



HUNGARIAN NATIONAL REPORT

Monitoring the online antigypsist hate speech

2024-2025

Authors: Trendl Fanni, Horváth Aladár

Monitoring crew, contributors: Fehér György,
Kovács Dezdemóna Mária, Markócs-Mezei Tímea,
Menyhért Attila, Végh Zoltán

Editor: Kiss Tibor Noé

Responsible for publication: Romaversitas Alapítvány,
Török Ildikó

Printed at:
Bolko-Print Kft.
7631, Pécs, Fűzes dűlő 23.



Abstract

The Together Against Antigypsyism Online (TAAO) project focuses to address and counter antigypsyist hate speech online by working with young Roma people, using a combination of training and monitoring in order to obtain two-fold results: the training methodology serving the purpose to achieve a direct change in attitudes and actions of a high number of young people, and the monitoring to contribute to better understanding and a more systemic change of policies through advocacy based on the results.

The following report represent a deliverable of the TAAO project in relation to the project's overall objectives:

- To reach out to young people in Germany, Hungary, Czech Republic, Romania, Bulgaria, Slovakia through training activities in universities or youth clubs to raise their awareness of the adverse impact of hate speech on Roma and to empower them to recognise and report hate speech online.
- To empower young Roma people in 6 countries to monitor hate speech in online media and report cases of hate speech to IT companies, national equality bodies and relevant state institutions to reinforce implementation of anti-discrimination legislation. Their work will lead to national datasets and several reported cases.

More in-depth, the reports offer insight of alignment to the short-term results, which are directly contributing to the aforementioned objectives:

- The training activities as well as the media monitoring by the young Roma monitors themselves will lead to a higher number of reported cases of hate speech and a higher number of antigypsyist content taken down from online platforms. During the project 3147 cases were monitored, 40 cases were reported overall at the six countries.
- Increased awareness of the negative impact of antigypsyist hate speech among the young people targeted, leading to a change of attitude among the young people.
- Increased self-confidence and skills of 40 young Roma who are participated at the training activities.

When it comes to the Roma people: The people most suffering from hate speech are obviously those targeted – Roma, citizens of Europe. The project will have a positive long-term impact on them if hate speech content is reduced in online media as a result of this project. Less hate speech online means less incitement to violence and less stereotyping by the majority society, leading to a general decrease of antigypsyism.

The present Report focuses on Hungary and is a result of a one-year-long work containing data gathered for 12 months (October 2024– September 2025) and training by 5 young people, aimed to investigate the adverse impact of and raise awareness of anti-Roma online hate speech.

With the results presented in this report, we aim to take a step further to learn about the presence and impact of antigypsyist online hate speech and to learn lessons that can be brought forward to make the future of an equal and inclusive society in a safe digital environment happen.

1. Introduction

1.1. Background and context

„Together against antigypsyism online (TAAO) in Europe”, investigates the extent, forms, and impact of online anti-gypsyism, along with the effectiveness of counter-actions, across six European nations: Bulgaria, Czech Republic, Germany, Hungary, Romania, and Slovakia. Conducted over a 12-month period from October 2024 to September 2025, the research employed a mixed-methods approach. It involved 30 young Roma and non-Roma monitors who collected quantitative data from 3147 cases using a standardized monitoring tool (MT) and conducted qualitative analysis of 54 selected case studies. The methodology included sentiment analysis across various topics and content styles, and aggressivity scales for both explicit and implicit forms of hate speech.

1.2. National context

Besides focusing on the goals of TAAO, our project aims, and initiatives were informed by the Hungarian context of anti-Roma hate speech. This is important not only because it shows the relevance of local thematicization of hate speech, but also because, due to the different historical, cultural contexts and the resulting different regulations in different countries, some degree of context-specific responses and advocacy is needed to achieve change.

In connection with hate speech in the Hungarian regulations, different areas of laws and their respective regulations each serve as a yardstick against online hate speech:

- Based on the civil law aspect, a member of the community can take action against hate speech and demand damages.¹
- Act CXXV of 2003 on Equal Treatment and the Promotion of Equal Opportunities is a law against segregation, discrimination, and harassment, and retribution.²
- Criminal law wording is key in the context of hate speech, which prohibits incitement to violence.³
- Media law states that no media content may be an incitement to hatred or exclusion.⁴

Various analyzes show that, apart from these numerous areas of law concerning hate speech, the Hungarian judicial practice interprets hate speech as extremely narrow. There have been attempts in the past to change this narrow standard, but this has several times failed in the Constitutional Court. In addition, just as there is no uniform concept of hate speech in international law, as demonstrated above, in the Hungarian context various stakeholders' measures and respective laws simultaneously apply in the issue of online hate speech. In the absence of a unified concept, the field of hate speech is always somewhat vague, and procedural doubts can arise in specific cases.

According to a report by the European Commission against Racism and Intolerance (ECRI)⁵, in addition to law enforcement

1 About Act V of 2013 on Civil Rights: <https://njt.hu/jogszabaly/2013-5-00-00>

2 Act CXXV of 2003: <https://net.jogtar.hu/getpdf?docid=A0300125.TV&targetdate=20170701&printTitle=2003.+%C3%A9vi+CXXV.+t%C3%B6rv%C3%A9ny>

3 Btk. 332. §: <https://buntetojog.info/kulonos-resz/btk-332-%C2%A7-kozossegelleni-uszitas/>

4 Act CIV of 2010 on the freedom of the press and the basic rules on media content: <https://net.jogtar.hu/jogszabaly?docid=a1000104.tv>

5 ECRI – Country Monitoring in Hungary. 2015. <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/hungary>

problems, the situation is further exacerbated by the fact that state action against anti-Roma hate speech is not only lacking in Hungary, but open anti-Roma hate speech is not limited to extremist parties and groups.

This omission in state practice is further aggravated by the fact that in 2015 the European Commission drew Hungary's attention to the requirement under Article 1(1)(a) of Council Framework Decision 2008/913/JHA, which mandates not only the criminalisation of incitement to violence but also of incitement to hatred. The provision requires Member States to take the necessary measures to ensure that—among other things—public incitement to violence or hatred 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, is punishable.

This view is reinforced by the 2020 principled opinion⁶ of Erzsébet Szalayné Sándor, Deputy Commissioner responsible for protecting the rights of nationalities living in Hungary. In this opinion, issued in connection with potential hate speech targeting the German nationality, she makes the following recommendations:

“I recommend that the Minister of Justice, in cooperation with the Prosecutor General, examine why the amendment of Section 332 of the Criminal Code—specifically the inclusion of ‘incitement to violence’ as a form of conduct—has not resulted in substantive change in the judicial practice concerning incitement against a community, and how the current approach could be modified, particularly in police procedures.

6 Principled Opinion No. 5/2020 on Statements Offending the German Nationality Community, on the Need to Combat Hate Speech, and on the Possibilities for Addressing Harmful Content Appearing on Video-Sharing Platforms.

I further recommend that the Minister of Justice continue to encourage and strengthen professional dialogue within the framework of the Digital Freedom Working Group on issues related to combating online hate speech and on the limitations of the operation of video-sharing platforms.”

As accentuated by the report, Hungary is “climate of impunity” is rampant. Influential figures’ degrading, racializing, sexualizing statements on vulnerable and marginalized groups are normalized in the public discourse. Several media platforms disseminate openly racist content without any restrictions. The report further states that despite its cruciality, online hatred does not receive enough attention, which is well exemplified by the fact that Hungary has not signed the Additional Protocol to the Convention on Cybercrime yet.

The frequently racist rhetoric against Roma by public and political actors clearly shows that the political climate in Hungary is permissive towards hate speech. However, as the juridical practice on hate speech is relatively narrow, many of these discriminatory statements barely has legal consequences.

Romaversitas Foundation previously was part of PECAO (Peer Education to Counter Antigypsyist Hate Speech Online) project, where important and relevant conclusion were published. To summarize these findings we can write that „in connection with the Anti-Roma hate speech, our observations show that due to its ubiquitous nature and actuality it is important to come up with up-to-date and complex solutions. We were witnessing an age in which online communication, which is global and accessible to all, has suddenly grown, making it difficult to keep pace with regulations to protect the dignity of human rights, both on international and national levels. Those who are cumulatively exposed to prejudice and exclusion, both at the institutional level and in everyday life,

may become the primary object of hatred in cyberspace as well. Thus, online hate speech is a crucial issue, for various reasons: it influences public opinion, fuels tension, and paves the way for discrimination and hate crimes. Online media plays a special role in spreading and inciting hate speech. It reinforces stereotypes, uses offensive language, denies or downplays anti-Roma sentiment. Through social media, hate speech reaches millions of people and allows perpetrators to incite hatred and violence anonymously.”⁷

2. Methodology

2.1. Purpose and design

The aim of the present methodology is to provide comprehensive data on the extent, reach, impact, authors and recipients of and reactions to antigypsyist hate speech online, in order to inform policies and actions to combat antigypsyist hate speech online.

Based on our research proposal, the conceptual framework underpinning the methodology draws upon a theoretical approach informed by previous studies conducted (PECAO, 2022). This framework is structured around the following key pillars: anti-gypsyism (2018), cyber hate/online hate speech targeting Roma, the phenomena of reverse racism and reverse victimhood (Bonilla-Silva, 2006; Taylor & Martinez, 2020), the spectrum of open, coded, or „new” racism versus blatant racism (Breazu & Machin, 2022) and counter-acting forms against online hate speech (Smith, 2023).

Employing a focused mixed-methods design, the research data is structured around two primary dimensions:

7 HUNGARIAN NATIONAL REPORT AND RECOMMENDATIONS 2021, <https://romaversitas.hu/gyuloletbeszedellen/>

a. To analyze the discursive repertoires employed against Roma individuals and communities, identifying the dominant forms in which anti-gypsyist online hate speech is disseminated. Concurrently, this dimension will pinpoint frequently recurring themes and subthemes, characterize various forms of online hate speech, assess levels of aggressiveness and forms of toxicity (particularly racially loaded toxic discourse), and ultimately, exploring the model of web-based racism.

b. To explore participants' experiences regarding the efficacy of reporting mechanisms, including new media platform codes of conduct, national institutional rules, and relevant EU legislation adopted and implemented in the six selected countries.

2.2. Participants:

The monitors, aged 20-35, represented diverse levels of formal education (e.g., students, volunteers, employed individuals) and were selected based on a targeted call and motivation interviews, considering ethnic criteria, geographical diversity, English proficiency, prior experience in discrimination-related projects, and access to time and internet resources. Considering all these criteria Romaversitas selected five young Roma monitors. Two women and three men. Three of them holding BA degree in social work and international relations, one is studying at law school and one has Ph.D at educatio science. All of them was previously part of one or more Romaversitas programmes which meant they had experience on antigypsism projects.

2.3. Tools and measures

Data collection and analysis was conducted using a mixed content analysis methodology, employing a purpose-built monitoring tool (MT) for online application (Google Form).

The research developed a methodology to collect and categorize a wide range of data on antigypsyist online hate speech and it was discussed and refined with the national researchers in online meetings to assure a common understanding and its scientific validity.

The MT (see Annex 1) comprises six key components: (1) general information regarding authors and recipients; (2) a general topic assessment and reach; (3) categorization of content types (visual, text) and their reach; (4) measurement of the intensity and frequency of observed phenomena, and their potential impact; (5) analysis of counter-speech responses and actions; and (6) personal observations, accounting for national specificities regarding the functions and extent of reporting mechanisms.

a. General Information: This section gathered data pertaining to the monitored social media platform, the identification of the specific internet page, the type of account (personal, public, group, or institutional/private company), general information about the title of the article or media post, details concerning its influence and popularity (e.g., numbers of likes, shares), and a direct link.

b. General Topic Assessment: This component assessed the topic (ranging from crimes purportedly committed by Roma to cultural and sports events), content type (e.g., news reports, advertising), and style (e.g., scientific, emotional, formal, fictional), alongside an evaluation of the content's perceived impact on a three-point scale (negative, neutral, positive).

c. Categorization of Content Types: This section addressed the relevance of any relationship between the content and its visual elements, categorized non-textual content types (images, videos, etc.), and evaluated the emotional impact of the non-textual content using a three-point scale (very little, somewhat, very much). Special attention was paid to non-textual forms such as memes, multimedia materials (short videos), and animations.

d. Intensity and Frequency of Content: Adapted from Siapera, Moreo, and Jiang (2018), this section described the forms of racist online hate speech across two dimensions scaled on three levels of perceived intensity (1 - no hate, 2 - weak, 3 - extremely strong): (1) Open discourse: crude epithets, racist slurs, grossly offensive, and dehumanizing utterances; and (2) Coded discourse: employing slang, circumlocutions, irony, ambiguity, pseudo-scientific references, resurgent white supremacist ideologies, „whataboutery,” narratives of elsewhere, bogus statistics, and metonymies. A third dimension, adapted from Lentin (2016), measured the level of aggressivity/toxicity of hate speech using a „frozen vs. motile” scale, also adapted to the same three-level intensity assessment. Racially loaded language was considered toxic when it conveyed messages that reinforced stereotypes, entrenched polarization, spread myths and disinformation, justified exclusion, stigmatization, and inferiorization, or reinforced exclusivist notions of national belonging and identity. Main forms of racially loaded toxic discourse included crude forms of racism (insults, slurs, criminal comparisons, direct denigration, racial stereotypes related to color, smell, and physical features) and notions of „nation” (ad hominem attacks, right-wing discourse, race, ethnicity, religion, and culture), alongside accusations of being „over sensitive” or „playing the race card,” being racist against white people, coming from cultures of victimhood, spreading reverse fake news, and employing bogus statistics or research published with dubious credentials, or simply „reciting truths”.

e. Counter-Acting: Adapted from Smith (2023), this section catalogued on a three-point scale (yes, no, not sure) the types of counter-speech actions undertaken by the monitor in response to the perceived content. It encompassed a wide range of possibilities, from sincere engagement and ignoring the comment to reciprocating in kind. Different actions were listed to observe

the counter-speech responses and actions, including commenting on the post and reporting to the platform or any legal institution.

f. Personal Observation (Narrative): This section described in detail the study cases selected by the monitor, outlining the main idea of the content and how it exemplified anti-gypsyist online hate speech, including how this content affected Roma people (correlated with possible national specificities). The monitor explained the reporting process, the actions taken, and provided supporting details (including comments, engagement metrics) and screenshots.

The MT has been filled in using an online form (Google Form) and all the cases were supervised and discussed by each of the national team members. Some of the cases were discussed together with the research coordinators from all the partner' countries in case clarification were needed

All data were collected via a collaborative Google Form to ensure data transparency for team members and to enable verification of uploaded content (via links and screenshots).

This report presents a content analysis of 481 cases collected over a 12-month period (October 2024 - September 2025) using a standardized monitoring tool (MT) instrument. The MT, tailored for social media monitoring activities, incorporated specific codes and descriptions informed by the research questions and project objectives. The content analysis, performed on the basis of the MT results, informs the findings presented in this report.

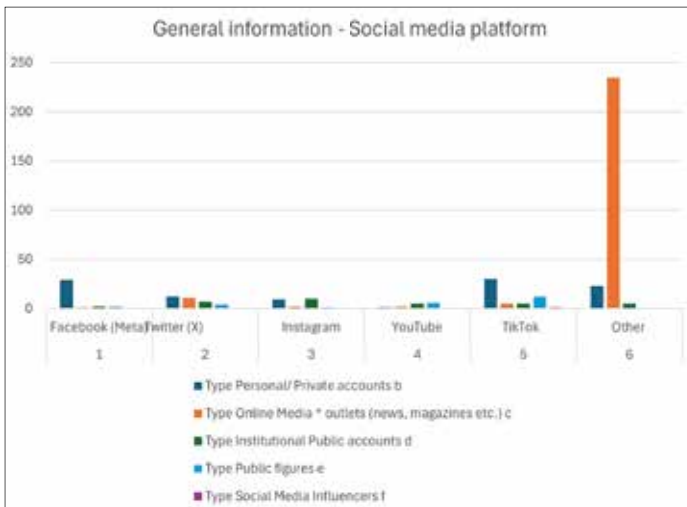
To ensure consistency and establish a common framework, we completed and transmitted a Synthesis Report to the research coordinator at agreed-upon intervals:

- 1st period: Oct-Dec 2024 - Report due January 2025
- 2nd period: Jan-Mar 2025 - Report due April 2025
- 3rd period: Apr-Jun 2025 - Report due July 2025
- 4th period: Jul-Sept 2025 - Report due October 2025

In accordance with project activities and objectives, the MT was applied subjectively, granting monitors the autonomy to select content and online platforms (including their personal accounts on Facebook, X, Instagram, TikTok, or YouTube) based on information appearing in their news feeds.

3. Results

3.1. Quantitative component



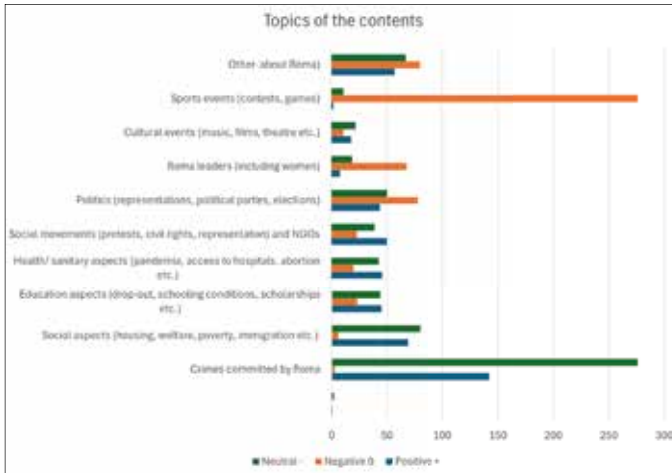
In the initial phase of the monitoring methodology, the primary focus was on collecting pertinent general information to contextualize the analysis, including the identification of the

main social media platforms utilized and the typology of user interactions. It is important to emphasize that the selection of specific platforms and websites for monitoring was a subjective decision made by five operators and their respective teams. Consequently, the data obtained should not be regarded as representative but rather as preliminary indicators and preferences among young operators regarding online social platforms (private or public accounts, social media influencers and public figures) and information sources (local and national online press). The operators' attention was also directed toward social networks frequently used by them, which explains the higher proportion of personal accounts (individual actors) compared to official news pages or public institutional accounts; the latter were analyzed only when taken over and disseminated online.

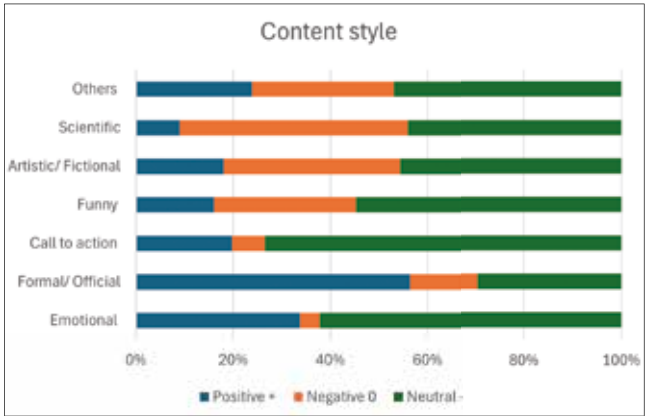
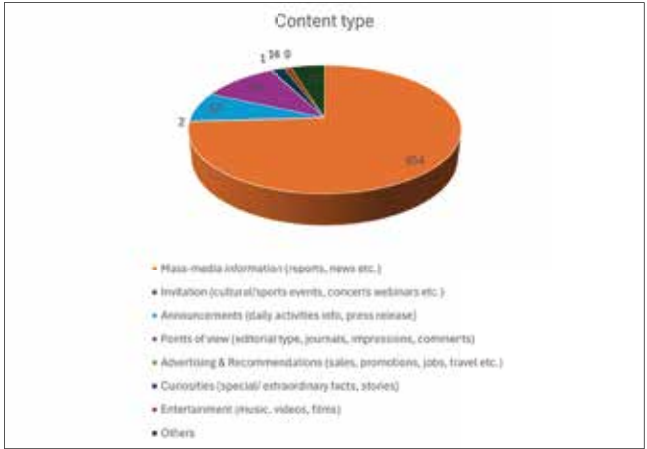
As illustrated in Figure 1 (General information – social media platform), the predominant categories of users involved in the research—identified as young Roma operators—were private accounts on Facebook, TikTok, X (former Twitter). Regarding frequency of use, Facebook remains the most widely utilized social media platform identified by the monitoring team, primarily due to its private and closed group features. The second most frequently used platform is TikTok, where hate speech propagation occurs predominantly through private accounts, followed by X (formerly Twitter), and YouTube. Concerning online news dissemination, Facebook/Meta serves as the principal source of information and is the primary channel through which public figures distribute their content, but in Hungary there were data collected from “Other” platforms known as far-right websites.

During the content analysis of the collected and monitored materials, we defined the following categories. Within each category, the monitors had to decide what overall effect the post/

article/video had: positive, neutral, or negative with regard to the Roma community. The chart below presents the results for the 481 cases from Hungary.

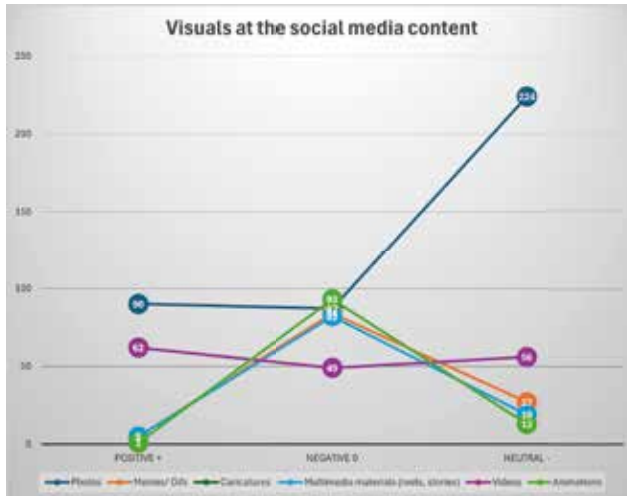


After reading the materials, it became apparent that the young participants mainly selected content appearing in news reports and features — materials that reach or can reach wide audiences. Regarding the style of the content, the young people also characterized what they read and saw on a positive–neutral–negative scale. The summary shows that the “emotional” style was the most typical with neutral content, followed again by the “emotional” style with positive content. It is important to highlight, however, that the highest number of negative mentions occurred in the case of content with a scientific theme and style.

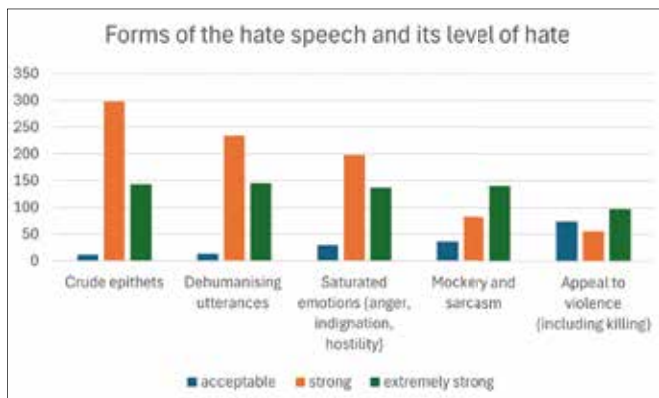


Most of the collected content included visual elements. These were also analyzed by the monitors according to type and positioned on the positive–neutral–negative scale.

We most frequently encountered images, the majority of which had neutral content, and most of the positive visual elements were also presented in photographs. Our data collectors encountered negative content primarily in animations.

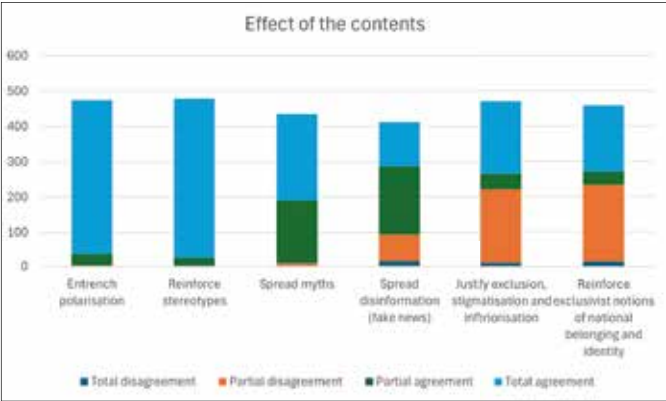
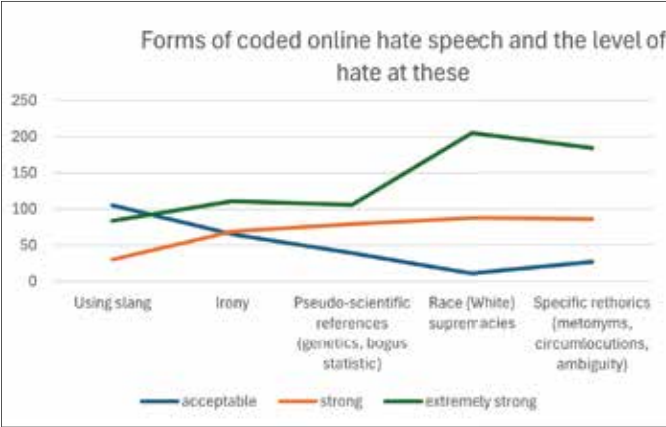


The collected materials were then classified by the Roma youth into different categories of hate speech, after which they determined, along three levels, how intense the hatred was in relation to the given content.



In order to gain a deeper understanding, the categorization of coded hate speech followed. The individual cases were also classified by the monitors along how strong they felt the hatred

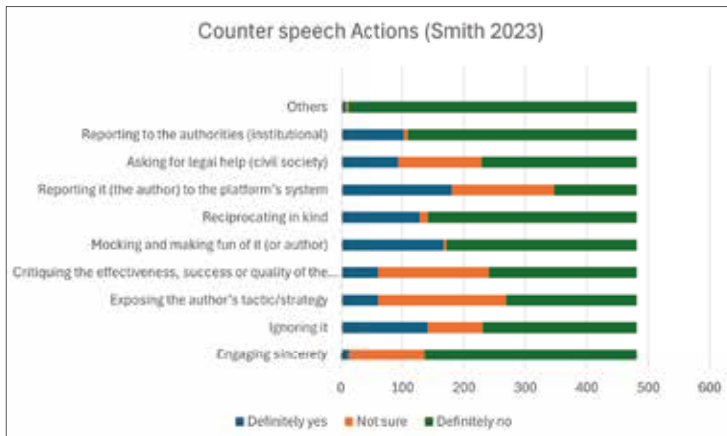
was in them on a three-point scale (acceptable-strong-extremely strong). The figure below shows that the explanations of racial and white supremacy predominate with the strongest hatred.



With regard to the impact of the content, the Roma young people who monitored agreed in most cases that the content they read strengthens polarization and the spread of stereotypes and myths in society, and that they justify the exclusion of the Roma community from the society as a whole to a very large extent, i.e. they question their belonging to the nation.

After a multifaceted analysis and assessment of the content, the monitoring young people had to answer the question on the online questionnaire about what and how they would react to that given content.

It is clear from their answers that in most cases we would not react and they would not take any action. During the meetings and workshops with them, we talked a lot about the reason for this. One of their responses was that they reported these contents to the given platform in many cases, but they did not receive a meaningful reaction to this. The other phenomenon that has been brought to our attention is that actors and content producers are becoming more and more sophisticated, which expressions and formulations are not able to filter by platforms. Users also use various hate speech content by circumventing the filters of the platforms, hiding them between code symbols. Thus, they have little chance of achieving real change and imposing real sanctions on hateful users.



3.2. *Qualitative component (The Narrative)*

During the monitoring period, we reviewed all cases every three months and collected the most typical ones. In this report, we would like to present one that characterizes the presence of hate speech in today's Hungary.

Toroczkai's statement at this video⁸ says: "there are certain subcultures where having children is not a problem, while the population of Hungarians is decreasing and less children are borning" - is a form of coded hate speech. Although it doesn't explicitly name the Roma community, in the Hungarian context it clearly alludes to them. Such statements express prejudice in a more indirect, euphemistic way, making them harder to legally or socially challenge.

This rhetoric implies that the Roma community irresponsibly has children and poses a problem for society, reinforcing stereotypes, stigmatizing the group, and fueling social division. Its "sophisticated" wording is dangerous precisely because it spreads hate in a hidden, less obvious form.

Even if it can be felt with common sense that a certain community is being blamed, hateful speech cannot be reported or punished in this form. This indicates persistent unmoderated antigypsyism, especially from political figures.

8 <https://www.tiktok.com/@toroczkai.laszlo/video/7522514008994352406>

4. Conclusion and recommendations

4.1. Overarching Cross-Cutting Observations:

- **Pervasive Normalization and Impunity:** Online anti-gypsyism is often normalized, and social media platforms frequently fail to enforce content moderation guidelines.
- **Rise of AI-Generated Hate Speech:** AI is increasingly leveraged to create stereotypical and derogatory content, often disguised as „parody,” further normalizing racism.
- **Political Instrumentalization:** Political actors and public figures actively contribute to generating or amplifying anti-gypsyist discourse for electoral gain.

These findings underscore the complex and multi-faceted nature of online anti-gypsyism, revealing significant challenges in content moderation, a pronounced negative framing of Roma-related issues, and a critical need for more effective counter-action and reporting mechanisms.

Tackling these challenges of online anti-gypsyism will require not only robust legal frameworks but also sustained collaboration between regulators, platforms, civil society, and Roma affected communities.

– **Strengthen Enforcement of Digital Regulations:** The study highlights a „pervasive normalization and impunity” due to platforms failing to enforce their own guidelines, and reporting mechanisms often yielding no action.

Action: Mandate and actively monitor the rigorous enforcement of the Digital Services Act (DSA) and AI Act

provisions specifically targeting anti-Roma hate speech and cyberhate by Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs).

– **Enhance Transparency and Accountability:** Platforms' current accountability appears insufficient, and there is a „significant dearth of research and concerted efforts” to combat anti-gypsyism. Greater transparency is crucial for public scrutiny and effective policy evaluation.

Action: Require social media platforms to publish detailed, granular transparency reports on the detection, moderation, and removal of anti-Roma hate speech, including country-specific data and breakdowns for both explicit and implicit forms.

– **Combat Political Instrumentalization:** Political instrumentalization of anti-gypsyist discourse for electoral gain was observed across Hungary, further legitimizing hatred.

Action: Develop ethical guidelines and codes of conduct for political parties and public figures regarding discourse concerning minority groups, with clear sanctions for amplifying anti-Roma narratives.

– **Support Monitoring and Reporting Mechanisms:** While formal reporting to platforms is the most favored active response, its effectiveness is low. Improving and supporting these mechanisms is vital.

Action: Invest in and standardize national reporting mechanisms for anti-gypsyism, ensuring they are accessible, effective, and provide timely feedback to users. Support civil society organizations, particularly Roma organizations, in monitoring and documenting online hate speech.

4.2. *Social Media Platforms*

– **Improve Content Moderation Effectiveness:** The study demonstrates that platforms frequently fail to act on reported anti-gypsyism, and that complex rhetorical figures are often not identified as overtly aggressive despite their harmful intent.

Action: Consistently enforce content moderation policies to specifically address the nuances of anti-gypsyism, including subtle, implicit forms like „irony,” „mockery and sarcasm,” and „pseudo-scientific references” that the study found to be highly aggressive.

– **Increase Human Oversight and Cultural Competence:** The subjective and context-dependent nature of hate speech identification, and the high prevalence of implicit forms, necessitates human intervention and oversight beyond AI-based moderation.

Action: Increase investment in diverse human moderators with cultural and linguistic expertise relevant to the European contexts studied, to accurately identify and address nuanced anti-Roma hate speech, especially in local languages.

– **Collaborate with Roma Organizations:** Roma organizations possess critical insights into the lived experiences and specific manifestations of anti-gypsyism.

Action: Establish formal partnerships with Roma organizations and experts to inform policy development, improve content moderation training, and develop counter-narratives and educational initiatives.

4.3. Roma Organizations

– **Empower Communities with Digital Literacy:** There is a high prevalence of „not sure” responses to counter-actions, and a potential „desensitization” to rhetoric like „appeal to violence,” suggesting a need for enhanced critical literacy.

Action: Develop and deliver training programs for Roma individuals, especially youth, on identifying, documenting, and reporting all forms of online anti-gypsyism, both explicit and implicit, to platforms and relevant authorities.

– **Sustain and Expand Monitoring Efforts:** The study highlights the value of data in informing evidence-based policies and actions, and a „dearth of research” in this specific area.

Action: Continue and expand systematic monitoring and data collection on online anti-gypsyism, ensuring that methodology captures nuanced forms and country-specific contexts. Collaborate with academic institutions for rigorous analysis.

– **Intensify Advocacy and Lobbying:** The evidence of pervasive normalization, ineffective moderation, and political instrumentalization provides a strong basis for advocacy.

Action: Utilize the study’s findings to advocate strongly with policymakers and media giants for robust legislative enforcement, increased platform accountability, and improved content moderation specific to anti-Roma hate speech.

– **Develop Counter-Narratives and Positive Content:** Negative sentiment heavily outweighs positive mentions in the discourse, and promoting positive content can help reshape public perception.

Action: Actively create and promote positive, culturally accurate, and empowering online content about Roma culture, history, and achievements to counteract negative stereotypes and foster a more inclusive online space.

5. References

- ERGO Network (2022). PECAO (Peer Education to Counter Antigypsyist Online Hate Speech) Synthesis Report on Antigypsyist Online Hate Speech, available at: <https://ergonetWORK.org/wp-content/uploads/2022/04/ERGO-Synthesysreportweb-FINAL.pdf>;
- Bonilla-Silva, Eduardo (2006). Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States. *Science and Society* 70 (3):431-434.
- Lentin, Alana (2016) 'Racism in Public or Public Racism: doing anti-racism in post-racial times', *Ethnic and Racial Studies*, 39(1): 33-48.
- Martinez-Fuentes, S., Umaña-Taylor, A. J., Jager, J., Seaton, E. K., & Sladek, M. R. (2020). An Examination of Ethnic-Racial Identity and U.S. American Identity among Black, Latino, and White Adolescents. *Identity*, 20(3), 208–223. <https://doi.org/10.1080/15283488.2020.1784177>
- Breazu, Petre & Machin, David (2022). Racism is not just hate speech: Ethnonationalist victimhood in YouTube comments about Roma during COVID-19, *Language in Society*, doi:10.1017=S0047404522000070.
- Smith, Philippa (2023). Online counterspeech and the targeting of digital discourses of racism in New Zealand in *Social*

Media and Society: Integrating the digital with the social in digital discourse Edited by Majid KhosraviNik, pp. 146-166, available at: <https://benjamins.com/catalog/dapsac.100.08smi>

Siapera, Eugenia, Moreo Elena and Jiang Zhou (2018). Hate Tracking and Monitoring Racist Speech Online (<https://www.ihrec.ie/app/uploads/2018/11/HateTrack-Tracking-and-Monitoring-Racist-Hate-Speech-Online.pdf>)

7. Appendices

Annex 1.

MONITORING TOOL FOR CONTENT ANALYSIS OF ONLINE HATE SPEECH

I. General information

Social Media platform	Type				
	Personal/Private accounts	Online Media * outlets (news, magazines etc.)	Institutional Public accounts**	Public figures***	Social Media Influencers
Facebook (Meta)					
Twitter (X)					
Instagram					
YouTube					
TikTok					
Other (please, specify)					

* Official page

** Governmental institution (Police, Ministry of Education, Governmental Agencies etc.), municipalities, companies, nongovernmental organisations (including church, schools, cultural)

*** Politicians, reputed professors/scientists, artists, musicians, journalists

Name (of the site/account/publication/channel	
Date of publication:	
Link or print screen	
Private or public	
Author (if the case)	

Reach (numbers)

Likes (appreciations)	Dislikes (including fury or sadness)	Shares	Comments	Visualisations (Views)

II. General Topic

Main topic	Positive +	Negative -	Neutral 0
Crimes committed by Roma			
Social aspects (housing, welfare, poverty, immigration etc.)			
Education aspects (drop-out, schooling conditions, scholarships etc.)			
Health/ sanitary aspects (pandemia, access to hospitals, abortion etc.)			
Social movements (protests, civil rights, representation) and NGOs			
Politics (representations, political parties, elections)			
Roma leaders (including women)			
Cultural events (music, films, theatre etc.)			
Sports events (contests, games)			
Other (about Roma)			

Content Type	Positive +	Negative -	Neutral 0
Mass-media information (reports, news etc.)			
Invitation (cultural/sports events, concerts webinars etc.)			
Announcements (daily activities info, press release)			
Points of view (editorial type, journals, impressions, comments)			
Advertising & Recommendations (sales, promotions, jobs, travel etc.)			
Curiosities (special/ extraordinary facts, stories)			
Entertainment (music, videos, films)			
Others			

Content Style	Positive +	Negative -	Neutral 0
Emotional			
Formal/ Official			
Call to action			
Funny			
Artistic/ Fictional			
Scientific			
Others			

Does it include visuals?	
Yes	No

Please, indicate the non-textual forms used (and their number, if more than one)	
Photos	
Memes/ Gifs	
Caricatures	
Multimedia materials (reels, stories)	
Videos	
Animations	
Others	

On a 3 level scale how do you appreciate the general frame of the visuals:

Positive +	Negative -	Neutral 0

III. Online Hate Speech – Intensity and levels of hate

Forms of explicit online hate speech	Levels of hate		
	1 acceptable	2 strong	3 extremely strong
Crude epithets			
Racist slurs			
Dehumanising utterances			
Saturated emotions (anger, indignation, hostility)			
Mockery and sarcasm			
Appeal to violence (including killing)			
Other			

Forms of coded online hate speech	Levels of hate		
	1 acceptable	2 strong	3 extremely strong
Using slang			
Irony			
Ambiguity			
Pseudo-scientific references (genetics, bogus statistic)			
Race (White) supremacies			
Specific rethorics (metonyms, circumlo- cutions, ambiguity)			
Other			

Does the content convey the next messages? In what extent?

Frozen vs. motile online hate speech (Lentin 2016)	Levels of (hate) toxicity			
	Total disagreement	Partial disagreement	Partial agreement	Total agreement
Entrench polarisation				
Reinforce stereotypes				
Spread myths				
Spread disinformation (fake news)				
Justify exclusion, stigmatisation and inferiorisation				
Reinforce exclusivist notions of national belonging and identity				
Other				

IV. Types of Counter speech

1. Does the post violate the common-sense rules? Yes/No
2. What type of actions would you do?

Counter speech Actions (Smith 2018)	Definitely yes	Not sure	Definitely no
Engaging sincerely			
Ignoring it			
Exposing the author's tactic/strategy			
Critiquing the effectiveness, success or quality of the author's arguments and behaviours			
Mocking and making fun of it (or author)			
Reciprocating in kind			
Reporting it (the author) to the platform's system			
Asking for legal help (civil society)			
Reporting to the authorities (institutional)			
Others			

In case you decide to counter act, please give evidence of your activity:

Type of evidence	Activity (short description: what, whom, why, date)	Examples (link or print-screen)
Text (comment to the post)		
Text (reporting to the platform)		
Text (reporting institutionally)		
Visual (Photo/ print screen)		

V. Reporting (personal observations):

Reporting - your personal observation. Please, write in a short paragraph some answers to the next guiding questions:

- What were your recommendations? to whom (platform, to the author, to the institution)
- What and when was the answer received? (mention if you not receive any)
- How was the decision made? (apologies were formulated, the post was erased; the author was blocked, the author was held accountable etc.)
- Are you satisfied or not with the answer/action? Explain why.
- Provide proofs: links, photos, print-screens etc.

In case, there are (just) **comments** of the post/ article/ image, please provide us some details:

Comments analysis	Frequency	Examples (if the case)
Number of comments	<2; 3-5; >6	
Number of authors of the comments	<2; 3-5; >6	
Type of language (aggressive vs. peaceful) – scale 5 levels	Aggressive	
Fake accounts (hidden identities)	1 – yes; 0 – no	
Intensity of the comments	no hate -0, mockery- 1, aggressive – 2, threats to killing - 3	
Persistence (insists in commenting with one or more persons)	1 – yes; 0 - no	
Types of popularity and influence of the comment (likes/ favourite/ shares) – 3 level scale	0 – nothing; 1 – weak (<2) 2 – Strong (>3)	

Annex 2.

SUMMARY OF THE SYNTHESIS REPORT OF THE MONITORING TOOL FOR CONTENT ANALYSIS OF ONLINE HATE SPEECH

Month	
Period	
Number of Monitoring Tools analysed	
Organization/ Country	

I. General information

Social Media platform	Type				
	Personal/ Private accounts b.	Online Media * outlets (news, magazines etc.) c.	Institutional Public accounts d.	Public figures e.	Social Media Influencers f.
Facebook (Meta) 1					
Twitter (X) 2					
Instagram 3					
YouTube 4					
TikTok 5					
Other (please, specify) 6					

II. General Topic

Main topic	
Crimes committed by Roma	a
Social aspects (housing, welfare, poverty, immigration etc.)	b
Education aspects (drop-out, schooling conditions, scholarships etc.)	c
Health/ sanitary aspects (pandemia, access to hospitals, abortion etc.)	d
Social movements (protests, civil rights, representation) and NGOs	e
Politics (representations, political parties, elections)	f
Roma leaders (including women)	g
Cultural events (music, films, theatre etc.)	h
Sports events (contests, games)	I
Other (about Roma)	j

On a scale how do you appreciate the general frame of the topic:

Positive +	Negative -	Neutral 0
1	2	0

Content Type		
Mass-media information (reports, news etc.)	a	
Invitation (cultural/sports events, concerts webinars etc.)	b	
Announcements (daily activities info, press release)	c	
Points of view (editorial type, journals, impressions, comments)	d.	
Advertising & Recommendations (sales, promotions, jobs, travel etc.)	e.	
Curiosities (special/ extraordinary facts, stories)	f.	
Entertainment (music, videos, films)	g.	
Others	h.	

Content Style		Positive +	Negative -	Neutral 0
		1	2	0
Emotional	a.			
Formal/ Official	b.			
Call to action	c.			
Funny	d.			
Artistic/ Fictional	e.			
Scientific	f.			
Others	g.			

Does it include visuals?	
Yes	No
1	0

Please, indicate the non-textual forms used (and their number, if more than one)	
Photos	a.
Memes/ Gifs	b.
Caricatures	c.
Multimedia materials (reels, stories)	d.
Videos	e.
Animations	f.
Others	g.

On a 3 level scale how do you appreciate the general frame of the visuals:

Positive +	Negative -	Neutral 0
1	2	0

III. Online Hate Speech – Intensity and levels of hate

Forms of explicit online hate speech Type A		Levels of hate		
		1 acceptable	2 strong	3 extremely strong
Crude epithets	a.			
Racist slurs	b.			
Dehumanising utterances	c.			
Saturated emotions (anger, indignation, hostility)	d.			
Mockery and sarcasm	e.			
Appeal to violence (including killing)	f.			
Other	g.			

Forms of explicit online hate speech Type B		Levels of hate		
		1 acceptable	2 strong	3 extremely strong
Using slang	a.			
Irony	b.			
Ambiguity	c.			
Pseudo-scientific references (genetics, bogus statistic)	d.			
Race (White) supremacies	e.			
Specific rethorics (metonyms, circumlocutions, ambiguity)	f.			
Other	g.			

Does the content convey the next messages? In what extent?

Frozen vs. motile online hate speech (Lentin 2016)		Levels of (hate) toxicity			
		Total disagreement 1	Partial disagreement 2	Partial agreement 3	Total agreement 4
Entrench polarisation	a.				
Reinforce stereotypes	b.				
Spread myths	c.				
Spread disinformation (fake news)	d.				
Justify exclusion, stigmatisation and inferiorisation	e.				
Reinforce exclusivist notions of national belonging and identity	f.				
Other	g.				

IV. Types of Counter speech

3. Does the post violate the common-sense rules? Yes (1) No (2)

4. What type of actions would you do?

Counter speech Actions (Smith 2018)		Definitely yes	Not sure	Definitely no
		1	2	3
Engaging sincerely	a.			
Ignoring it	b.			
Exposing the author's tactic/strategy	c.			
Critiquing the effectiveness, success or quality of the author's arguments and behaviours	d.			
Mocking and making fun of it (or author)	e.			
Reciprocating in kind	f.			
Reporting it (the author) to the platform's system	g.			
Asking for legal help (civil society)	h.			
Reporting to the authorities (institutional)	i.			
Others	j.			

In case you decide to counter act, please give evidence of your activity:

Type of evidence		Activity (short description: what, whom, why, date)	Examples (link or print-screen)
Text (comment to the post)	a.		
Text (reporting to the platform)	b.		
Text (reporting institutionally)	c.		
Visual (Photo/ print screen)	d.		

This part will be selected from the database and addressed by the national team.

V. Narratives Reporting:

Choose 1-3 cases you find representative (or exceptional) for the period you monitored (e.g. electoral campaigns, accidents, cultural event etc)

Or/and

If you identified a pattern of narrative (e.g. Roma children should or are forced to stay in segregated schools and neighbourhoods) that is disseminated for more than 5 days and on several channels (media and online)

1. Please, write in a paragraph some answers to the next guiding questions:

Explain and provide details: what is the topic, who is the main sender and who are the secondary senders/multipliers of the messages, what is channel (and the other channels – international, for example), what are the reactions. Make a short paragraph (half page – 500 words) on the case.

Name (of the site/account/publication/channel)	
Date of publication:	
Private or public	
Author (if the case)	

Multiply it accordingly to your case.

Reach (you may provide numbers)

Likes (appreciations)	Dislikes (including fury or sadness)	Shares	Comments	Visualisations (Views)

If it is for one study case

2. In some cases, there are (just) **comments** of the post/ article/ image, please provide us some details:

Comments analysis	Frequency/numbers	Examples (if the case)
Number of comments		
Number of authors of the comments		
Type of language (aggressive vs. peaceful)		
Fake accounts (hidden identities)		
Intensity of the comments (mockery, aggressive, threats to killing etc.)		
Persistence (insists in commenting with one or more persons)		<i>It came after with formal apologies</i>
Types of popularity and influence of the comment (likes/ favourite/ shares) – strong/weak		

Write your observations here:

3. Reporting - your personal observation.

Please, write in a paragraph some answers to the next guiding questions:

- What were your recommendations? to whom (platform, to the author, to the institution)
- What and when was the answer received? (mention if you not receive any)
- How was the decision made? (apologies were formulated, the post was erased; the author was blocked, the author was held accountable etc.)
- Are you satisfied or not with the answer/action? Explain why.

Other background information if you consider relevant (you may insert photos, print screens, citations, references, links)