MEDIA LITERACY AND CRITICAL THINKING: DEFENDING AGAINST HATE SPEECH

Branislav Sančanin[1]
Aleksandra Penjišević[2]

ABSTRACT

Continuous media literacy represents a key defensive determinant that clearly distinguishes between a constructive critical narrative and hate speech as an unacceptable media output. Excessive legal regulation, through over-norming, can in practice severely undermine the right to freedom of expression, creating broad space for both censorship and self-censorship. A particular challenge is posed by hate speech published on social media, in terms of its detection, dissemination, removal, and the imposition of sanctions proportionate to the harm caused. The primary research was conducted to examine how and to what extent the citizens of the Republic of Serbia perceive media content and its influence, with special emphasis on hate speech. The research sample was based on opportunistic, convenient sampling and included residents of the Republic of Serbia from each of the four regions (n=186). Data were collected from March to June 2025. The findings confirm the importance of media literacy and critical thinking in the fight against hate speech, which can have devastating effects on individuals, communities, and society as a whole. Respondents emphasized that the absence of sanctions proportionate to the harm inflicted on individuals or groups is one of the main reasons for the persistence of hate speech in the media.

KEYWORDS: *media literacy, hate speech, critical thinking, social media*

MEDIJSKA PISMENOST I KRITIČKO RAZMIŠLJANJE: ODBRANA OD GOVORA MRŽNJE

SAŽETAK

Medijska pismenost je kontinuirani proces koji se pozicionira kao ključna odbrambena determinanta neophodna za razumevanje razlike između konstruktivnog kritičkog narativa i govora mržnje, kao nedopustivog medijskog proizvoda. Rigidna zakonska regulativa, svojim prekomernim normiranjem, u praksi može ozbiljno da uruši pravo na slobodu izražavanja, stvarajući širok prostor za cenzuru i autocenzuru. Poseban izazov predstavlja govor mržnje objavljen na društvenim mrežama, s aspekta njegovog detektovanja, distribucije, uklanjanja i sankcija srazmernih pričinjenoj šteti. Primarno istraživanje je sprovedeno s ciljem ispitivanja na koji način i u kojoj meri stanovnici Republike Srbije doživljavaju medijske sadržaje i nji-

1 Professor, The College of Applied Studies for Management and Business Communication, Belgrade, Serbia, phone number: +381658981898, https://orcid.org/0000-0002-4255-2634; e-mail: sancanin.branislav@gmail.com

2 Associate professor, University "Union – Nikola Tesla", Cara Dušana 62-64, Belgrade, Serbia, phone number: +38163275674, https://orcid.org/0000-0002-0898-6818; e-mail: aleksandra.penjisevic@gmail.com

hov uticaj, s posebnim naglaskom na govor mržnje. Istraživački uzorak je bio opportunistic, convenient sampling, stanovnika Republike Srbije u svakom od četiri regiona (n=186). Podaci su prikupljani od marta do juna 2025. godine. Ustanovljena je važnost medijske pismenosti i kritičkog razmišljanja u okviru borbe protiv govora mržnje koji može da ima razorne uticaje na zajednicu, pojedince i društvo u celini. Ispitanici navode da je izostanak sankcija srazmernih nanetoj šteti pojedincu ili grupama, jedan je od najvažnijih razloga prisutnosti govora mržnje u medijima.

KLJUČNE REČI: *medijska pismenost, govor mržnje, kritičko razmišljanje, društveni mediji*

INTRODUCTION

Despite the absence of a universally accepted definition of hate speech, the term broadly refers to "any kind of communication in speech, writing, or behavior 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, color, descent, gender, or other identity factors" (Guterres, 2019).

Hate speech and toxic communication on the internet are exhibiting an exponential growth trend. Munn (2020) observes that explanations for this phenomenon often propose either technical solutions (e.g., automation) or non-technical ones (e.g., human content moderation), or regard hate speech as a natural expression of human behavior.

A study conducted in 16 countries between August 2022 and September 2023, involving 8,000 participants, revealed that two out of three individuals had encountered hate speech. The highest incidence was reported in India, where 85% of respondents stated they frequently came across hate speech, followed by Bangladesh (84%) and South Africa (79%). Hate speech and/or disinformation were perceived as particularly prevalent on Facebook and TikTok (Fleck, 2024). Within the European Union, 80% of individuals reported encountering hate speech online, while 40% stated they had felt threatened or attacked on social media platforms (Gagliardone et al., 2015).

AlKhamaiseh (2021) emphasizes that discrimination and hate speech in the media not only offend the sentiments of individuals or communities they target, but may also contribute to crimes against them, incite armed conflict, and even legitimize or encourage acts of violence against specific ethnic or national groups. Furthermore, such speech can fuel violence against particular demographic groups, including women, children, refugees, minorities, or political opposition figures.

Due to the nature of their responsible role in the public sphere, journalists are frequently targeted by hate speech. The growing number of incidents is cause for concern, as the consequences of such verbal attacks directly threaten the freedom of journalistic expression and undermine the public's right to timely and accurate information. A study conducted in Germany by Obermaier, Hofbauer, and Reinemann (2018) found that hate speech directed at journalists is a growing problem that negatively affects public attitudes toward media professionals. Hate speech targeting journalists can provoke negative emotions such as anxiety and anger, but it may also reinforce a sense of purpose in their professional work. The authors observed that journalists are more likely to adopt coping strategies to manage the emotional and cognitive impact of hate speech than to engage in actions aimed at preventing further incidents.

The aim of this study is to examine how and to what extent different social groups evaluate content on selected social media platforms, with a particular focus on the recognition and analysis of hate speech. Given the harmful effects that hate speech can have on individuals, communities, and society as a whole, special attention is devoted to promoting media literacy and critical thinking. This paper focuses on identifying the competencies necessary for recognizing, analyzing, and critically assessing media content as key strategies for mitigating the adverse consequences of digital communication.

The digitalization of media and the widespread use of social networks, characterized by eroded credibility and vague legislative and ethical standards, have acted as catalysts in the spread of hate speech, creating an environment marked by fear, division, and exclusion. In this context, efforts in the digital environment should be understood as part of a broader strategy aimed at producing positive outcomes, based on the firm belief that safety in virtual space should not be viewed as a privilege, but as a global and universal norm (Sančanin et al., 2025).

The author Poni Alice JameKolok (n.d.) outlines five approaches to combating hate speech:

- Education on media ethics: Education should focus on journalists' rights and freedoms, as well as their role in fostering and promoting peaceful societies. This is essential because journalists must possess the necessary knowledge and skills to identify hate speech and counter its messages.
- Encouraging conflict-sensitive reporting and multicultural awareness campaigns: These campaigns should emphasize the importance of respecting cultural and traditional diversity. Journalists are expected to adhere to professional standards and be able to write, broadcast, and engage with people impartially.
- Regulating social media: Regulation should aim to enhance education on media laws and ethics, without infringing upon the right to freedom of the press.
- Encouraging victims and witnesses to report hate speech-related crimes: Hate speech often remains invisible because victims may not know where to report such incidents or may not even realize they are victims of hate speech.
- Ending impunity for hate crimes: Establishing editorial units to monitor and evaluate trends in hate speech, and raising awareness among key institutions and civil society, can significantly contribute to holding perpetrators accountable.

Keen and Georgescu (2018) emphasize the necessity of establishing educational processes in the broad and complex field that encompasses media literacy within the educational system, highlighting the importance of acquiring competencies for the effective use of media and information. In this context, Santiago del Pino and Goenchea Permistina (2020) argue that the integration of media education into initial teacher training curricula could facilitate a transformation of social engagement through education into critical literacy among secondary school students, thereby contributing to the eradication of the growing prevalence of violent cultural patterns.

DATA, METHODOLOGY AND STYLIZED FACTS

The primary research was designed to examine how and to what extent the citizens of the Republic of Serbia assess media content and the ways in which it affects them, with particular emphasis on hate speech. In relation to this objective, the following research questions were posed:

Q1. How can technology and digital tools contribute to the promotion of media literacy and critical thinking as a means of countering hate speech?

Q2. What are the potential consequences of exposure to content containing hate speech on individuals' states and behavior?

Data were collected using *Google Forms*. The research sample was based on opportunistic, convenient sampling of residents from each of the four regions of the Republic of Serbia. A pilot study was conducted with 30 participants. Following an analysis of the pilot study, which included the assessment of content validity for all measured aspects, the final version of the questionnaire was developed and administered to a new group of 186 participants ($n = 186$). Data collection took place between March and June 2025.

The questionnaire consisted of two parts. The first part included general demographic questions, while the second part required respondents to assess, using a Likert scale, the influence of media content on individuals, including themselves, with particular attention to hate speech, as well as their own level of digital media literacy (1 = I don't agree at all, 7 = I completely agree).

To address the research questions, descriptive statistics, measures of dispersion, and measures of skewness were calculated. Hypotheses were tested using non-parametric techniques (Chi-Square Test, Mann-Whitney U Test). The data were processed using the SPSS statistical software package.

EMPIRICAL RESULTS AND DISCUSSION

The study included 186 participants, of whom 48.9% identified as male and 51.5% as female. In terms of regional distribution, 24.3% were from the Vojvodina region, 27.6% from Belgrade, 20.0% from Southern and Eastern Serbia, and 28.1% from the Šumadija and Western Serbia region.

Reflecting on the week prior to completing the survey, the internet—including watching videos, playing games, and using social media—had the greatest impact on respondents in the following ways:

1. *It kept me from feeling bored*: the average rating was 4.45, with a mode of 7 and a median of 5. The distribution shows moderate negative skewness and moderate variability (CV = 43.39%).
2. *It kept me from feeling lonely*: the average rating was 2.99, with a mode of 1 and a median of 2. The distribution exhibits strong positive skewness and very high variability (CV = 70.57%).

For further details, see Table 1.

Table 1.: Statistics

		It kept me from feeling bored	It kept me from feeling lonely	It made me feel depressed	It left me with anxious thoughts and feelings	It caused problems with my friends	It made me feel angry
N	Valid	164	166	158	157	157	158
	Missing	22	20	28	29	29	28
Mean		4.5427	2.9880	1.5063	1.6115	1.3694	1.7025
Median		5.0000	2.0000	1.0000	1.0000	1.0000	1.0000
Mode		7.00	1.00	1.00	1.00	1.00	1.00
Std. Deviation		1.97018	2.11771	1.11015	1.20708	1.06390	1.27449
Skewness		-.347	.621	2.686	2.559	3.582	2.371
Std. Error of Skewness		.190	.188	.193	.194	.194	.193
Percentiles	25	3.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	50	5.0000	2.0000	1.0000	1.0000	1.0000	1.0000
	75	6.0000	5.0000	1.2500	2.0000	1.0000	2.0000

The Mann-Whitney U test did not reveal any statistically significant differences between male and female respondents for any of the following items: *It kept me from feeling bored*, *It kept me from feeling lonely*, *It made me feel depressed*, *It left me with anxious thoughts and feelings*, *It caused problems with my friends*, and *It made me feel angry*.

According to respondents, inappropriate content is most frequently found on the social media platforms TikTok, YouTube, and Facebook.

When asked about the media and hate speech, the largest proportion of respondents (44.02%) answered *I don't know*, 21.2% stated it is *Prohibited*, while 34.8% believe it is *Permitted*.
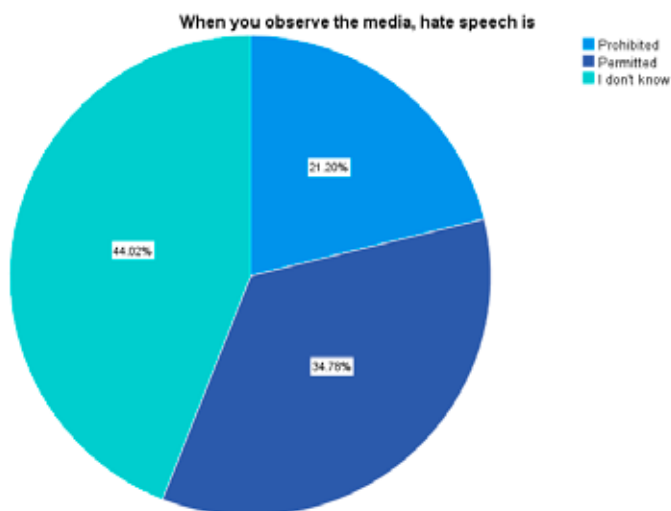
Graph 1 When you observe the media, hate speech is Source: Authors' calculation

Among respondents who believe that hate speech in the media is permitted, the reasons cited were: "no one is punished" (54.0%), "everyone uses it" (28.6%), and "it aims to strengthen and empower us" (17.5%).
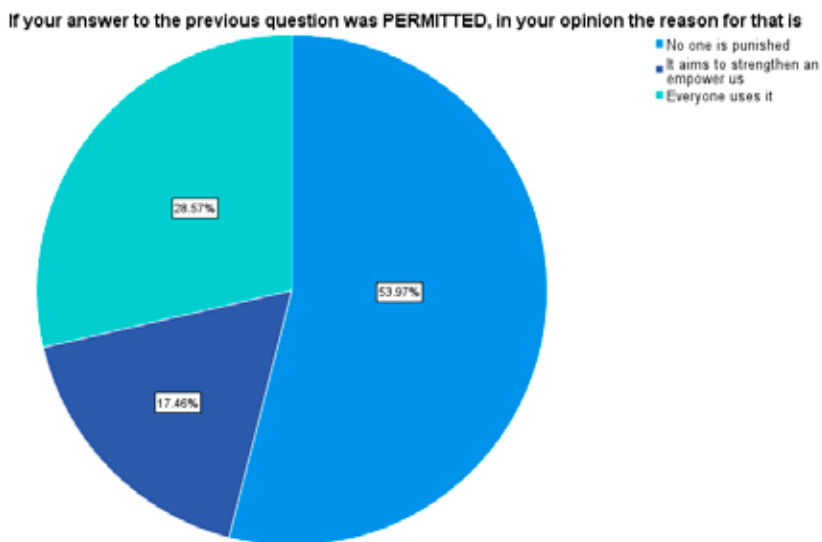
Graph 2 Razlozi za prisustvo govora mržnje u medijima Source: Authors' calculation

The Chi-Square test revealed a significant association between gender and the assessment of hate speech in media content, $\chi^2(2, n = 184) = 6.621$, $p = 0.037$, *phi* = 0.189. The belief that hate speech in the media is permitted was expressed by 43.3% of male respondents, compared to 26.6% of female respondents.

The Chi-Square test also demonstrated a significant association between the respondents' region and their assessment of hate speech in media content, $\chi^2(6, n = 183) = 18.315$, $p = 0.005$, *phi* = 0.224. The perception that hate speech is permitted in the media was most frequently reported by respondents from the Vojvodina region (53.3%) and the Belgrade region (37.3%).

Table 2.: Crosstab

			When you observe the media, hate speech is			
			Prohibited	Permitted	I don't know	Total
Region	Šumadija and West-ern Serbia	Count	13	14	25	52
		% within Region	25.0%	26.9%	48.1%	100.0%
		Adjusted Re-sidual	.8	-1.3	.7	
	Southern and East-ern Serbia	Count	5	6	24	35
		% within Region	14.3%	17.1%	68.6%	100.0%
		Adjusted Re-sidual	-1.1	-2.4	3.2	
	Belgrade	Count	13	19	19	51
		% within Region	25.5%	37.3%	37.3%	100.0%
		Adjusted Re-sidual	.9	.5	-1.2	
	Vojvodina	Count	8	24	13	45
		% within Region	17.8%	53.3%	28.9%	100.0%
		Adjusted Re-sidual	-.7	3.1	-2.4	
Total		Count	39	63	81	183
		% within Region	21.3%	34.4%	44.3%	100.0%

Source: Authors' calculation

The Chi-Square test revealed a significant association between gender and the perceived reasons for the presence of hate speech in the media, $\chi^2(2, n = 63) = 9.836$, $p = 0.009$, *phi*= 0.387. Among female respondents, 76.9% identified the lack of punishment as the main reason for the presence of hate speech, compared to 37.8% of male respondents.

The Chi-Square test also indicated a significant association between the respondents' region and the perceived reasons for the presence of hate speech in the media, $\chi^2(6, n = 62) = 25.250$, $p < 0.001$, *phi*= 0.451. The belief that hate speech is permitted in the media because "no one is punished" was shared by 76.9% of respondents from the Vojvodina region, 61.5% from Šumadija and Western Serbia, 40.0% from Southern and Eastern Serbia, and the lowest percentage was recorded among respondents from the Belgrade region (16.7%).

Table 3.: Crosstab

			If your answer to the previous question was PERMITTED, in your opinion the reason for that is			
			No one is punished	It aims to strengthen and empower us	Everyone uses it	Total
Region	Šumadija and Western Serbia	Count	8	1	4	13
		% within Region	61.5%	7.7%	30.8%	100.0%
		Adjusted Residual	.7	-1.1	.2	
	Southern and Eastern Serbia	Count	2	0	3	5
		% within Region	40.0%	0.0%	60.0%	100.0%
		Adjusted Residual	-.6	-1.1	1.6	
	Belgrade	Count	3	9	6	18
		% within Region	16.7%	50.0%	33.3%	100.0%
		Adjusted Residual	-3.7	4.3	.5	
	Vojvodina	Count	20	1	5	26
		% within Region	76.9%	3.8%	19.2%	100.0%
		Adjusted Residual	3.2	-2.4	-1.4	
Total		Count	33	11	18	62
		% within Region	53.2%	17.7%	29.0%	100.0%

Source: Authors' calculation

CONCLUSIONS

Achieving zero tolerance toward incitement to violence and discrimination in the public sphere, as well as advocacy of any form of hostility in media content, is attainable only through continuous media education and the improvement of legal regulations and ethical standards. A rigid legislative framework maintains the position that the existence of hate speech is conditioned by the intent to incite and advocate hatred, meaning that the mere distribution of such content is not sufficient to cross the threshold that would qualify the act as a criminal offense. Numerous and diverse manifestations of hate speech on social media platforms are particularly significant for legal and ethical analysis, given the evidentiary challenges and the complexity of enforcing available legal remedies. Social media significantly contributes to the creation of "echo chambers," making individuals and groups even more vulnerable.

The findings of this study highlight the importance of media literacy and critical thinking. The internet—especially activities such as watching videos, playing games, and using social media—has had the greatest impact on preventing boredom and feelings of loneliness among individuals. This perception was shared by respondents of both genders and from all regions. Inappropriate content is most frequently disseminated via social media platforms such as TikTok, YouTube, and Facebook. The research revealed that 34.8% of respondents believe that hate speech in the media is permissible. The most significant reason identified for the dissemination of hate speech through media content is the lack of

sanctions, a view more frequently expressed by women and respondents from the Vojvodina region.

The authors emphasize the importance of encouraging the entire media industry to remain committed to promoting and affirming media literacy. Likewise, media consumers are urged to follow this example and adopt a critical distance toward information that is often the product of user-generated content, shaped by personal creativity and subjective inspiration.

REFERENCES

- AlKhamaiseh, M. (2021). The problem with hate speech: How the media has fuelled its rise. Aljazeera Media Institute. Available at: https://institute.aljazeera.net/en/ajr/article/1697 (Accessed date: June 10, 2025)
- Fleck, A. (2024). Hate Speech: 2 in 3 People Often Encounter Hate Speech Online. *Statista*. Available at: https://www.statista.com/chart/33299/online-hate-speech-encounters/ (Accessed date: June 20, 2025)
- Gagliardone, I., Gal, D., Alves, T., Martinez, G. (2015). *Countering online Hate Speech*. UNESCO. Paris, France.
- Guterres, A. (2019). United Nations Strategy and Plan of Action on Hate Speech. Available at: https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan %20 of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf (Accessed date: April 25, 2025)
- Keen, E. & Georgescu, M. (2018) Orientaciones – Manual para combatir el discurso de odio en internet a través de la educación en derechos humanos. Instituto de la juventud (INJUVE). Available at: https://www.injuve.es/sites/default/files/2019/07/publicaciones/orientaciones.pdf (Accessed date: May 13, 2025)
- Munn, L. (2020). Angry by design: toxic communication and technical architectures. *Humanities and Social Sciences Communications*, *7*(1), 1-11. https://doi.org/10.1057/s41599-020-00550-7
- Obermaier, M., Hofbauer, M., & Reinemann, C. (2018). Journalists as targets of hate speech. How German journalists perceive the consequences for themselves and how they cope with it. SCM Studies in Communication and Media, 7, 499–524. https://doi.org/10.5771/2192-4007-2018-4-499
- Poni Alice JameKolok (n.d). 5 ways to counter hate speech in the Media through Ethics and Self-regulation. UNESCO. https://webarchive.unesco.org/web/20190621005258/https://en.unesco.org /5-ways-to-counter-hate-speech
- Sančanin, B, Penjišević, A. & Stefanović, D. (2025). Countering Hate Speech in the Media: Challenges and Solutions. *Knowledge – International Journal*, 69(1), 237-242. Paper presented at XLVIII International Scientific Conference Knowledge without borders – Istanbul, Turkey, April 3-6, 2025. Available at: https://ojs.ikm.mk/index.php/kij/article/ view/7228 (Accessed date: April 21, 2025)
- Santiago del Pino, M. & Goenchea Permistina, C. (2020). Critical Literacy and the Media: Teachers in Training in the Face of Hate Speech. *Investigaciones Sobre Lectura*, 14, 181-196. https://doi.org/10.37132/isl.v0i14.328