



Report of the Project “Hate Speech on the Internet. A comparative analysis of European and American legal regulations”

Hate speech is among one of the most widespread problems on the Internet¹ with which several entities operating and regulating its space are struggling. Combating the proliferation of online hate speech, which continues to record upward trends, is one of the challenges facing Internet regulators, which are both Internet platforms and state authorities. The increased use of online media as a major means of communication and information transport is particularly relevant in this regard. The project aimed to analyze individual cases of hate speech in the Internet space appearing in case law and under US law and European regulations.

These are significantly different systems, applying a different range of restrictions, definitions, or understandings of fundamental rights and freedoms, particularly relevant to this analysis. This is due, of course, to the different historical experiences of the two continents, which directly affect the formed legal systems involving the definition and regulation of fundamental rights and freedoms. The originality and differences in these

¹ Benedek W., Kettemann M. C Freedom of expression and the Internet, Council of Europe Publishing, 2013, p. 83.

systems undoubtedly derive the greatest advantage of the present analysis, namely its comprehensiveness and ability to provide innovative perspectives.

Thus, at the outset, it should be emphasized that from the dissimilarity of these systems and the restrictions they impose on freedom of speech, there is a different system of evaluation and analysis of European and American jurisprudence. First, European jurisprudence explicitly refers to hate speech as a non-protected phenomenon, while in the US system, hate speech is permitted and protected under the regulations of the First Amendment to the US Constitution. Secondly, in American jurisprudence, we do not find a reference to hate speech as such, while in European jurisprudence we do. The case law analyzed by the authors of the research project is primarily the case law of the European Court of Human Rights, a conventional judicial body upholding human rights and freedoms, and the case law of state courts (mainly appellate courts) of the United States.

One of the important elements of the research conducted, which served as a starting point, was to establish the definition of hate speech itself. The most common one that appears within the framework of the discussion of hate speech is the definition that was adopted under the Council of Europe General Assembly Recommendation No. R 97 (20). According to it, hate speech is “any form of expression that disseminates, incites, promotes or justifies racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerance expressed in the form of aggressive nationalism or ethnocentrism, discrimination or hostility towards minorities, migrants, or persons from migrant communities”². Hate speech on the Internet thus refers to the use of discriminatory, offensive, or harmful language or expressions directed at individuals or groups based on attributes such as race, ethnicity, religion, gender, sexual orientation, or other protected characteristics. In addition, it can include speech, gestures, images, or any form of expression that promotes violence, discrimination, or hostility against individuals or groups based on their protected

² Council of Europe General Assembly Recommendation No. R 97 (20).

attributes³. The dissemination of false data often accompanies it, the promotion of misinformation, and the spread of hate ideology. In addition, it leads to the exacerbation of social conflicts, and verbal or physical violence, as well as restricts freedom of expression and can contribute to social exclusion.

US

In the United States, under a broad interpretation of freedom of speech, hate speech is usually protected except when it directly incites violence or poses an imminent threat. In other words, freedom of speech is usually protected even when it expresses controversial or unpopular views, but there is a limit that is crossed when it directly incites violence or creates a real threat to others. The Supreme Court has systematically held that the Constitution's First Amendment provides protection even for the content of speech that may be considered offensive or hateful. However, it is the context and whether the speech directly incites violence or poses a direct threat to other individuals that is critical. Unlike countries in Europe, the United States does not have specific laws that criminalize hate speech as a separate offense under federal or state laws, due to the broad scope of free speech. However, some forms of hate speech, such as harassment or threats, can be prosecuted under existing laws on incitement, defamation, or harassment. For this reason, there is a lack of judicial decisions that directly refer to hate speech, but only concepts related to it in the US system that take the form of hate speech.

Thus, the analysis is made of judgments that treat hate speech in the European sense, and within the framework of U.S. regulations are understood as cyber harassment (digital intimidation), cyberbullying (digital harassment), cyberstalking (digital stalking), or are related to social media platforms, discrimination or restriction of access within the Internet.

³ Doyle A., Free Speech And Why it Matters, Constable & Robinson, 2021, p. 88-89.

The issue of hate speech remains the subject of ongoing debate and controversy in American society. Some advocate stricter laws to protect vulnerable groups, while others stress the importance of maintaining broad protections for free speech, even if it contains offensive or hateful expressions. Interestingly, however, the issue of hate speech is one of the oldest public issues, dating back to the Maryland Toleration Act of 1649⁴. The Toleration Act was one of the first laws touching on hate speech in the context of religious tolerance. However, the issue has evolved over the years. Currently, in the United States, hate speech cannot be directly regulated by the government due to the constitutional right to free speech, which is protected by the First Amendment. The U.S. Supreme Court has repeatedly affirmed that most content considered hate speech in other Western countries is legally protected speech under the First Amendment⁵. In a unanimous decision in *Matal v. Tam* (2017), the Supreme Court affirmed that there is no exception for “hate speech” from free speech rights, and the US government cannot discriminate against speech based on the views of the speaker⁶. The court said the Lanham Act's disparagement clause violates the First Amendment's free speech clause⁷. The judges stressed that while speech that demeans based on race, ethnicity, gender, religion, etc., is hateful, one of the most important aspects of freedom of speech is to protect the expression of thoughts that may be unpopular or unwanted. Guaranteeing the prohibition of hate speech, while preserving freedom of political expression, is extremely challenging due to the subjectivity and variation in the potential interpretation of such speech. In practice, defining the line between acceptable political expression and hate speech is often ambiguous

⁴ M.W. McConnell, *America's First Hate Speech Regulation*, „Constitutional Commentary” 1992, nr 9(1), p. 17–23.

⁵ M. Israel, (1999). *Hate Speech and the First Amendment*, „Journal of Contemporary Criminal Justice” 1999, ne 15(1), p. 101.

⁶ Part of Judge Alito's Opinion: Speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express “thoughts we hate.” <https://supreme.justia.com/cases/federal/us/582/15-1293/#tab-opinion-3749204> (accessed 10.07.2023).

⁷ *Matal v. Tam*, www.oyez.org/cases/2016/15-1293 (accessed 10.07.2023).

and controversial. What one person may view as political criticism, another may see as an attack on his or her identity or social group, leading to controversy and dispute. Different individuals may interpret the definition of hate speech in different ways, which poses a challenge when trying to establish uniform regulations without infringing on freedom of speech. Developing a precise definition of hate speech, which would additionally be universal and cross-border in nature, seems practically impossible, as there is a risk that the scope of such a definition could also include legitimate speech, which could lead to its silencing.

An analysis of the jurisprudence of U.S. courts, in the predominant part of state courts (appellate divisions), shows that the dominant principle is the protection of hate speech as speech protected by the First Amendment. Consequently, laws that violate this protection are declared illegal. In addition, the rulings show that the courts seek a narrowing interpretation of the laws at issue in certain cases in such a way that they do not violate the First Amendment. However, it is impossible to draw a single general conclusion from the analyzed case law, due to the variation about the acts and regulations that are subject to evaluation within it. One can only try to identify a kind of catalog related directly or indirectly to hate speech. This catalog should include key statements for jurisprudence, such as a government official cannot prohibit hate speech on his website (Brian Davison v. Phyllis Randal⁸), the overly broad scope of the Cyberbullying Act criminalizing hate speech violates the First Amendment, (The People v. Marquan M.⁹), it is permissible to use hate speech as evidence in a case (Urbanski v. State¹⁰). The database of summaries of judgments created as part of the implementation of this research project is intended to show the fullest possible picture of the case law in the area under discussion.

⁸ Case of Brian Davison v. Phyllis Randall, No. 17-2002 (4th Cir. 2019): <https://law.justia.com/cases/federal/appellate-courts/ca4/17-2002/17-2002-2019-01-09.html> (accessed 10.07.2023).

⁹ Case of People v. Marquan M. 2014 NY Slip Op 0488: <https://law.justia.com/cases/new-york/court-of-appeals/2014/139.html> (accessed 10.07.2023).

¹⁰ Case of Sean Urbanski v. State of Maryland, No. 1318: <https://mdcourts.gov/data/opinions/cosa/2022/1318s20.pdf> (accessed 10.07.2023).

EUROPE

In Europe, laws banning the spread of hate speech were mainly aimed at combating political racism, anti-Semitism, and the rise of fascist parties. Initially, these bans stemmed from a reaction to the traumatic experience of World War II and the stigma it left behind in terms of spreading hate and promoting totalitarian regimes. Efforts to counter racism and xenophobia derive from the deep-rooted principle of respect for human dignity and the fundamental freedoms and rights on which the legal structure of post-war Europe is based¹¹.

The main regulation on which the study is based is the European Convention on Human Rights, specifically Article 10 and Article 17. According to Article 10(1) of the Convention, everyone has the right to freedom of expression, including freedom to hold opinions and to receive, and impart information and ideas without interference from public authorities and regardless of national borders¹². This article protects individuals from state interference with their freedom of expression, ensuring respect for the rule of law. Restrictions on this freedom must meet a three-element test that assesses their compatibility with democratic values and other rights and principles. The three-element test is designed to examine the proportionality, necessity, and justification for restrictions on freedom of expression on the Internet. It takes a broader and more specific shape in the Internet space.

Freedom of expression also applies to online forms of expression, and the Internet - like other mass media - influences the way freedom of expression is understood. However, the interpretation and regulation of online aspects of freedom of expression are specific and require in-depth analysis. According to the aforementioned three-element test, restrictions on freedom of expression must be proportionate, necessary, and driven by legitimate objectives, such as state security, territorial integrity, public safety, protection of the health,

¹¹ A. Śledzińska-Simon, *Mowa nienawiści – wprowadzenie*, [w:] *Mowa nienawiści a wolność słowa. Aspekty prawne i społeczne*, red. R. Wieruszewski et al., Warsaw 2010,

¹² Article 10 of the Convention for the Protection of Human Rights and Fundamental Freedoms drawn up in Rome on 4 November 1950, subsequently amended by Protocols Nos. 3, 5 and 8 and supplemented by Protocol No. 2 (Journal of Laws of 1993, No. 61, item 284, as amended).

morals, and/or reputation of others, protection of confidential information and preservation of the seriousness and impartiality of the judiciary¹³. The European Court of Human Rights stresses that states have a certain margin of discretion in establishing restrictions, but must take into account the principles of non-discrimination. Restrictions must be based on existing legal provisions and take into account the legal context, the factual circumstances, the nature of the speech, its purpose, and the sender. In addition, restrictions should not have a chilling effect on freedom of expression and should balance the interests of individuals and the public interest. As stated in Article 17 of the Convention, nothing in the Convention may be interpreted as granting the right to take action to nullify the rights and freedoms guaranteed by the Convention. Article 17 serves to prevent the abuse of convention rights and applies in cases where certain actions go beyond the rights and harm the values of a democratic society¹⁴.

The European Court of Human Rights also examines issues related to hate speech in its online manifestations. The Court's jurisprudence includes online hate speech as one of the types of hate speech constituting its key manifestations. The Court recognizes that under Article 17 of the Convention, the protection of fundamental rights cannot be granted when this right is abused to violate other rights and freedoms protected by the Convention. The Court's jurisprudence analyzed in this Draft shows that Article 17 applies only in exceptional situations to prevent the use of a Convention right in a manner contrary to the values of the Convention¹⁵. About Article 10 of the Convention, Article 17 should only be applied when speech is intended to deviate from the purpose of Article 10 - by using the right to freedom of expression in a manner contrary to the values of the Convention¹⁶. Under these assumptions,

¹³Article 10(2), *Ibidem*.

¹⁴ Article 17 of the Convention for the Protection of Human Rights and Fundamental Freedoms drawn up in Rome on 4 November 1950, subsequently amended by Protocols Nos. 3, 5 and 8 and supplemented by Protocol No. 2 (Journal of Laws of 1993, No. 61, item 284, as amended).

¹⁵ Nowicki M. A., art. 10, *Komentarz do Konwencji o ochronie praw człowieka i podstawowych wolności [w:] Wokół Konwencji Europejskiej. Komentarz do Europejskiej Konwencji Praw Człowieka*, ed. VIII, Warsaw 2021,

¹⁶ K. Warecka, *Mowa nienawiści to nie wolność wyrażania opinii. Omówienie orzecznictwa ETPC na podstawie art. 10 Konwencji o prawach człowieka*, LEX/el. 2019.

speech that incites hatred or violence or is intended to violate rights and freedoms protected by the Convention does not fall under the protection of freedom of expression. These expressions may include the denial of rights such as the right to life, freedom of religion, prohibition of discrimination (including sexual orientation), or democratic values of the rule of law. It should be emphasized that the Court considers individual cases, analyzing the context and purpose of the speech and its impact on the rights and freedoms of others. There is a difference between political criticism and hate speech, which the Court takes into account when assessing compliance with the Convention.

It is clear from the case law reviewed that the Court treats the invocation of interference with freedom of expression (Article 10 of the Convention) as a clear abuse of Article 17 of the Convention. It also recognizes extremely often that there is indeed a violation of the freedom of expression guaranteed by the Convention, but the interference is justified and therefore legal (Sanchez v. France¹⁷, Smajić v. Bosnia and Herzegovina¹⁸). In addition, it should also be noted that the Court has determined the responsibility of websites for third-party comments (Delfi AS v. Estonia¹⁹) and agreed on the possibility of holding accountable the person under whose post-hateful comments appear (Sanchez v. France).

COMPARISON AND CONCLUSIONS

The project's analysis of two different systems of regulating and protecting hate speech through the protection of freedom of expression can contribute to improving the applied solutions both on national, European, and transnational grounds. Striking a balance between protecting against hate speech and preserving freedom of expression on the Internet is a challenge for European jurisprudence as well as European national and EU legislators. There is a real danger that an overly broad interpretation of hate speech could lead to excessive

¹⁷ Judgment of the ECtHR (Grand Chamber) of 15.05.2023 in Sanchez v. France, (Application no. 45581/15)

¹⁸ Judgment of the ECtHR of 16.01.2018 in Smajić v. Bosnia and Herzegovina, (Application no. 48657/16)

¹⁹ Judgment of the ECtHR (Chamber, Section I) of 10.10.2013 in DELFI AS v. Estonia, (Application no. 64569/09)

restrictions on freedom of expression online. The flexibility of the applicable liability and the openness to various categories of restrictions may result in an arbitrary approach to classifying certain types of speech as hate speech. As a result, there is a need to balance carefully between ensuring protection against discrimination and verbal violence and preserving freedom of expression and pluralism of opinion in the online space.

The application of the guidelines introduced in the case law, and restrictions on publishing and sharing content on online platforms can blatantly undermine Internet users' freedom of expression. Moreover, it is particularly controversial because of the importance of freedom of communication and freedom of expression online. For this reason, the European system in this regard should draw in some ways on US solutions to narrow the scope of restrictions only in situations of obvious violations that are deemed necessary under the existing legal system and culture. In doing so, it should stick to the framework set by the three-element test with particular attention to its online application. On the other hand, the U.S. approach seems almost shocking compared to the European one.

The primary concern is the extremely broad scope of First Amendment protection and the dangers it entails. Yes, free speech should be central to the legal system, but certain limitations are necessary to respect its ideas and role. While hate speech is generally protected, there are some limited exceptions, such as incitement to violence, true threats, obscenity, child pornography, defamation, and speech that incites/ incites or causes an immediate and present danger. In the case of the U.S. system, the protection of free speech seems overly broad and could potentially have a detrimental effect on the recipients or addressees of such content. The main instrument for combating hate speech is therefore social and educational campaigns. Thus, it turns out that through a kind of combination of the two systems, one can potentially achieve the most harmonious solution, under which adequate protection and scope of freedom of expression are established while combating hate speech.

Given the magnitude of the phenomenon, which is judged generally due to the global nature of the Internet, it seems that both European and US solutions are failing in their effectiveness. This is primarily about their deterrent and potentially chilling nature²⁰. While the fight against traditional manifestations of hate speech seems to be having an effect, in its online manifestations the measures taken by several involved actors are insufficient and fail. Therefore, there is a need for concerted action, drawing on shared experiences and solutions to apply such policies, regulations, and education that will have real effects and reduce the scale of the phenomenon.

ONLINE PLATFORMS

In line with the original objectives of the research project, an analysis of the solutions adopted by social media themselves was also conducted. It was limited to social networks: Facebook, Twitter²¹ and YouTube. Each of these platforms has regulations and community standards that define the scope of restricted content. Through these regulations, the sites influence freedom of expression on the Internet with a particular focus on the spaces of their online platforms. As online market giants, they indirectly influence the shape of the entire online space through the creation of their policies and regulations. Not only do they create their definitions of hate speech, distinguishing the characteristics on which such speech is based, but they also determine the penalties that threaten such publications. Due to the cross-border nature of freedom of online expression, it was necessary to take into account how the major issue is regulated by the large corporations that dominate the Internet market. After all, it is social media that is the main venue for abuses or violations of freedom of expression and the rights of others. Consequently, they are the main creators of the online environment.

²⁰ We are referring to the internet's chilling effect on freedom of expression as a phenomenon whereby users choose not to express their opinions or participate in online discussions due to fear of consequences such as attacks, hejt or repression.

²¹ From 24 July 2023, Twitter is the X platform.

Facebook, which is part of the Meta Platforms group, has established community standards that set guidelines for publishing content on the platform. These guidelines include rules of conduct, regulations on prohibited content, and guidelines for behavior toward other users, which aim to ensure a safe and positive online environment for all who use it. Facebook prohibits hate speech and imposes consequences for violations of these standards. As part of its community guidelines, the company explicitly states that content that promotes hate, discrimination, or violence based on personal characteristics, such as racism, sexism, homophobia, or religious intolerance, is not allowed on the platform. Violation of these rules can result in removal of content, restriction of access to platform features, and even suspension or deletion of a user's account. Decisions made by Facebook are final and arbitrary, based on their standards and not on a specific legal system²².

Twitter also has rules and policies governing published content. It bans hate speech, defining a narrow range of what is unacceptable. Twitter takes steps to limit the publication of hateful content and content that violates community standards by monitoring and removing it. Violations of these rules can result in reduced account visibility, suspension, or removal of content.

The third platform that is the subject of analysis within this report is YouTube. YouTube, as a video publishing platform, also has regulations on hate speech and several other regulations and policies on publishing potentially harmful content. Compared to other platforms, YouTube has a broad catalog of factors affected by the ban on promoting violence and hate. These include: "age, caste, disability, ethnicity, gender identity and expression, nationality, race, immigrant status, religion, biological/cultural sex, sexual orientation, being

²² The principles underpinning Facebook and other Meta products are set out, among others, in Facebook's Community Standards: <https://transparency.fb.com/pl-pl/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards> (accessed 10.07.2023).

a victim of a violent act or related person, veteran status”²³. Violations of these rules may result in the removal of the video or channel of the user in question.

The regulations used by the above platforms are designed to ensure a safe online environment. Private platforms such as social media can set their own rules on hate speech and have the right to remove or restrict content they deem offensive or harmful. They thus influence the determination of the scope of online freedom of expression as one of the main regulators of the online environment. However, online platforms face challenges in effectively moderating hate speech due to the vast amount of content, the global nature of the Internet, and the difficulty in distinguishing between legitimate free speech and harmful speech.

In conclusion, hate speech is a significant and increasingly emerging phenomenon that is not protected by freedom of expression under the European Convention on Human Rights, but is protected by the First Amendment to the US Constitution. For this reason, there are almost extremely different approaches to the issue at hand by the European and US systems. The regulation of hate speech is complicated because it raises the question of the balance between the protection of freedom of speech and the protection of the rights and dignity of individuals. US and European regulations and jurisprudence, in addition to online platforms, are making efforts to define and properly regulate hate speech, which differ significantly from each other. As a result, the line between legitimate online expression and hate speech can be subjective and unclear.

²³ Misrepresentation rules:

https://support.google.com/youtube/answer/10834785?hl=pl&ref_topic=10833358&sjid=612835591289554021-EU (accessed 10.06.2023).