

COUNTERING HATE SPEECH IN THE MEDIA: CHALLENGES AND SOLUTIONS

Branislav Sančanin

The College of Applied Studies for Management and Business Communication, Belgrade, Serbia

sancanin.branislav@gmail.com

Aleksandra Penjišević

Union – Nikola Tesla University, Belgrade, Serbia, aleksandra.penjisevic@gmail.com

Dragana Stefanović

Faculty of Economics and Engineering Management, Novi Sad, Serbia, draganavstefanovic@gmail.com

Abstract: Daily, explicit or implicit exposure to violence in the media fosters real-life violence, shaping attitudes, values, and behavior over time. This paper aims to highlight the existence and key drivers of hate speech, identify the most frequently targeted groups, and emphasize the urgent need for decisive action against all forms of discriminatory narratives and hate crimes before they become deeply rooted in society. The lack of preventive measures against hate speech and hate crimes creates fertile ground for undermining peace efforts and fueling new tensions and conflicts. The digitalization of media and the widespread use of social networks, often marked by diminished credibility and unclear legislative and ethical guidelines, have accelerated the spread of hate speech, fostering an environment of fear, division, and exclusion. In this context, efforts to create a safer digital space should be understood as a fundamental necessity rather than a privilege. Achieving tangible progress requires two key actions: the coordinated enhancement of media and user capacities to recognize hate speech and the legal and ethical regulation of the media space to prevent the dissemination of harmful content.

Keywords: hate speech, pyramid of hate, discriminatory narrative, media.

1. INTRODUCTION

Although there is no universal definition, hate speech—fundamentally opposed to freedom of expression, is a powerful driver of discrimination against vulnerable groups in society, fostering hatred, hostility, and aggression. The *Recommendation of the Committee of Ministers of the Council of Europe No. R97(20), Principle 6* states that “national legislation and practice in the field of hate speech should take into account the role of the media in transmitting information and ideas that expose, analyse and clarify the nature of specific cases of hate speech and the phenomenon as a whole, as well as the right of the public to receive such information and ideas” (Council of Europe, 1997). In this context, it is recommended that national legislation clearly differentiate between the responsibility of the author of hate speech and that of the media and media professionals who, through their actions, promote or facilitate its spread.

Hate speech can serve as a tool to intimidate or silence minority groups, human rights defenders, and representatives of non-governmental organizations. As a complex and multidimensional phenomenon, it poses significant threats to the stability of democratic societies. It is important to recognize that such destructive rhetoric affects not only the dignity and human rights of individuals but also entire minority or targeted groups. History has shown that hate speech has been deliberately used to mobilize groups and societies against each other, leading to violent escalation, hate crimes, war, and genocide (Council of Europe, 2022).

In an effort to address this pressing issue, the United Nations designated June 18 as the International Day Against Hate Speech.

Hate speech directly contradicts the values of tolerance, inclusion, and diversity, as well as the core principles of human rights. Moreover, such rhetoric can expose individuals and groups to discrimination, abuse, violence, and even social and economic exclusion. Addressing and combating hate speech, therefore, requires a holistic approach and the collective mobilization of society (United Nations, 2023).

The European Court of Human Rights defines the right to freedom of expression in Article 10 as follows: “Everyone has the right to freedom of expression; this right includes freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.” However, respect for this right also entails certain responsibilities, and it may be subject to legal restrictions and penalties, but only when necessary in a democratic society (Kovačević, 2023).

Freedom of speech and expression is a fundamental aspect of human rights, but its boundaries extend only as far as the rights and freedoms of others. In other words, restrictions are determined by the interests of individuals or society as a whole. As stated in the Constitution of the Republic of Serbia (Official Gazette of the Republic of Serbia No. 98/2006 and 115/2021, 46): “Freedom of expression may be limited by law if necessary to protect the rights and

reputation of others, to preserve the authority and impartiality of the court, and to protect public health, morals in a democratic society, and national security.”

Hate speech is becoming increasingly prevalent on social networks. In response, the Canadian KnowledgeFlow Foundation (2024), as part of its efforts to create a safe and secure digital environment for all citizens, proposes five key steps to combat hate speech online.

Recognize

The first step in establishing a barrier against hate speech is recognizing it, which can be categorized into three main forms:

- **Dehumanization and Demonization:** Dehumanization involves portraying members of a group as less than human, while demonization labels them as a superhuman threat that must be neutralized. Both techniques serve to justify violence and persecution.
- **Violence and Incitement to Violence:** This category includes any rhetoric that explicitly calls for or encourages violence against members of a group.
- **Early Warning:** This form of speech does not necessarily involve dehumanization or violent language but manifests as subtle generalizations that can contribute to negative stereotyping, potentially escalating into more dangerous discourse.

Responses

A key step in combating hate speech online is to respond to it. While it may not always be possible to change the views of those who spread hate speech, challenging misleading claims can encourage others to think critically and question harmful narratives.

Support

Fighting hate speech also means ensuring the safety and well-being of its victims. This involves showing solidarity with those targeted and offering support to help counter the effects of online harassment.

Report

Most social media platforms provide tools to report accounts or posts that violate their code of conduct. Additionally, users can block harmful content to prevent further exposure.

Learn

The final step in countering hate speech is staying informed and aware of the harmful consequences of sharing or engaging with hateful content.

Hate speech is embedded in the *pyramid of hate*, a model illustrating how bias, discrimination, and violence escalate. This theoretical and visual framework demonstrates how the seeds of hatred, once planted, can quickly evolve from biased attitudes into hate speech, discrimination, and even murder—ultimately, in extreme cases, leading to genocide (Rodriguez Martinez et al., 2024).

Figure 1. Pyramid of Hate



Source: Anti-Defamation League, 2018

The pyramid illustrates the progression of biased behaviors, increasing in severity from the bottom to the top. While behaviors at each level negatively impact individuals and groups, those at the upper levels pose life-threatening consequences. Importantly, the existence of higher-level behaviors is sustained by those at the lower levels.

Journalists and other media professionals play an essential role in democratic societies, making it crucial that they have the editorial freedom and independence necessary to fulfill their duties. This is why protecting journalists, especially those targeted by hate speech, is emphasized in the *Recommendations of the Committee of Ministers of the Council of Europe to Member States* (2016):

“It is alarming and unacceptable that journalists and other media actors in Europe are increasingly exposed to threats, harassment, surveillance, intimidation, arbitrary deprivation of liberty, physical attacks, torture, and even murder for their research, opinions, or reporting, especially when their work is directed at uncovering abuse and human rights violations.”

The trial before the International Military Tribunal in Nuremberg (1945–46) marked the first conviction in history for what is now recognized as hate speech. Julius Streicher was sentenced to death based on his writings and speeches, which targeted a specific social group, the Jewish people, on ideological grounds. His conviction for crimes against humanity was rooted in the belief that, through his relentless promotion of hatred, primarily via the magazine he personally edited, he played a significant role in the psychological "infection" of the German public, spreading years of brutal propaganda.

2. LITERATURE REVIEW

Different national approaches to defining hate speech, along with concerns over restrictions on freedom of speech, are the central issues explored by Antić (2023) in his work *Sanctioning Hate Speech on the Internet: In Search of the Best Approach*. He argues that these divergent perspectives present the greatest obstacle to establishing effective mechanisms against hate speech. At the same time, requiring online platforms to implement self-regulatory mechanisms for removing hate speech content has raised fears of inappropriate censorship and undue restrictions on freedom of expression.

Similarly, Knol Radoja and Bodul (2022) emphasize the challenges of protecting human rights in the digital space, highlighting the inadequacy of legal frameworks at both national and international levels. Striking the right balance between media freedom and individuals' right to protection from hate speech remains highly complex. Živanovski (2014) suggests that media intermediaries should promptly respond to inappropriate content but cautions that preemptively filtering user comments, as per the positions of the European Court of Human Rights and the Court of Justice of the EU, is incompatible with internet freedom of expression.

Two primary models have emerged in addressing hate speech in communication: the European and the American approaches. In the U.S. legal system, hate speech is considered part of free expression, whereas in Europe, it is viewed as an abuse of this fundamental right. The core disagreement lies in defining the boundaries of free speech, as even well-intended efforts to curb hate speech can risk evolving into totalitarian restrictions on differing opinions and criticism.

Europe is increasingly prioritizing the dignity of social groups over absolute freedom of speech. Ristivojević (2008) warns that restrictions on free expression are becoming more numerous, extensive, and frequently enforced. He argues that only public hate speech, speech that incites hatred and infringes on the honor and dignity of social groups, warrants criminal legal intervention. In most cases, such speech manifests as a combination of both elements, with hatred typically incited through insults, belittlement, or slander.

Pukallus and Arthur (2024), in their study *Combating Hate Speech on Social Media: Applying Targeted Regulation, Developing Civil-Communicative Skills, and Utilizing Local Evidence-Based Anti-Hate Speech Interventions*, conclude that hate speech is not harmless but a powerful driver of group radicalization. This, they argue, is why it has become a central issue in the “digital-communicative battlefield.” Addressing this challenge requires: (a) regulating and redesigning social media platforms to prevent their role in amplifying hate speech and (b) empowering citizens, both individual users and local civil associations, to recognize and actively combat online hate speech.

Dagen (2022), in his study *Legal Challenges of the Croatian Legislature with Regard to Hate Speech and the Commission of Certain Criminal Offenses*, examines cases of the (mis)use of freedom of expression - one of the highest human, social, and humanistic evolutionary values. The author highlights how, as early as the 1920s, dysfunctions and anomalies in the public media space began to emerge, particularly with the rise of modern electronic platforms that reach large audiences. The abuse of freedom of expression has since led to the spread of hate speech and the commission of certain criminal offenses.

Hate speech and fake news have proven to be highly disruptive in African countries, weakening fragile political systems and undermining peace, stability, and efforts toward national cohesion. Bukar, Mohammed, and Zannah

(2021) highlight that this phenomenon is particularly toxic in the Nigerian public sphere, yet the concepts remain insufficiently understood. In their study, which deliberately sampled data from student transcripts and audiovisual sources, the authors conclude that Nigerians may unconsciously engage in hate speech. However, they emphasize that this is not a new issue but a long-standing problem rooted in unresolved socioeconomic and political imbalances.

3. HATE SPEECH IN THE MEDIA IN THE REPUBLIC OF SERBIA

Although the existing legal framework and longstanding commitment to zero tolerance for hate speech in the media are in place, research indicates that these formal provisions are neither sufficiently effective, dissuasive, nor proportionate. Analyzing individual texts alone is not enough to determine the presence of hate speech; instead, a comprehensive assessment of both content and context over an extended period is necessary to accurately evaluate editorial policy.

The Constitution of the Republic of Serbia (*Official Gazette of the Republic of Serbia, No. 98/2006 and 115/2021, 46*) guarantees “freedom of opinion and expression, as well as the freedom to seek, receive, and disseminate information and ideas through speech, writing, images, or any other means.” Raising public awareness of the importance of respecting pluralism, highlighting the dangers of hate speech, ensuring publicly accessible codes of ethics for public media services, providing training for journalists and editors on non-discrimination and hate speech, and promoting professionalism are among the key practical recommendations for enhancing media standards (Krstić, 2020, 63).

The European Commission, in its 2020 report, concluded that media freedom and freedom of expression in Serbia had not improved. The report highlighted that the editorial landscape favored tabloidization and that journalistic standards remained low. In its six-month working document (*Non-paper*), published on June 11, which reviewed progress in chapters 23 and 24 (judiciary and human rights) of the EU accession negotiations, the Commission emphasized the need to improve media conditions and strengthen the independence of the Regulatory Authority for Electronic Media (REM) to effectively safeguard media pluralism (Belgrade Center for Human Rights, 2021).

The fourth report of the European Commission against Racism and Intolerance (ECRI), published in late June 2024, expressed concern over the persistence of hate speech in political and public discourse, particularly online. The groups most frequently targeted include Roma and other ethnic minorities, LGBTI individuals, refugees, and migrants. The report also noted a troubling paradox: nearly half of those responsible for hate speech belong to professions and institutions that should be protected from it, including media workers, politicians, state officials, and political parties (Krstić, 2024).

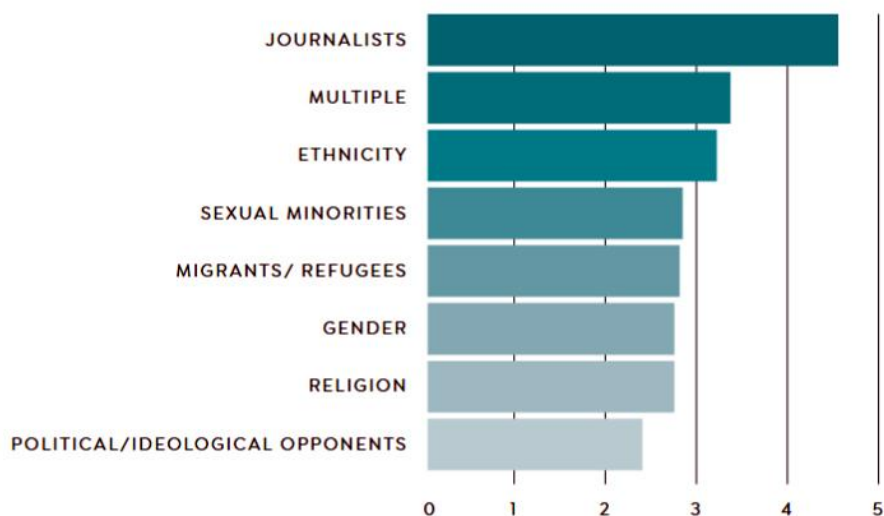
A study by the Center for Media Professionalization and Media Literacy (2024), titled “*Communicative Aggression in Serbia*”, conducted from September 1 to 30, 2024, found that in just one month, the most widely read print and online media published 21,638 texts containing elements of labeling, hate speech, and sensationalism. Compared to a similar study conducted in 2019 with an almost identical representative sample, three key findings emerged: a continuous rise in aggressive communication - predominantly in online media, evolving trends in labeling and hate speech, and an increasingly intense militaristic narrative that further fuels and amplifies communicative aggression. Hate speech, rooted in years of accumulated prejudice, anxiety, and distorted values, primarily targets individuals based on their origin, gender, sexual orientation, and religious affiliation. The anonymity and lack of direct confrontation on the internet and social networks significantly contribute to its spread. Stigmatization is further fueled by the accessibility and appeal of online content, its rapid sharing without critical scrutiny, and the minimal costs associated with its distribution.

A study on hate speech in Serbia, conducted by the Western Balkans Media Diversity Institute from January 15 to April 11, 2022, aimed to assess media coverage of hate speech, analyze prevalent narratives, and identify key sources of hate and discriminatory discourse in Serbian media. The research monitored newspapers, television, online news portals, radio, and social networks. The average severity score across all cases was 2.89, with hate speech targeting journalists receiving the highest average score of 4.5. Reported incidents were rated on a scale from 1 to 6: (Jovanović & Anđušić, 2022)

- **Disagreement** – Rhetoric that challenges an idea on a cognitive level, including questioning claims, beliefs, or ideologies of a group and attempting to change them.
- **Negative Actions** – Rhetoric that highlights negative but nonviolent actions associated with a specific group.
- **Negative Character** – Rhetoric that includes nonviolent insults or negative characterizations of individuals or groups.
- **Demonization and Dehumanization** – Rhetoric that attributes subhuman or superhuman traits to a targeted individual or group.

- **Incitement to Violence** – Rhetoric that advocates for or encourages physical harm, whether directly or implicitly.
- **Death** – Rhetoric that explicitly calls for the killing or eradication of a group.

Graph 1. Average scores per type of group/individual targeted by hate speech



Source: Jovanović & Anđušić, 2022

Journalists face a challenging professional and ethical dilemma: should they report on hate speech, potentially amplifying harmful content, or fulfill their duty to inform the public by exposing and analyzing such messages? This dilemma becomes even more complex when hate speech escalates into violence, murder, or protests. In such cases, journalists are compelled to report on the events while simultaneously navigating the challenge of addressing the underlying causes without inadvertently legitimizing or spreading the very rhetoric that incited them.

3. CONCLUSIONS

Hate speech has a particularly devastating impact on young people, who are often the primary targets of a precisely crafted narrative that violates human rights. Therefore, future efforts should focus on those who have the potential and the responsibility to drive change. These individuals must play a crucial role in recognizing problems early, raising awareness, and actively opposing harmful hate rhetoric. The threats embedded in hate speech are increasingly accelerating from inappropriate communication to unacceptable manifestations of aggression, physical violence, and crime. Addressing this persistent issue requires solutions not only at a global level but also within local and regional contexts. This approach is grounded in the understanding that counter-discourse is one of the most effective strategies in combating the many complex challenges posed by hate speech. It also acknowledges that displacing targeted individuals or groups from digital spaces to local settings does not eliminate the consequences of virtual hate speech. Enhancing media literacy, fostering critical thinking, challenging prejudices and stereotypes, and promoting tolerance and empathy provide a strong and necessary foundation for combating hate speech. This multidisciplinary approach has the potential to make both digital and traditional media safer and more secure through a sustained, evidence-based fight against hate speech.

REFERENCES

- ADL. 2018. Nota adaptada de Pyramids of hate (Figure), por Anti-Defamation League. Available at: <https://www.adl.org/sites/default/files/documents/pyramid-of-hate.pdf> (Accessed date: February 28, 2025).
- Antić, P. (2023). Sanctioning hate speech on the internet: in search of the best approach. *Pravni zapisi*, 14(1), 74-100. <https://doi.org/10.5937/pravzap1-43118>
- Beogradski centar za ljudska prava. (2021). *Ljudska prava u Srbiji 2020 – pravo, praksa i međunarodni standardi ljudskih prava*.

- Bukar, A.A., Mohammed, S, Zannah, M. (2021). Conceptual discourse on hate speech and fake news in Nigeria's public sphere: some implications for national security. Annual African Council for Communication Education (ACCE) Conference, Faculty of Communication, Bayero University, Kano, 25–28.05.2021.
- Centar za profesionalizaciju medija i medijsku pismenost - CEPROM (2024). Kako smo od poželjno negativnog stigli do obavezno agresivnog. (28/12/2024). Available at: <https://www.ceprom.rs/2024/12/28/kako-smo-od-pozeljno-negativnog-stigli-do-obavezno-agresivnog/> (Accessed date: March 1, 2025)
- Council of Europe (1997). Recommendation No. R (97) 20 of the Committee of Ministers to member states on “hate speech” adopted on 30 October 1997
- Council of Europe. (2022). Combating hate speech. Available at: <https://rm.coe.int/prems-083822-gbr-2018-recommendation-on-combating-hate-speech-memorand/1680a710c9> (Accessed date: December 25, 2024).
- Dagen, T. (2022). Legal challenges of the Croatian legislature with regard to the hate speech and committing certain criminal offenses. *Zbornik radova Pravnog fakulteta u Novom Sadu*, 56(3), 927-946. <https://doi.org/10.5937/zrpfns56-40442>
- Gujaničić, D. (2023). Jedinstveni slučaj Julijusa Štrajhera – prvo međunarodno suđenje u istoriji za govor mržnje. *Politička revija*, 78(4), 51-69. <https://doi.org/10.5937/pr78-45978>
- Jovanović, I., & Anđušić, A. (2022). Monitoring report on hate speech in Serbia. Media Diversity Institute and Media Diversity Institute Western Balkans. Available at: https://www.reportingdiversity.org/wp-content/uploads/2022/07/MRHS_Serbia.pdf (Accessed date: March 1, 2025)
- Knol Radoja, K., Bodul, D. (2022). Kako zaustaviti plasiranje neistinitih informacija i govor mržnje na internetu? *Zbornik radova Pravnog fakulteta u Nišu*, 61(95), 135-151. <https://doi.org/10.5937/zrpfns1-40003>
- KnowledgeFlow. (2024). 5 Positive strategies to counter hate speech online. Available at: <https://knowledgeflow.org/5-ways-you-can-help-counter-hate-speech-online/> (Accessed date: March 7, 2025)
- Kovačević, M. (2023). Sprečavanje govora mržnje u praksi Evropskog suda za ljudska prava. *Megatrend revija*, 20(2), 79-92. <https://doi.org/10.5937/MegRev2302079K>
- Krstić, I. (2020). *Izveštaj o upotrebi govora mržnje u medijima u Srbiji*. Available at: <https://rm.coe.int/hf25-hate-speech-serbian-media-srp/1680a2278f> (Accessed date: January 31, 2025)
- Krstić, I. (2024). Prisutnost govora mržnje u medijima u Srbiji. Savet Evrope. Available at: <https://rm.coe.int/3-prisutnost-govora-mrznje-u-medijima-u-srbiji-2024/1680b3d563> (Accessed date: March 6, 2025)
- Official Herald of the Republic of Serbia, Nos. 98/2006 and 115/2021, 46. Constitution of the Republic of Serbia.
- Pukallus, S., & Arthur, C. (2024). Combating Hate Speech on Social Media: Applying Targeted Regulation, Developing Civil-Communicative Skills and Utilising Local Evidence-Based Anti-Hate Speech Interventions. *Journalism and Media*, 5(2), 467-484. <https://doi.org/10.3390/journalmedia5020031>
- Recommendation CM/Rec(2016)4 of the Committee of Ministers to member States on the protection of journalism and safety of journalists and other media actors (Adopted by the Committee of Ministers on 13 April 2016 at the 1253rd meeting of the Ministers' Deputies)
- Ristivojević, B. (2008). Granica između slobode govora i dostojanstva društvenih skupina u odabranim evropskim krivičnim pravima – govor mržnje. *Zbornik radova Pravnog fakulteta u Novom Sadu*, 42(3), 211-246.
- Rodriguez Martinez, P., Martinez Joya, L., & Villegas Lirola, F. (2024). Hate-Speech Countering by Immigrant and Pro-Immigrant Associations in Almeria (Spain). *Social Sciences*, 13(1), 33. <https://doi.org/10.3390/socsci13010033>
- United Nations (2023). Hate Speech – Impact and prevention. Available at: <https://www.un.org/en/hate-speech/impact-and-prevention/why-tackle-hate-speech> (Accessed date: March 4, 2025)
- Živanovski, N. (2014). Različiti tretman govora mržnje u procesima komunikacije u Evropi i SAD. *CM – časopis za upravljanje komunikacijama*, 9(32), 57-68. <https://doi.org/10.5937/comman1432057Z>