*, School of Communication and Media, n.o.
26. Temperman, J., 2015, *Religious Hatred and International Law*, Cambridge, Cambridge University Press.
27. Trier, T., Samasile, E., 2005, 'Towards Ratification' Conference on the Framework Convention for the Protection of National Minorities, European Centre for Minority Issues.
28. Ubanga, C., 2016, *Hate Speech in Cyberspace: Why Education Is Better than Regulation*, Social Science Research Network.
29. Vukčević, I., 2021, Facebook Oversight Board's Decision on the Indefinite Suspension of Donald Trump's Account, *Pravni zapisi*, Vol. XII, No. 1, pp. 295–311.
30. Walker, J., 2010, *Section 13 of the Canadian Human Rights Act, Anti-Hate Laws and Freedom of Expression*, Ottawa, Ontario, Library of Parliament.
31. Walker, C., Akdeniz, Y., 1999. The Governance of the Internet in Europe with Special Reference to Illegal and Harmful Content, *Criminal Law Review*, December Special Edition, pp. 15–19.
32. Watson, K. M., 2020, The United States' Hollow Commitment to Eradicating Global Racial Discrimination, *Human Rights*.
33. Wilson, A. R., Land, M., 2021, Hate Speech on Social Media: Content Moderation in Context, *Connecticut Law Review*, Vol. 52, No. 3, pp. 1029–1242.
34. Windisch, S. *et al.*, 2022, Online Interventions for Reducing Hate Speech and Cyberhate: A Systematic Review, *Campbell Systematic Reviews*, Vol. 17, No. e1133, pp. 1–17.
35. Yar, M., 2018, A Failure to Regulate? The Demands and Dilemmas of Tackling Illegal Content and Behavior on Social Media, *International Journal of Cybersecurity Intelligence and Cybercrime*, Vol. 1, No. 1, pp. 5–20.

## LEGISLATIVE SOURCES

1. Canadian Criminal Code, R.S.C., 1985, c. C-46.
2. Constitution of the Republic of Serbia, *Official Gazette of the RS*, No. 98/06.
3. Law on the Prohibition of Discrimination in the Republic of Serbia, *Official Gazette of the RS*, No. 22/09.
4. Criminal Code, *Official Gazette of the RS*, Nos. 85/05, 88/05 – corr., 107/05 – corr., 72/09, 111/09, 121/12, 104/13, 108/14, 94/16 and 35/19.
5. U.S. Constitution, Bill of Rights, First Amendment, Religion, Speech, Press, Assembly, Petition (1791).
6. New York State Senate Bill S8298B.
7. Communication Decency Act, 47 U.S.C. (1996).
8. German Criminal Code, *Federal Law Gazette I*, p. 3322, as last amended by Article 2 of the Act of 19 June 2019, *Federal Law Gazette I*.
9. Network Enforcement Act (Netzdurchsetzungsgesetz, NetzDG), *Federal Law Gazette I*, p. 3352 ff.
10. Protection of Young Persons Act, (Jugendschutzgesetz), *Federal Law Gazette I*.

11. Broadcasting Interstate Agreement, (Rundfunkstaatsvertrag), dated 31 August 1991 (incorporating the third amendment adopted between 26 August and 11 September 1996).
12. Gesetz zur Änderung des Netzwerkdurchsetzungsgesetzes, Art. 2, d) 9, (<http://perma.cc/9W8E-GSWM>, 15 September 2022).
13. Code pénal Version consolidé au 1 janvier 2020.
14. Proposition de loi visant à lutter contre les contenus haineux sur internet, Texte adopté n° 419.
15. Loi du 29 juillet 1881 sur la liberté de la presse, Version en vigueur au 03 octobre 2022.
16. UN General Assembly, *International Covenant on Civil and Political Rights*, UN doc. Resolution 2200A (XXI) (16 December 1966).
17. UN General Assembly, *International Convention on the Elimination of All Forms of Racial Discrimination (ICERD)*, UN doc. Resolution 2106 (21 December 1965).
18. Human Rights Committee, *Resolution 20/8 on the Internet and Human Rights*, UN doc. A/HRC/RES/20/8 (16 July 2012).
19. Human Rights Committee, *General Comment No. 34*, UN doc. CCPR/C/GC/34 (12 September 2011).
20. Human Rights Committee, *Concluding observations on the seventh periodic report of Germany*, UN doc. CCPR/C/DEU/CO/7 (30 November 2021).
21. African Union, African Charter on Human and Peoples' Rights (27 June 1981).
22. Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, ETS 5 (4 November 1950).
23. Council of Europe, Committee of Ministers, Recommendation No. R (97) 20 (30 October 1997).
24. Council of Europe, Convention on Cybercrime, ETS No. 185 (8 November 2001).
25. Council of Europe, Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, ETS 189 (28 January 2003).
26. Council of Europe, Explanatory report to the Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, ETS 189 (28 January 2003).
27. Council of Europe, Committee of Ministers, Recommendation Rec (2001) 8 of the Committee of Ministers to member states on self-regulation concerning cyber content (5 September 2001).
28. Council of Europe, Committee of Ministers, Declaration on freedom of communication on the Internet, the Committee of Ministers (28 May 2003).
29. Council of Europe, Committee of Ministers, Recommendation Rec (2008) 6 of the Committee of Ministers to member states on measures to promote the respect for freedom of expression and information with regard to Internet filters (26 March 2008).
30. European Parliament and the Council Decision No. 854/2005/EC of 11 May 2005, establishing a multiannual Community Programme on promoting safer use of the Internet and new online technologies.



31. Council Framework Decision Amending Framework Decision 2002/475/JHA of 6 November 2007 on combating terrorism: Impact Assessment.
32. Council Framework Decision 2008/913/JHA of 28 November 2008, on combating certain forms and expressions of racism and xenophobia by means of criminal law.
33. Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services.
34. EU Directive 2006/24/E on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/E., *Official Journal of the European Union* L 105/54.
35. European Commission Communication COM (2021) 777 final of 9 December 2021, A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime.

## CASE LAW

### European Court of Human Rights

1. ECtHR, *Handyside v. the United Kingdom*, No. 5493/72, Judgment of 7 December 1976.
2. ECtHR, *Magyar Jeti ZRT v. Hungary*, No. 11257/16, Judgment of 4 December 2018.

### Canada

1. Canadian Human Rights Tribunal, *Citron et al. v. Zündel*, T.D. 1/02 2002/01/18, Decision of 18 January 2002.

### U. S.

1. U.S. Supreme Court, *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942), Decision of 9 March 1942, at 572.
2. U.S. Supreme Court, *Roth v. United States*, No. 582, Decision of 24 June 1957.
3. U.S. Supreme Court, *New York v. Ferber* No. 81–55, Decision of 2 July 1982.
4. U.S. Supreme Court, *Beauharnais v. Illinois*, No. 118, Decision of 28 April 1952.
5. U.S. Supreme Court, *Watts v. United States*, No. 21528, Decision of 25 September 1968.
6. U.S. Supreme Court, *Virginia v. Black*, No. 01–1107, Decision of 7 April 2003.
7. District Court for the Northern District of California, *Yahoo!, Inc. v. La Ligue Contre Le Racisme* 169 F. Supp. 2d 1181 (N.D. Cal. 2001), Decision of 7 November 2001, 1188–90.

### France

1. Conseil constitutionnel, Décision n° 2020–801 DC du 18 juin 2020.

2. Tribunal de Grande Instance (High Court), *UEJF and Licra v. Yahoo! Inc and Yahoo France*, RG 05308 *UEJF and Licra v. Yahoo! Inc and Yahoo France*, RG 05308, Decision of 22 May 2000.

## REPORTS AND PUBLICATIONS

1. UN General Assembly, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN doc. A/74/486 (9 October 2019).
2. Secretary-General, *United Nations Strategy and Plan of Action on Hate Speech: detailed guidance on implementation for United Nations field presences* ([http://digitallibrary.un.org/record/3889286?ln=zh\\_CN](http://digitallibrary.un.org/record/3889286?ln=zh_CN), 22 August 2022).
3. Secretary-General, *The fight against racism, racial discrimination, xenophobia and related intolerance and the comprehensive implementation of and follow-up to the Durban Declaration and Programme of Action*, UN doc. A/58/313 (22 August 2003).
4. World Summit on the Information Society, Geneva 2003 – Tunis 2005, *Declaration of Principles: building the information society: a global challenge in the new millennium*, UN doc. WSIS-03/GENEVA/DOC/4 (12 December 2003).
5. *Tunis Agenda for the Information Society*, UN doc. WSIS-05/TUNIS/DOC/6(Rev.1)-E (18 November 2005).
6. OSCE-ODIHR, 2010, *Report of OSCE-ODIHR activities on hate on the internet* (<https://www.osce.org/files/f/documents/8/7/73461.pdf>, 24 September 2022).
7. Commission on Security and Cooperation in Europe, 2008, *Hate in the Information Age, Briefing of the Commission on Security and Cooperation in Europe*, (<https://www.govinfo.gov/content/pkg/CHRG-110jhrgr75645/html/CHRG-110jhrgr75645.htm>, 23 September 2022).
8. OSCE, Permanent Council, *Decision No. 633. Promoting Tolerance and Media Freedom on the Internet* (11 November 2004).

## INTERNET SOURCES

1. Cuevas, S., 2009, Neo-Nazi Rallies Provoke “Anger, Fear”, *NPR* (<https://www.npr.org/2009/11/07/120129726/neo-nazi-rallies-provoke-anger-fear>, 08. 12. 2022).
2. Statista Research Department, 2022, U.S. teens encountering hate speech on social media 2018, by type (<https://www.statista.com/statistics/945392/teenagers-who-encounter-hate-speech-online-social-media-usa/>, 13. 09. 2022).
3. Laub, Z., 2019, *Hate Speech on Social Media: Global Comparisons*, Council on Foreign Relations (<https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons>, 13. 09. 2022).
4. Yosie, M., 2012, Section 13 of the Canadian Human Rights Act Repealed!?, *Slaw* (<https://slaw.ca/2012/06/14/section-13-of-the-canadian-human-rights-act-repealed/>, 16. 09. 2022).

5. Petaković, M., 2015, Presuda zbog govora mržnje na internetu, *Blic* (<https://www.blic.rs/vesti/hronika/presuda-zbog-govora-mrznje-na-internetu/lg3etnr>, 16. 09. 2022).
6. United Nations, *United Nations Establishes Working Group on Internet Governance*, PI/1620 (<https://press.un.org/en/2004/pi1620.doc.htm>, 18. 09. 2022).
7. European Commission, 2022, *EU Code of Conduct against online hate speech: latest evaluation shows slowdown in progress* ([https://ec.europa.eu/commission/presscorner/detail/da/ip\\_22\\_7109](https://ec.europa.eu/commission/presscorner/detail/da/ip_22_7109), 01. 12. 2022)
8. Facebook's Terms of Use and User Conduct Policy (<https://www.facebook.com/legal/terms>, 23. 09. 2022).
9. Blogger Report Abuse Page ([https://www.blogger.com/report?hl=en&visit\\_id=637995231999443134-3081178069&rd=1](https://www.blogger.com/report?hl=en&visit_id=637995231999443134-3081178069&rd=1), 12 September 2022).
10. Jugendschutz.net (<https://www.jugendschutz.net/en/>, 03. 10. 2022).
11. International Cyber Security Protection Alliance (<https://www.cybersecurityintelligence.com>, 11. 05. 2022).

## SANKCIONISANJE GOVORA MRŽNJE NA INTERNETU: U POTRAZI ZA NAJBOLJIM PRISTUPOM

Petar Antić

### APSTRAKT

Bezgranična priroda interneta, različiti nacionalni pristupi razumevanju šta je govor mržnje, kao i opasnost od ograničenja slobode govora otežavaju razvoj odgovarajućih mehanizama protiv ove pojave. Ograničenja međunarodnog prava u davanju univerzalne definicije govora mržnje zbog različitih nacionalnih pristupa slobodi izražavanja osujetila su pokušaje da se napravi efikasan međunarodni ugovor koji bi se bavio ovim pitanjem. Nametanje obaveza internet portalima da uspostave samoregulatorne mehanizme za brisanje sadržaja govora mržnje izazvalo je zabrinutost zbog neprimerene cenzure i ograničenja slobode izražavanja. Ovaj rad se fokusira na izazove sa kojima se susrećemo u borbi protiv govora mržnje na internetu i na moguće mehanizme za suzbijanje ove pojave.

**Ključne reči:** govor mržnje, sloboda izražavanja, internet, onlajn, mediji, distribucija.

Article History:

Received: 27 February 2023

Accepted: 7 June 2023