

Table of contents

executive summary	3
1. introduction	4
1.1. background and context	4
1.2. significance of the study	6
2. methodology	8
2.1. purpose and design	8
2.2. participants	10
2.3. tools and methods	11
3. results	10
3.1. quantitative component	13
3.1.1 Item: General information – authors/ recipients	13
3.1.2 Item: General topic – reach (QII a)	15
3.1.3. Item: Content (visual, texts) – reach	25
3.2 intensity and frequency of hate – impact	30
3.2.1. Explicit online hate speech (A)	30
3.2.2. Coded online hate speech (B)	34
3.2.3. impact of online hate speech	37
3.3 counteracting	41
3.4. qualitative component (the narrative)	47
4. limits of the study	54
5. conclusions	55
6. recommendations	56
6.1 policymakers	56
6.2 social media platforms	57
6.3 roma organisations	58
references	59

author: Ileana Rotaru
date of publication: 2026
designer: Inga Sabanova

printer: MEDIAPROCESS, Chaussée de Louvain
775 - 1140 Brussels



This publication was funded by CERV-2023-CHAR-LITI-SPEECH. The European Commission's support for the production does not constitute an endorsement of the contents, which reflect the views of the authors only, and the European Commission cannot be held responsible for any use which may be made of the information contained there.

executive summary

This multi-country study, "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 monitors who collected quantitative data from 3147 cases using a standardised monitoring tool (MT) and conducted qualitative analysis of 54 selected case studies. The methodology included sentiment analysis across various topics and content styles, as well as aggressivity scales for both explicit and implicit forms of hate speech.

Key quantitative findings revealed a significant prevalence of perceived negative sentiment in media content pertaining to Roma, constituting 40.58% of 15,080 topic-related instances, compared to positive framing, which was notably low at 14.49%. Country-specific analyses showed distinct patterns, with Hungary, Germany, and Romania leaning towards negative reporting (43.95%, 63.26%, and 50%, respectively).

**“ONLY 2 OUT OF 40
REPORTED CASES
RESULTED IN ACTION”**

In terms of hate speech aggressivity, "weak" (moderate) forms were most frequent for explicit hate speech, yet "Appeal to violence" surprisingly often fell into the "low aggressivity" category, suggesting potential desensitisation among monitors. Implicit forms, such as "Irony" and "Race (White) supremacies," were prominent in higher aggressivity levels. Social media platforms, particularly Facebook and TikTok, were identified as primary vectors for the dissemination of anti-gypsyist content, with visual elements frequently accompanying negative narratives.

Qualitative insights further delineated the manifestations of online anti-gypsyism, highlighting "Group Differentiation" as the most prevalent theme, alongside "Insults, Disgust, Violence, and Punitiveness" and "Deviance and Criminality." The study also identified the concerning rise of AI-generated hate speech and the political instrumentalisation of anti-Roma discourse. A critical finding regarding counteractions was the overwhelming lack of effective moderation: only 2 out of 40 reported cases resulted in a positive outcome, despite a strong reliance on formal platform reporting mechanisms. Overall, there was low active engagement from individuals in counter-speech, with high levels of uncertainty and refusal to act.

In conclusion, the study underscores that online anti-gypsyism is a pervasive, nuanced, and frequently unaddressed problem within the European digital space. The widespread ineffectiveness of current content moderation and reporting mechanisms fosters a climate of impunity. These findings carry significant implications for policymakers, urging them to strengthen enforcement of digital regulations, enhance platform accountability for hate speech, and address the specific challenge of AI-generated content. For social media platforms, the study calls for improved, culturally competent content moderation and transparent reporting. Roma organizations are encouraged to empower communities with digital literacy, sustain monitoring efforts, and intensify advocacy for systemic change.

1. introduction

1.1. background and context

This report is part of a multi-country study within the TAAO project, and it directly addresses the European Union's priority of **safeguarding fundamental values and rights** by focusing on the critical issue of online anti-Roma hate speech and hate crime.



Hate speech, as a manifestation of anti-gypsyism, merits particular attention due to its multiplier effect: it shapes public opinion, exacerbates social tensions, and facilitates discrimination and hate crimes. Online media plays a crucial role in disseminating and inciting hate speech by reinforcing stereotypes, employing offensive language, and denying or trivialising antigypsyism. **Through social media, hate speech reaches millions, enabling perpetrators to anonymously incite hatred and violence, often triggering offline hate crimes.**

As emphasised in the European Commission Communication "A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime" (2021), "Hate speech and hate crime affect not only the individual victims and their communities, causing them suffering and limiting their fundamental rights and freedoms, but also society at large. Hate undermines the very foundations of our society. It weakens mutual understanding and respect for diversity on which pluralistic and democratic societies are built."

The 10-12 million Roma in Europe constitute the most discriminated ethnic minority in the region. Recent years have seen an escalation of anti-Roma hate speech, notably amplified during the Covid-19 pandemic (e.g., the "Roma as super spreaders of the virus" narrative, as per the ERGO Network Report, 2020) and the war in Ukraine (e.g., "Roma refugees as taking advantage of social benefits"). Addressing online hate speech against Roma is thus essential for upholding their right to dignity and non-discrimination, concurrently safeguarding the right to freedom of expression for all. While antigypsyist hate speech, as a form of racist and xenophobic expression, has been categorised as illegal under an EU Council Framework Decision of 2018, there remains a significant dearth of research and concerted efforts to combat this specific manifestation of online racism.

Building upon previous research conducted within the European area (ERGO Network Report PECAO, 2022) that has focused on addressing and countering anti-gypsyist online hate speech based on the lived experiences of young Roma people, their engagement with online media spaces, their personal social media accounts, and their digital footprints, the present research aims to achieve the following general objectives:

- To provide comprehensive data on the **extent, reach, impact, authors, recipients, and reactions related to anti-gypsyist hate speech online**, thereby informing evidence-based policies and actions designed to combat this phenomenon.

- To **increase the number of reported cases of antigypsyist hate speech** through the dissemination of information regarding available reporting mechanisms.

The present study has a very clear European dimension. It includes **six national organisations** from Germany (Amaro Drom), the Czech Republic (ROMEIA), Slovakia (Roma Advocacy and Research Centre - RAVS), Hungary (Romaversitas), Romania (Nevo Parudimos) and Bulgaria (Integro Association) and one European umbrella organisation (The European Roma Grassroots Organisations Network – ERGO).



1.2. significance of the study

The virtual public sphere is consequently saturated with **racist and discriminatory content**, often disguised as ironic "jokes." Antigypsyism, in particular, continues to be a contentious topic in public discourse, not only regarding its definition and regulation but also as a deeply ingrained, conscious attitude within the collective consciousness, whose manifestations are increasingly amplified by the digital environment.

Defined as a specific form of racism, antigypsyism is an ideology rooted in racial superiority, institutional dehumanisation, and is expressed through violence, hate speech, exploitation, and stigmatisation. It encompasses a broad spectrum of attitudes and practices, with implicit or dynamic manifestations that permeate traditional and new media.

The 7th evaluation of the EU Code of Conduct on countering illegal hate speech online (2022) indicated removal rates of 69.6% for content advocating murder or violence against specific groups, and 59.1% for defamatory content targeting identified groups. This evaluation further identified antigypsyism, xenophobia (including anti-migrant hatred), and sexual orientation as the most frequently reported grounds for hate speech, with antigypsyism constituting the largest proportion at 16.8%.

At the European level, antigypsyism, including anti-gypsyist hate speech, is widely acknowledged based on reports and studies, yet the results felt in everyday life by the Roma citizens are very little addressed and ignored.

A data collection exercise from the **"Peer Education to Counter Antigypsyist Hate Speech" (PECAO) project**, which served as a predecessor to this initiative, analysed 1795 cases across nine countries. Its primary finding was the pervasive nature of racially loaded toxic discourse within online public spaces. This discourse was widely disseminated by diverse categories of users and prosumers, representing varied cultural and educational backgrounds, who often demonstrated limited awareness or empathy towards the victimised "other" regarding their online conduct. Usually, **this type of behaviour is not taxed or reported by social media platforms' algorithms**, nor by individual reactions to it. They are seen and classified in a soft manner as "innocent jokes" (when it is about more visual content), or as tacit agreement with aggressive comments/content.

More than that, the report "Challenging Digital Antigypsyism" (2023) by the European Roma Rights Centre, collecting and analysing antigypsyist hate speech data from Albania, Serbia, Turkey and Ukraine, confirms these conclusions.

The central message from the ERRC's Challenging Digital Antigypsyism project was the importance of agency. "By forming digital activist communities, by taking action to monitor, record, and report examples of anti-Roma hate speech, and by holding political authorities and social media platforms publicly to account, these teams of volunteers have performed an exemplary civic duty in defending democracy".

The proliferation of **online antigypsyism and other forms of cyberhate poses a significant challenge to the protection of fundamental rights** and the integrity of public discourse within the European Union. In response, the EU has enacted landmark legislation, notably the Digital Services Act (DSA) and the Artificial Intelligence Act (AI Act), to enhance accountability among social media platforms and foster a safer online environment for EU citizens.

One of this project's objectives is to examine how these regulations address online antigypsyism and cyberhate, focusing on their mechanisms for holding platforms accountable. The Digital Services Act, operational since early 2024 for Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs), directly targets illegal and harmful content, which unequivocally includes antigypsyism and cyberhate. Its provisions establish a robust framework that compels platforms to take greater responsibility for the content they host.

Due Diligence and Notice-and-Action Mechanisms:

The DSA mandates that platforms implement accessible, effective mechanisms for users to report illegal content. Upon receiving such a notice, platforms are legally obliged to act "without undue delay," assess the content's legality under EU or national law, and remove it if it is found to be illegal. This directly empowers EU citizens, including Roma individuals, to flag antigypsyist content and holds platforms accountable for their responsiveness. The previous analysis of reported cases, revealing widespread platform inaction, underscores the critical need for this provision.

Risk Assessments and Mitigation for VLOPs/VLOSEs:

VLOPs/VLOSEs are mandated to conduct annual systemic risk assessments, including risks to fundamental rights (such as non-discrimination) and public discourse. Online antigypsyism, given its pervasive nature and documented impact, constitutes a systemic risk that these platforms must identify and actively mitigate. This necessitates changes to their content moderation policies, recommender systems, and algorithmic amplification strategies to prevent the spread of such hate speech.

Codes of Conduct: The DSA encourages the development of voluntary Codes of Conduct for specific types of illegal content, which could include dedicated provisions to combat antigypsyism and promote best practices across the industry.

Transparency and Explanations:

Platforms must provide clear, specific reasons for any content moderation decisions, including removal or non-removal. They are also required to publish regular transparency reports on content moderation activities, including the number of pieces of illegal content reported and the actions taken. For antigypsyism, this means platforms must articulate why content like dehumanising slurs or calls for violence is or is not deemed a violation, allowing for greater public scrutiny and accountability.

User Redress Mechanisms:

The DSA provides avenues for users to appeal content moderation decisions, both internally within the platform and externally through certified out-of-court dispute-resolution bodies. This provides EU citizens with robust avenues for recourse when platforms fail to adequately address antigypsyism.

Enforcement:

Digital Services Coordinators (DSCs) at the national level and the European Commission are endowed with significant enforcement powers, including the ability to levy substantial fines (up to 6% of global annual turnover) for non-compliance. This financial penalty creates a strong incentive for platforms to uphold their responsibilities toward EU citizens.

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: antigypsyism (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, 2018).

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

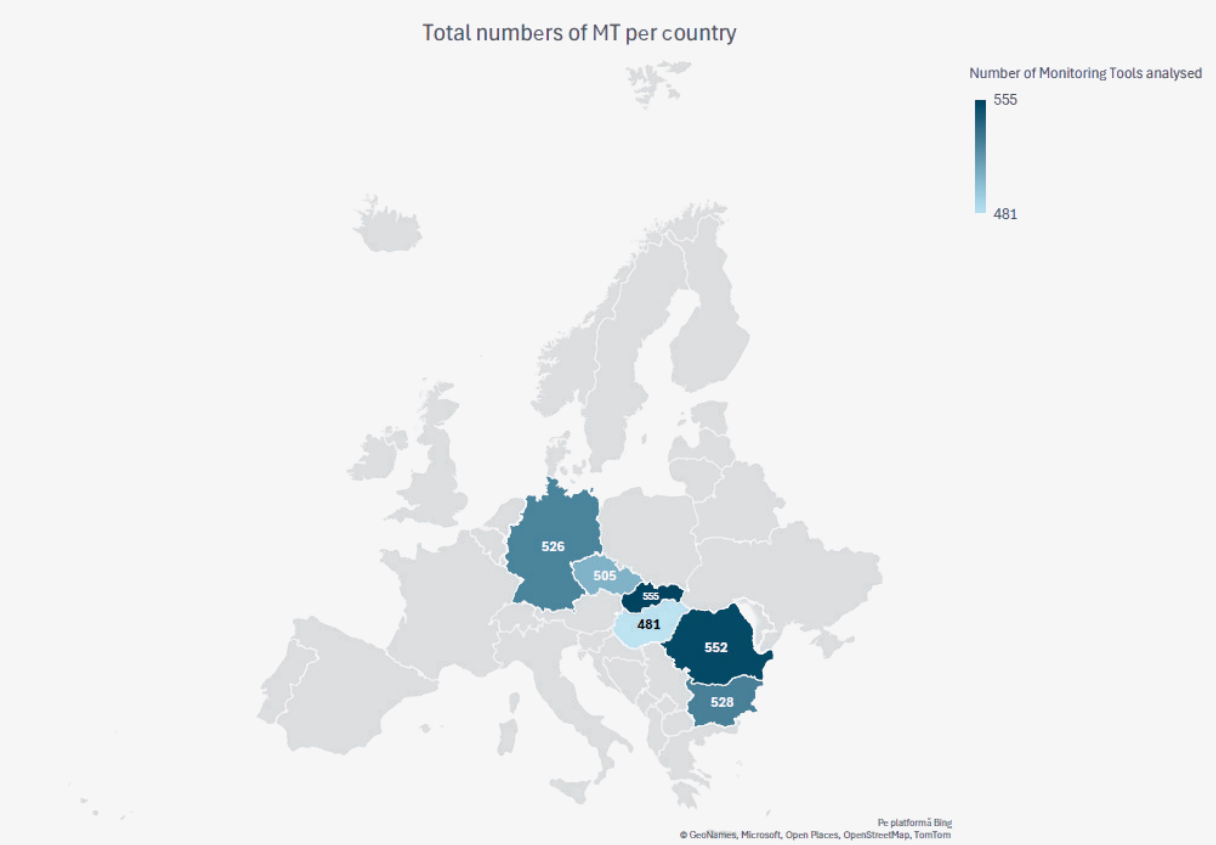
- a. To analyse **the discursive repertoires employed against Roma individuals and communities**, identifying the dominant forms in which antigypsyist online hate speech is disseminated. Concurrently, this dimension will pinpoint frequently recurring themes and subthemes, characterise various forms of online hate speech, assess levels of aggressiveness and forms of toxicity (particularly racially loaded toxic discourse), and ultimately, **explore 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.

Data collection and analysis were conducted using a mixed-methods content analysis, employing a purpose-built monitoring tool (MT) within an online application (Google Form). A comparative perspective was adopted along two dimensions: a national dimension, comparing data across the 2 dimensions (evidence-to-action), and a European dimension, comparing findings across the 6 countries participating in the project.

The research developed a methodology to collect and categorise a wide range of data on antigypsyist online hate speech, which was discussed and refined with national researchers in online meetings to ensure a common understanding and scientific validity.

The MT comprises six key components: (1) general information regarding authors and recipients; (2) a general topic assessment and reach; (3) categorisation 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.

This report presents **a content analysis of 3147 cases collected over a 12-month period** (October 2024 - September 2025) using a standardised monitoring tool (MT) instrument (see Annex 1). The MT, tailored for social media monitoring activities, incorporated specific codes and descriptions informed by the research questions and project objectives. The content analysis, based on the MT results, informs the findings presented in this report. Findings are contextualised within the broader situation of each country, as detailed in six corresponding National Reports.



Each partner organisation employed the same MT, adapted to national specificities by five trained and counselled young Roma monitors, who were familiarised with the instrument during a dedicated training and piloting period, in accordance with social research ethics and standards. These **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. 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. Approximately eight completed MTs per person per month were included in the periodical Synthesis Reports (see Annex 2).

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

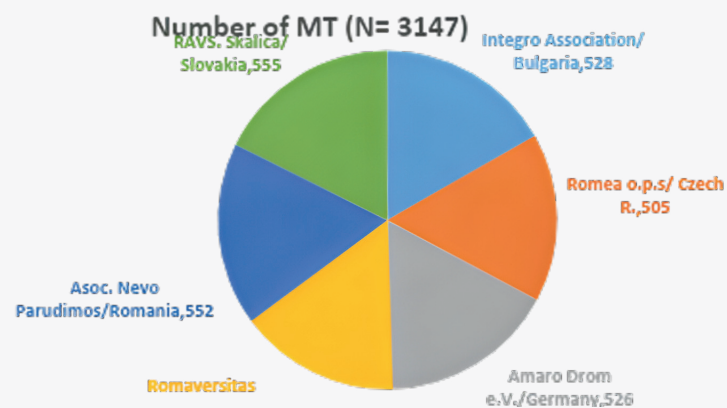
- 1st period: Oct-Dec 2024 - Report due January 2025
- 2nd period: Jan-Mar 2025 - Report due April 2025
- 3rd period: Apr-Jun 2025 - Report due July 2025
- 4th period: Jul-Sept 2025 - Report due October 2025

The National Synthesis Reports were used to create a common database, facilitating the present analysis and cross-national comparative analysis using a common language, despite the monitoring tools being completed in the national language of each young Roma monitor. At the same time, each of the 4 trimestral Reports included a selection of 2-3 cases that detailed the study cases selected by the national team and were considered representative of antigypsyist online hate speech and of counteracting it.

Case sampling relied on non-probabilistic methods (convenience sampling), random selection, and the operator's subjective judgment. This approach aimed to provide insight into the typical content Roma youth encounter on their personal accounts and to highlight the constant emotional pressure they may face in the online public sphere. A secondary objective was to observe how media algorithms target Roma users based on their preferences, consumption patterns, and user profiles (see Personal Observations section).

While the results achieved a degree of expectancy and saturation, they should not be generalised to the entire Roma community. However, the **total number of cases (N=3147 across six countries) provides a solid foundation for developing and advocating** for social, digital, and educational policies to combat the proliferation of anti-gypsyist online hate speech and foster a safer, more inclusive, and creative online public space. (Figure 1).

Figure 1. Total number of monitoring instruments applied by each organisation during one year period.



Source: Author's own contribution

2.2. participants

In each of the six participating countries, a team of five young Roma hate speech monitors employed the established methodology to collect data over a 12-month period (October 2024 - September 2025). These 30 young Roma monitors, trained and experienced in media monitoring of online anti-gypsyism, were familiarised with the monitoring instrument during a one-month piloting phase (September 2024). A designated national coordinator in each country was responsible for compiling and analysing national data, producing a corresponding national report, and providing ongoing discussion and support to the young Roma monitors throughout their data collection.

Within each of the six countries (**Romania, Bulgaria, Slovakia, Hungary, Germany, and the Czech Republic**), the national team of five utilised the methodology to collect at least 40 cases per month per country, equating to at least eight cases per month per monitor. The project anticipated collecting a **minimum of 480 cases per country per year**.

2.3. tools and methods

The data are collected and analysed using a content analysis research method based on the monitoring tool (MT) instrument applied online. The first 5 parts of MT were analysed quantitatively using frequency and sentiment analyses based on the 3-point scale (positive, neutral, and negative grading, from no hate, weak, to extremely strong levels of perceived hate). The last part (Personal observations) involved a qualitative content analysis of the selected cases reported by the national groups, grounded in a post-positivist Grounded Theory approach (Molnar, 2022).

The MT consists of 6 main parts:

- **general information** – authors/ recipients of the social media platform;
- **general topic** – reach;
- **types of content (visual, text)** – reach; measurement;
- **intensity and frequency** – impact;
- **types of counteract** and
- **personal observations** – (national specificities regarding the functions of reporting mechanisms) – extent; background information – reactions (after reporting activities).

The monitoring tool (MT) employed in this study comprised six key components designed to systematically capture and analyse online anti-gypsyist hate speech:

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. Categorisation of Content Types:

This section addressed the relevance of any relationship between the content and its visual elements, categorised 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 polarisation, spread myths and disinformation, justified exclusion, stigmatisation, 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 (2018), 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 counter-speech responses, including commenting on the post and reporting to the platform or to any legal institution.

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).

The monitoring tool is annexed to the present Report (Annex 1).

f. Personal Observation (Narrative):

This section describes in detail the study cases selected by the monitor, outlining the main idea of the content and how it exemplifies 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 was filled in via an online form (Google Form), and all cases were supervised and discussed by each national team member. Some of the cases were discussed together with the research coordinators from all the partner countries in case clarification was needed

3. results

3.1. quantitative component

3.1.1 Item: General information – authors/ recipients

In the initial phase of the monitoring methodology, the primary focus was on collecting pertinent general information to contextualise the analysis, including identifying the main social media platforms used and the typology of user interactions. It is important to emphasise 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 the social networks they frequently used, which explains the higher proportion of personal accounts (individual actors) compared to official news pages or public institutional accounts; the latter were analysed only after being taken over and disseminated online.

As illustrated in Figure 2 (Online platform users), the predominant categories of users involved in the research—identified as young Roma operators—were private accounts on Facebook, TikTok and X (former Twitter), followed by YouTube and Instagram. Regarding frequency of use, Facebook remains the most widely used social media platform identified by the monitoring team, primarily due to its private and closed-group features. Additionally, Roma individuals with verified accounts—according to the established instructions—reported instances of online hate speech in five out of six countries, with Hungary being the exception, accounting for 1,014 indicators—approximately one-third of the total reported cases. The second most frequently used platform is TikTok, where hate speech is predominantly propagated through private accounts (14%), followed by X (formerly Twitter) with 9% and YouTube with 5%. Regarding online news dissemination, Facebook/Meta is the principal source of information and the primary channel through which public figures distribute their content, particularly in Slovakia, Bulgaria, and Romania.

“ONLINE SPACES REMAIN DOMINATED BY NEGATIVE REPRESENTATIONS.”

Furthermore, TikTok has gained significant prominence among user preferences for social media influencers in the Czech Republic, and for public figures in Romania and Hungary. In Romania, Instagram is notably favoured for engagement with social media influencers (SMIs), whereas in Germany, Twitter and YouTube are preferred platforms for similar interactions. Institutional public accounts are most frequently engaged in Slovakia and Romania via Facebook, while in Hungary, Instagram is the platform predominantly used for official institutional communication, according to our monitors' reports.

The provided data for item 1 (Figure 2) details social media platform engagement across five distinct user categories and six European countries (Bulgaria, Czech Republic, Germany, Hungary, Romania, Slovakia), revealing a diverse and often country-specific landscape of digital interaction.

Analysis of **Personal/Private accounts** demonstrates Facebook (Meta) as a predominantly utilised platform, recording the highest engagement in Bulgaria (269 instances), the Czech Republic (219), and Slovakia (206). However, TikTok exhibits a significant presence, particularly in Germany (116) and Romania (128), indicating its growing relevance for individual users. Notably, Slovakia also has a substantial number of private YouTube accounts (116). Conversely, Hungary records a comparatively lower overall engagement for private accounts (104 total).

For **Online Media** outlets, Facebook (Meta) remains a primary channel, notably in Slovakia (28) and Bulgaria (24). A striking observation is Hungary's overwhelming reliance on "Other" platforms (235 out of 256 total online media engagements), suggesting a significant proportion of media consumption occurs outside the explicitly listed major social media platforms in this country. Germany also displays a diversified media consumption, with YouTube (14) and "Other" (23) contributing notably. The Czech Republic shows remarkably low engagement with listed social media among online media outlets, with only 1 observation.



In the realm of **Institutional Public accounts**, Facebook (Meta) remains a key platform in Slovakia (10) and Romania (6). Interestingly, Instagram is the leading platform for institutional communication in Hungary (10), whereas Germany's institutional presence is more distributed across platforms.

Regarding **Public figures and Social Media Influencers (SMIs)**, the data highlights a clear shift towards visually-centric, dynamic content platforms. TikTok exhibits strong engagement for public figures in both Hungary (12) and Romania (14). For SMIs, Instagram dominates in Romania with 41 reported instances, significantly surpassing other platforms within that country for this category. The Czech Republic also shows a strong preference for TikTok among SMIs (26). Germany, in contrast, demonstrates a more distributed landscape for SMIs, with notable engagement on Twitter (X) (29) and YouTube (22), reflecting a diverse influencer ecosystem.

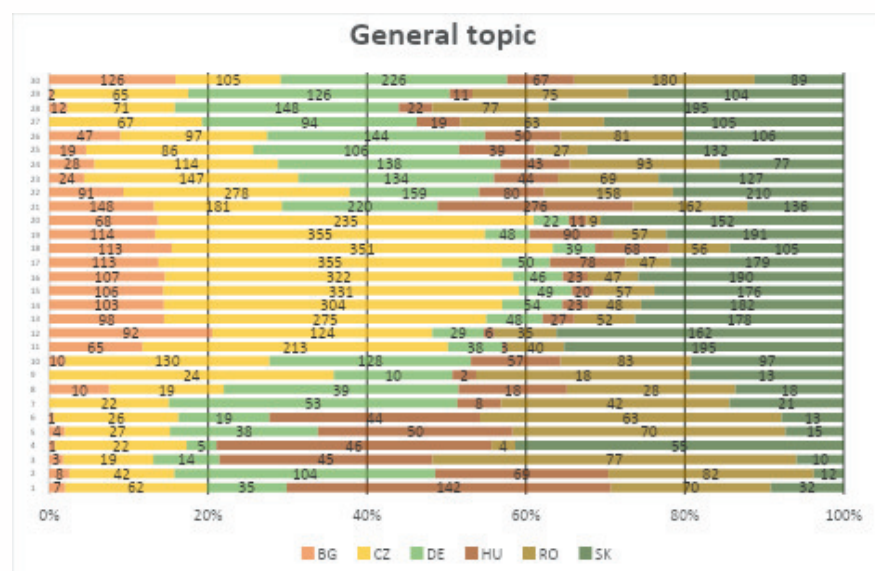
3.1.2 Item: General topic – reach (QII a)

In the second phase of this research, the primary objective was to identify the salient thematic categories pertaining to the Roma population within the analysed media content. This identification was complemented by an in-depth evaluation of the sentiment associated with these themes, employing a structured sentiment analysis scale. The monitoring period, spanning from October 2024 to September 2025, was notably influenced by concurrent contextual factors, including various electoral campaigns (e.g., presidential and parliamentary elections in Romania), cultural and religious holidays, and significant sporting events. The principal thematic categories selected for analysis were systematically classified as follows:

- **Criminality Attributed to Roma:** Encompassing issues such as theft, murder, begging, and vandalism.
- **Social Aspects:** Covering topics like housing, welfare, poverty, and immigration.
- **Educational Aspects:** Addressing school dropout rates, schooling conditions, and scholarship provisions.
- **Health and Hygiene Issues:** Including the impact of healthcare conditions, access to hospital and medical services, births and abortions.
- **Social Movements and Civil Society:** Focusing on protests, civil rights advocacy, and the role of non-governmental organisations.
- **Politics:** Pertaining to political representation, party affiliations, and electoral processes.
- **Roma Leaders:** Highlighting the involvement and roles of community leaders, including women.
- **Roma Culture:** Exploring various aspects of Roma cultural identity.
- **Other Issues/Events:** A residual category for miscellaneous topics concerning Roma.

For each identified thematic category, operators were instructed to assess its sentiment using a three-point scale (positive, neutral, and negative). This methodology not only facilitated accurate thematic classification but also aimed to capture how the repetition or specific portrayal of themes or events within media content might generate additional connotations or distorted interpretations that affect Roma individuals. (Figure 3)

Figure 3. An overview of the perceived main topics of the media content



Source: Author's own contribution

The dataset comprises 15 080 reported instances across all sentiments, countries, and topics. A preliminary overview reveals the following distribution:

Total Positive Mentions: 2186 (14,49%)

Total Negative Mentions: 6120 (40,58%)

Total Neutral Mentions: 6774 (44,92%)

Neutral sentiment instances constitute the largest proportion (44.92%) of the data, closely followed by Negative (40.58%), and then Positive mentions (14.49%). This suggests a predominant tendency towards factual or uncoloured reporting, with negative framing more prevalent than positive framing when sentiment is explicitly expressed. This implies that, while much discourse is objective, sentiment is more often critical or problematic than affirmative or celebratory.

“NEGATIVE THEMES OUTNUMBER POSITIVE REPRESENTATIONS.”



Analysis of Positive Sentiment

The total count of positive mentions across all countries and topics is 2186. This category has the lowest overall frequency among negative and neutral sentiments, indicating that positive coverage or discourse related to these topics is less common.

Among the countries, Romania (RO) recorded the highest total number of positive mentions (537), driven notably by 'Education aspects' (77), 'Social aspects' (82), and 'Other facts' (83). Following closely, Hungary (HU) (481 instances) with 'Crimes committed' (HU: 142). Germany (DE) (445 instances) also exhibited significant positive reporting on Social aspects (DE: 104). Bulgaria (BG) registered the lowest total positive instances (44), with several topics, such as 'Roma leaders (including women)', 'Sports events', and 'Health/sanitary aspects', receiving negligible or zero positive mentions. This suggests a significant disparity in the propensity to report positively across the participating countries.

Regarding topics, 'Social aspects' (317) and 'Education aspects' (168) received notable positive attention. 'Roma leaders (including women)' (146 instances) also received notable positive attention. In contrast, 'Sports events (contests, games)' (67 instances) recorded the fewest positive mentions overall, suggesting this area is less commonly framed positively in the analysed discourse or less covered in social media content. Regarding specific topics, 'Other facts' emerged as the most frequently covered positively (583 instances across all countries), largely influenced by high counts in CZ (130), DE (128), and SK (97). This broad category seems to capture various positive, non-categorised events or information. 'Crimes committed' surprisingly also shows a notable positive count (348), primarily due to a high number from Hungary (142). This might indicate reports on successful crime prevention, resolutions, or positive community responses to crime.

Analysis of Negative Sentiment

With a total of 6120 instances, negative sentiment accounts for a substantial portion of the discourse, exceeding positive mentions by a significant margin. This indicates a general tendency towards framing these topics in a critical or problematic light.

Germany (DE) reports the highest aggregate negative mentions (1 495 instances), with significant contributions from 'Crimes committed' (220), 'Social aspects' (159), and 'Health/sanitary aspects' (138). Slovakia (SK) (1,281 instances) and the Czech Republic (CZ) (1,211 instances) also show substantial negative reporting. Conversely, Hungary (HU) records the lowest total negative instances (651), suggesting a less prevalent negative framing there than in other nations, particularly for 'Crimes committed' (276) and 'Social aspects' (80).

Czech Republic (CZ) exhibits the highest negative mentions on 'Social aspects' (278), while Slovakia (1281 mentions) encounters exceptionally high numbers across nearly all topics, particularly 'Roma leaders (including women)' (105), 'Sports events (contests, games)' (104), and 'Cultural events' (195). This suggests a generally more negative media landscape or reporting style in SK for the subjects examined. For Bulgaria, the listed topics registered a high level of negative framing.

'Crimes committed' emerges as the most frequently covered negative term (1,123 instances in total), largely driven by high counts in HU (276), DE (220), and CZ (181). Following this, 'Social aspects' (976 instances) also exhibits significant negative associations. These findings suggest that 'Crimes committed' and 'Social aspects' are often portrayed with critical or problematic undertones. In contrast, 'Sports events (contests, games)' (383 instances) received fewer negative mentions overall, although it still represents a considerable volume of negative discourse.

Analysis of Neutral Sentiment

Neutral sentiment registered a total of 6774 mentions. This suggests that a significant portion of reporting is factual, descriptive, or devoid of explicit positive or negative connotations.

Czech Republic (CZ) exhibits an exceptionally high volume of neutral content (2,865 instances), dominating nearly all topics, particularly 'Roma leaders (including women)' (355), 'Sports events (contests, games)' (355), and 'Cultural events' (351). This suggests a strong tendency toward factual, non-emotive reporting in CZ for the subjects examined. Slovakia (SK) (1,710 instances) and Bulgaria (BG) (979 instances) also show high neutral content. Conversely, Hungary (HU) records the fewest total neutral instances (349), indicating that discourse in HU is less frequently neutral than in other countries.

'Roma leaders (including women)' (822 instances) and 'Sports events (contests, games)' (855 instances) register the highest number of neutral mentions, primarily driven by the high counts in CZ. 'Politics/elections' (735 instances) also demonstrates a considerable neutral volume. These findings indicate that these topics are often presented in a straightforward, uncoloured manner. In contrast, 'Crimes committed' (554 instances) and 'Social movements and ngo' (739 instances) also show substantial neutral reporting, reflecting their factual reporting aspects.

A deeper comparative analysis reveals intricate relationships between sentiment categories across countries and topics.

- **Bulgaria** exhibited the lowest positive share (2.89%), a relatively high negative component (32.69%) and 64.4% neutral mentions. BG presents the lowest positive counts overall, with a relatively higher proportion of negative mentions for topics like 'Crimes committed' (148) and 'Social aspects' (91) compared to its other sentiment categories.
- The **Czech Republic** exhibits the highest proportion of neutral sentiment (64.1% of its total instances), alongside a substantial negative component (27.09%) and a relatively low positive component (8.8%). This suggests a predominantly factual discourse with significant critical undertones. CZ demonstrates a remarkably high volume of neutral reporting (2865), particularly on 'Roma leaders (including women)' (355) and 'Sports events' (355).
- **Germany** exhibited 18,83% positive, 17,9% neutral, and 63,26% negative mentions. DE presents a notable balance on negative reporting, with a strong positive presence in 'Other facts' (128) and 'Social aspects' (104).
- **Hungary** exhibited 32.27% positive, 23.56% neutral, and 43.95% negative mentions. This profile suggests a more polarised discourse, with strong positive or negative framing rather than neutrality. HU shows the high negative counts and a high proportion of positive reporting for 'Crimes committed' (276 negative), respectively (142).
- **Romania** exhibited 27.25% positive, 22,74% neutral, and 50% negative mentions. RO shows a negative sentiment distribution, though with a high proportion of neutral mentions (448).
- **Slovakia** exhibited 8.72% positive, 39.03% negative, and 52.18% neutral mentions. SK has a significant negative representation in 'Crimes committed' (136) and 'Social aspects' (210), but also substantial neutral reporting (195 for 'Crimes committed').



Crimes committed: Across all countries, this topic is covered with predominantly negative sentiment, 55.54%. Notably, this topic attracts a high volume of neutral reporting 27.35%, suggesting a factual approach, yet also significant negative mentions, particularly in HU (276) and DE (220), contrasting with positive reports mainly from HU (142). Nevertheless, 'Crimes committed' represents the main topic related to Roma.

Education aspects: Across all countries, this topic received 12.07% positive, 39.18% negative, and 48.77% neutral mentions. Positive mentions are moderate, while negative and neutral reports are substantial. CZ leads in negative (147) and neutral framing (275), while SK (127) and DE (134) contribute significantly to negative reporting, too.

Social movements and NGOs: Across all countries, this topic received 15.08% positive, 30.25% negative, and 54.65% neutral mentions. This topic exhibits moderate positive, high negative, and substantial neutral mentions. SK shows prominent negative framing (132), while CZ (331) and BG (106) contribute significantly to neutral reports.

Roma leaders (including women): Exhibits low positive coverage (146), high neutral (882), and significant negative (348) sentiment. This implies that while their activities are often reported factually, explicit positive framing is limited, and negative portrayal is common. This topic has a very low positive representation (0 in BG), is frequently framed negatively, especially in SK (105), and shows varying levels of neutral reporting across countries, with a notably low count RO (47).

Social aspects: Across all countries, this topic received a predominant negative mention of 56.05% points. This topic shows a notable balance of sentiments, with high negative (976) and neutral (448) mentions, as well as a considerable number of positive mentions (317), particularly in DE (104) and RO (82).

Health/ sanitary aspects: Across all countries, this topic received 9.92% positive, 36.679% negative, and 53.28% neutral mentions. Characterised by low positive mentions and high negative and neutral reporting. SK (55) stands out with a significant positive count for this topic, in contrast to high negative counts in CZ (114), DE (138), and RO (93).

Politics/ elections: Across all countries, this topic received 11.64% positive, 36.81% negative, and 51.54% neutral mentions. This category generally sees low positive and high negative and neutral mentions. DE consistently leads in negative framing (144), while CZ (322) and SK (190) show relatively balanced reporting, though still predominantly neutral.

Cultural events: Across all countries, this topic received 9.5% positive, 37.79% negative, and 52.69% neutral mentions. Positive mentions are moderate (132), while negative reporting is high (525), especially from SK (195) and DE (148). Substantial neutral instances from CZ (351) and BG (113). This suggests cultural events are sometimes presented negatively despite their positive nature.

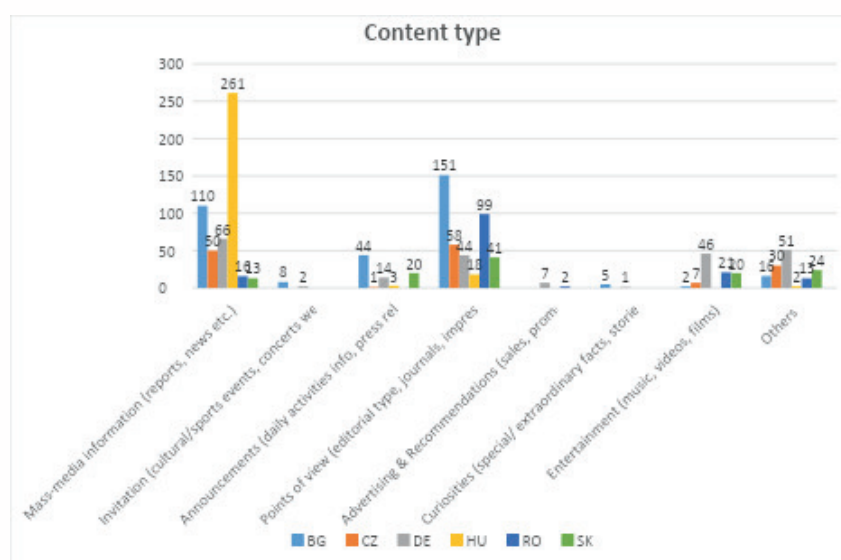
Sports events (contests, games): Across all countries, this topic received 5.1% positive, 29.3% negative, and 65.5% neutral mentions. Similarly, this topic registers low positive engagement (67) and high neutral reporting (855), particularly in CZ (355). Negative reporting is registered in DE (126) and SK (104), while BG registers 0 positive mentions on this topic.

The analysis highlights significant variations in sentiment distribution both across countries and across topics. Neutral reporting is the most prevalent form, suggesting a foundational layer of factual dissemination. However, **negative sentiment consistently outweighs positive sentiment**, particularly pronounced in countries like Germany (DE), Hungary (HU) and Romania (RO). Countries like Bulgaria (BG), the Czech Republic (CZ), and Slovakia (SK) exhibit a comparatively lower incidence of negative framing and a higher proportion of neutral reporting, indicating divergent media portrayals. Constantly, topics like ‘Crimes committed’ (2025 mentions), ‘Social Aspects’ (1741), ‘Politics/elections’ (1426) and ‘Education’ (1391) gathered the highest scores, while, conversely, ‘Health’ (1340) and ‘Sports’ (1305) received the lowest numbers of mentions. These findings underscore the importance of nuanced, country- and topic-specific approaches when evaluating public discourse and media representation related to these sensitive areas and to the perception of the media content addressed by the Roma young monitors. While neutral reporting constitutes a significant portion, explicit positive framing is notably low (14.49% overall, with some zero positive mentions in Bulgaria). Conversely, negative sentiment is prominent (40.58%), especially in Germany, the Czech Republic and Slovakia, indicating a tendency towards critical or problematic portrayal. More broadly, media coverage across all topics leans toward neutral and negative reporting, with factual accounts prevalent, but when sentiment is expressed, negative framing significantly outweighs positive. This suggests that, while factual reporting exists, issues pertinent to the Roma community and broader social topics are frequently framed critically or lack positive contextualization in the analysed media landscape.

In the context of the analysed message content, its categorisation within specific genre classifications was undertaken. This categorisation framework encompasses the following types: informational mass media (e.g., reports, news articles); invitations; announcements; opinion-based content (e.g., editorials, journalistic impressions, comments); promotional and recommendation materials (e.g., sales, promotions, job listings, travel offers); anomalous or intriguing content (e.g., extraordinary facts, anecdotal narratives); entertainment media (e.g., music, videos, films); and miscellaneous categories. This classification was utilised to systematically characterise the nature and intent of the discourse under investigation. The results are presented in Figure 4.

“VISUAL CONTENT PLAYS A KEY ROLE IN SPREADING ANTIGYPSYIST NARRATIVES.”

Figure 4. Content type distribution per country

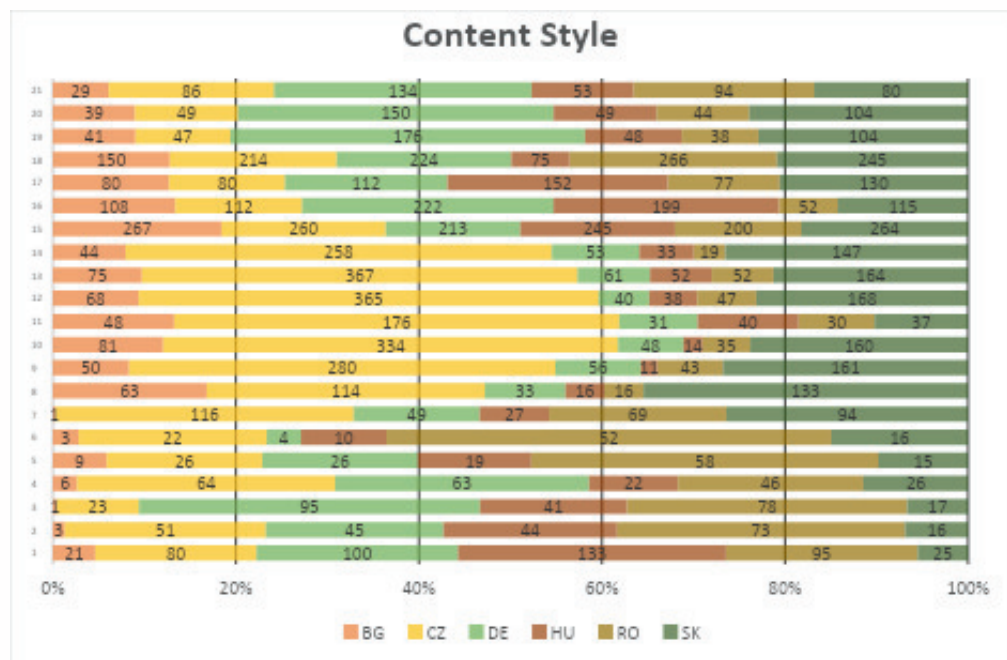


Source: Author's own contribution

Online media constituted the largest category of mentions (516), followed by expressed opinions and commentary (411). Reports from Germany and Hungary predominantly focused on general media content, whereas those from Bulgaria, Romania, the Czech Republic, and Slovakia primarily highlighted expressed opinions.

The next dimension presents a comparative analysis of content types categorised by sentiment (Positive, Negative, and Neutral). The content types examined include Emotional, Formal/Official, Call to Action, Funny, Artistic/Fictional, Scientific, and Others. The objective is to identify prevalent sentiment patterns associated with specific content types and country-specific reporting tendencies, utilising a scientific analytical framework. (Figure 5).

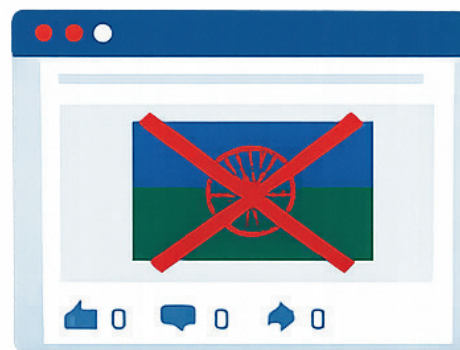
Figure 5. An overview of the content frame using a sentiment scale



Source: Author's own contribution

The dataset comprises 11,272 instances across all available content types, sentiments, and countries. The aggregated sentiment distribution is as follows:

- Positive Sentiment: 1,784 instances (15.82%)
- Neutral Sentiment: 4,061 instances (36.02%)
- Negative Sentiment: 5,427 instances (48.14%)



This overall distribution indicates a strong predominance of negative sentiment, accounting for nearly half of the analysed content, followed by neutral sentiment. Positive sentiment represents the smallest proportion, suggesting that content is more frequently framed critically or objectively rather than affirmatively or constructively. Negative content substantially outweighs positive content, suggesting a bias towards critical or problem-oriented framing when sentiment is explicitly conveyed. It is important to note that neutral data for the "Others" content type was not provided in the dataset and is thus excluded from these neutral totals.

Analysis of Positive Sentiment

Positive sentiment instances total 1,784, representing the smallest proportion of the analysed data.

Romania (RO) recorded the highest volume of positive content (471 instances), constituting 33.9% of its total content. This positive output in RO was significantly driven by 'Emotional' (95), 'Formal/Official' (73), 'Call to action' (78), and 'Scientific' (52) content. Conversely, Bulgaria (BG) exhibited the lowest positive content contribution (44 instances), accounting for only 2.5% of its total content, suggesting very limited positive discourse across all examined content types in this country. Germany (DE) and the Czech Republic (CZ) also showed substantial positive contributions (382 instances each).

The 'Emotional' content type registered the highest number of positive instances overall (454), suggesting that emotionally resonant content is most frequently associated with positive framing. This was particularly evident in Hungary (HU, 133), Germany (DE, 100), and Romania (RO, 95). The 'Others' content type also showed a strong positive presence (356 instances). In contrast, 'Scientific' content yielded the fewest positive mentions (107), with extremely low numbers in BG (3) and DE (4), suggesting that scientific discourse is rarely presented positively in this context.

**“ONLINE
ANTIGYPSYISM
REMAINS
WIDESPREAD AND
INSUFFICIENTLY
ADDRESSED.”**



Analysis of Neutral Sentiment

Neutral sentiment accounted for 4,061 instances, making it the second-most frequent sentiment category.

Czech Republic (CZ) demonstrated an exceptionally high volume of neutral content (1,894 instances), accounting for 46.4% of its total content. This high neutrality is evident across various content types, particularly 'Artistic/Fictional' (365), 'Scientific' (367), and 'Call to action' (334), which are particularly prominent. This suggests a pervasive neutral framing in the content types analysed within CZ. Conversely, Hungary (HU) registered the lowest volume of neutral content (204 instances), indicating comparatively less neutral discourse. Slovakia (SK) (970 instances) also showed substantial neutral contributions.

'Scientific' content globally recorded the highest number of neutral instances (771), largely propelled by CZ (367) and SK (164). Following this, 'Artistic/Fictional' (726) and 'Call to action' (672) also exhibited significant negative associations. These findings suggest that factual or action-oriented content is often framed objectively. 'Funny' content, while still having a considerable number of neutral mentions (362), had the lowest overall neutral count among the specified content types. 'Emotional' content recorded the second lowest neutral count (375), suggesting that emotional content is less frequently presented without a strong sentiment associated with.

Analysis of Negative Sentiment

Neutral sentiment instances totalled 5,427 mentions, representing the largest proportion of the data. This highlights a pervasive tendency to frame content in a critical or problem-oriented way.

Germany (DE) reported the highest volume of negative content (1,231 instances), representing 22.68% of all negative sentiment. This was driven by significant contributions from 'Formal/Official' (222), 'Funny' (224), and 'Emotional' (213) content. Slovakia (SK) (962 instances) and Hungary (HU) (768 instances) also show substantial negative reporting. Bulgaria (BG) records the lowest total negative instances (685), although this still constitutes a considerable volume. Romania (RO) registered the second-lowest total number of negative instances (771), indicating a significant, albeit lesser, focus on negative reporting. Hungary (HU) also exhibited a high proportion of negative content, with more than 800 mentions.

The 'Emotional' content type exhibited the highest number of negative instances globally (1,449), with consistent high contributions from BG, CZ, DE, HU, RO, and SK (ranging from 200 to 267). 'Funny' content also showed a very high volume of negative reporting (1,174). This suggests that much emotional and humorous content, as well as official communications, are frequently used to convey negative sentiment. 'Scientific' content recorded the lowest negative count (435), indicating that when scientific content is included, it is less likely to be presented without sentiment than emotional or funny content. As previously noted, negative data for the "Others" content type was not provided.

The analysis reveals distinct sentiment profiles for each country:



- **Bulgaria:** Displays the highest proportion of negative sentiment (63.2% of its total instances), with a substantial neutral component (39.8%) and the lowest positive share (4.1%). This indicates a highly critical discourse in BG.
- **Czech Republic:** Exhibits a high proportion of neutral sentiment (46.4% of its total instances), alongside a substantial negative component (37.7%) and a relatively low positive share (9.4%). This suggests a predominantly factual discourse with significant critical undertones.
- **Germany:** Exhibits the highest proportion of negative sentiment (63.6% of its total instances), with a substantial neutral component (16.6%) and a moderate positive share (19.7%).
- **Hungary:** Shows a significant negative sentiment share (821 instances), a moderate positive sentiment share (296), and the lowest neutral sentiment share (204). This profile suggests a more polarised discourse with strong negative or positive framing.
- **Romania:** Exhibits the highest positive share (471) among all countries. Possesses a comparatively with Hungary the same trend of polarised sentiment profile, with a strong positive share, alongside substantial neutral (242) and negative (771) distribution.
- **Slovakia:** Shows a significant negative bias (1042 instances), second only to CZ (848), balanced by a substantial neutral share (970) and a low positive contribution (209).

Specific content types display characteristic sentiment distributions:

Emotional & Funny: Both are largely negative (63.6% and 66.6% respectively), suggesting that emotional discourse is often used to convey critical or problematic messages. Humour and emotional expression are often used in a critical, satirical, or problematic way.

Artistic/Fictional: These content types are predominantly negative (1174 mentions in total), suggesting that even the artistic and cultural domains are used critically and problematically by Roma community members.

Others: Based on available data (excluding neutral), this category has the highest positive proportion (39.1%) but also a significant negative share (60.9%), indicating a diverse range of content that can be either highly positive or highly negative depending on its specific nature.

Formal/Official & Call to Action: These types show a strong negative tendency (36.6% and 43.1% respectively), with 'Call to action' having the highest neutral share. Negative content is substantial, while positive content is moderate. This suggests that official communications or calls for action are often framed in response to problems or critical situations.

Scientific: This content type has a notably low positive representation (107 instances) but a high proportion of neutral (771) and negative sentiment (435). This might indicate that scientific discourse often focuses on problems, challenges, or critical evaluations.

“ONLINE ANTIGYPSYISM IS OFTEN NORMALISED, REINFORCING A CLIMATE OF IMPUNITY.”

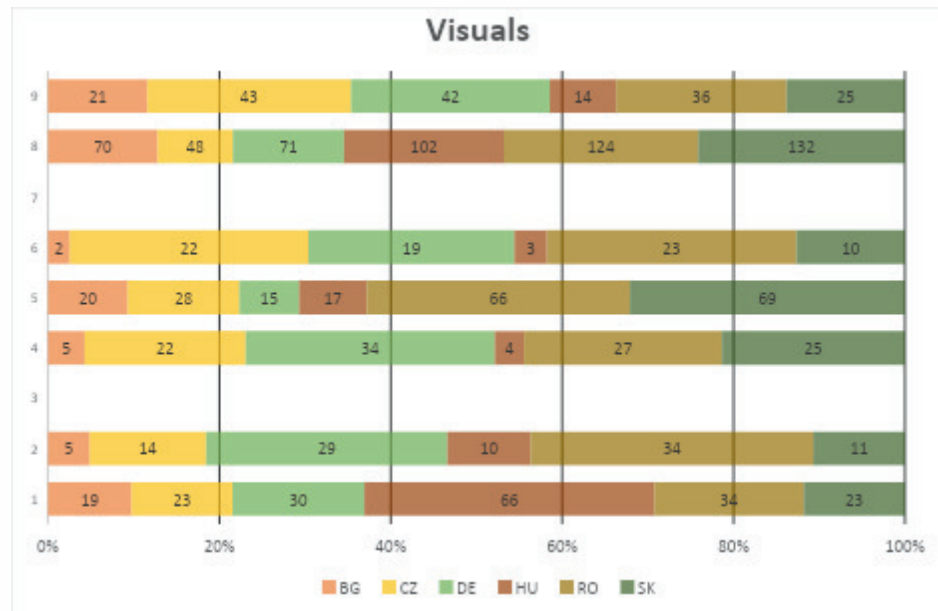
The analysis reveals a complex interplay between content types, national reporting tendencies, and sentiment. While neutral framing predominates overall, negative sentiment is consistently more frequent than positive sentiment, particularly within factual, action-oriented, artistic, and scientific content. Countries such as the Czech Republic, Bulgaria, Germany, and Slovakia exhibit a pronounced negative bias.

The dominance of negative sentiment in 'Emotional', 'Funny', 'Artistic/Fictional', and 'Scientific' content types suggests that these content types are frequently used to convey criticism or highlight problems. These findings underscore the importance of considering both national media cultures and the inherent nature of content when analysing sentiment, as different content types are clearly predisposed to specific emotional valences in public discourse. The lack of neutral data for the "Others" content type limits the completeness of the sentiment profile for this category.

3.1.3. Item: Content (visual, texts) – reach

The data indicate the frequency with which media-reported content includes visual components across the six countries (Figure 6).

Figure 6. An overview of the non-textual forms within the monitored content



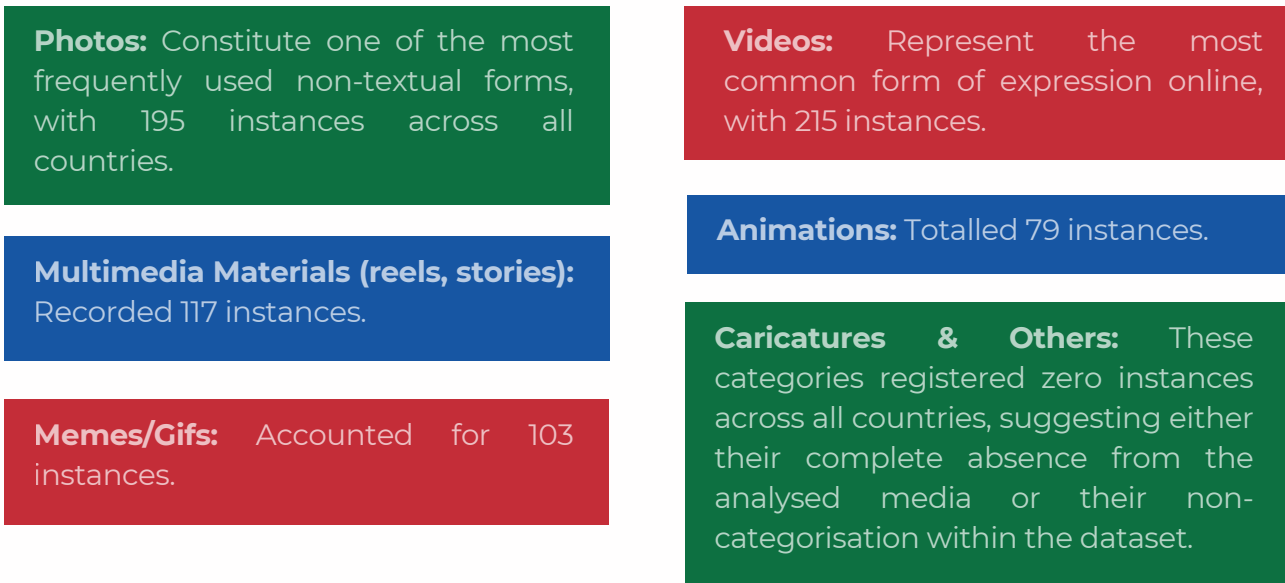
Source: Author's own contribution

A total of 709 articles were analysed for visual inclusion. Out of these, 547 (77.15%) included visuals, while 181 (25.52%) did not. This demonstrates a general propensity for media articles in the surveyed countries to incorporate visual elements.

- **Romania (RO):** Exhibited the highest absolute number of articles with visuals (184) and the highest proportion of visual inclusion (77.5% of its total articles). This suggests a strong reliance on visual content in Romanian media.
- **Slovakia (SK):** Followed closely with 138 articles including visuals, representing 84.1% of its total articles, indicating a very high rate of visual integration.
- **Hungary (HU):** Demonstrated the highest percentage of visual inclusion (87.9%), with 102 out of 116 articles featuring visuals. This country shows a clear preference for visually supported media.
- **Germany (DE):** Maintained a high inclusion rate (62.8%), with 71 articles containing visuals out of 113.
- **Bulgaria (BG):** Showed the lowest absolute number of articles with visuals (70) but a respectable inclusion rate of 76.9%.
- **Czech Republic (CZ):** Recorded 48 articles with visuals, representing 52.7% of its total, which is the lowest proportional inclusion rate among the analysed countries. This suggests that a significant portion of Czech media content might be predominantly text-based.

In summary, Hungary, Slovakia, and Romania show a particularly strong tendency to include visuals, with over three-quarters of their articles containing such elements. The Czech Republic, while still including visuals in over half its articles, does so at a noticeably lower rate than the other nations.

This section analyses the distribution of specific non-textual forms, such as photos, memes/gifs, caricatures, multimedia, videos, and animations, across the six countries.



Romania (RO) exhibited the highest overall usage of non-textual forms, particularly dominating in Videos (66 instances), Photos (34 instances), Memes/Gifs (34 instances), and Animations (23 instances). This highlights a diverse and prominent adoption of visual media in RO.

Slovakia (SK) also demonstrated high usage of Videos (69 instances), Photos (23 instances), and Multimedia materials (25 instances).

Hungary (HU) showed a strong preference for Photos (66 instances), significantly more than any other country, but had comparatively low use of other non-textual forms, such as Multimedia (4) and Animations (3).

Czech Republic (CZ) used a diverse range of non-textual forms, with high numbers of Videos (28 instances), Multimedia materials (22 instances), Animations (22 instances), and Photos (23 instances), indicating a varied visual strategy.

Germany (DE) maintained a substantial presence in Photos (30 instances) and Multimedia (34), but was less prominent in Videos (15) and Animations (19).

Bulgaria (BG) showed the lowest overall usage of non-textual forms, with Photos (19 instances) and Videos (20 instances) being the most frequent, and very limited use of Memes/Gifs (5), Multimedia (5), and Animations (2).

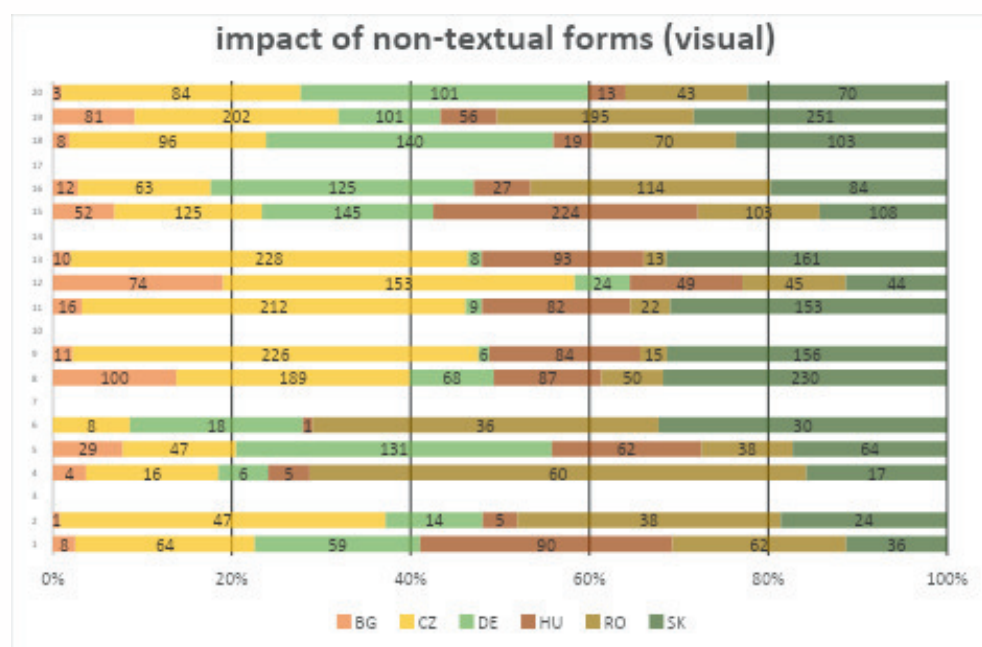
The analysis of visual content reveals a predominant trend towards the inclusion of non-textual forms in media articles across the six European countries, with an average of over 80% of articles incorporating visuals. Hungary, Slovakia, and Romania exhibit the highest proportion of visual integration, suggesting a media landscape that strongly leverages visual communication. Conversely, the Czech Republic displays the lowest rate of visual inclusion, indicating a more text-centric approach in a significant portion of its media.

Regarding non-textual forms, **Photos** and **Videos** are the most ubiquitous. Romania stands out for its comprehensive adoption of various non-textual forms, including a strong presence across photos, videos, memes/gifs, and animations. Hungary, while having a high overall visual inclusion rate, particularly emphasises photos over other multimedia formats. The complete absence of 'Caricatures' and 'Others' from the dataset across all countries suggests a specific focus of the data collection on other visual categories or their marginal use in the analysed media.

The presented data offers valuable insights into the quantitative aspects of visual media utilisation, highlighting country-specific preferences and overall trends in the incorporation of non-textual elements into media articles.

In a in-depth analysis of the dataset, we present a detailed breakdown of different visual content types (Photos, Memes/Gifs, Multimedia materials, Videos, and Animations) categorized by sentiment (Positive, Negative, Neutral) across six European countries: Bulgaria (BG), Czech Republic (CZ), Germany (DE), Hungary (HU), Romania (RO), and Slovakia (SK). (Figure 7).

Figure 6. An overview of the non-textual forms within the monitored content



Source: Author's own contribution

The aggregated data across all visual types and countries totals 6,456 instances. The sentiment distribution is as follows:

- **Positive Sentiment: 1,020 instances (15.79%)**
- **Neutral Sentiment: 2,618 instances (40.55%)**
- **Negative Sentiment: 2,818 instances (43.64%)**

This overall distribution indicates a strong predominance of negative sentiment, constituting the largest proportion of analysed visual content, closely followed by neutral sentiment. Positive sentiment represents the smallest proportion. This suggests that visual content is most frequently framed critically or objectively rather than affirmatively or constructively.

Analysis of Positive Sentiment in Visual Content

Positive sentiment accounts for the smallest proportion of visual content (15.79%), suggesting that visual media are less frequently used to convey explicit positivity.

- **Germany (DE) and Romania (RO)** lead in the absolute number of positively framed visuals (228 and 234 instances, respectively). In DE, 'Videos' (131) and 'Photos' (59) are key contributors. In RO, 'Multimedia materials' (60), 'Photos' (62), 'Videos' (38), and 'Animations' (36) contribute significantly to positive visual content.
- **Hungary (HU)** also shows substantial positive visual engagement (163 instances), with 'Photos' (90) and 'Videos' (62) being the primary drivers.
- Conversely, **Bulgaria (BG)** recorded the lowest overall positive visual instances (50), with 'Videos' (29) and 'Photos' (8) being the most common forms. Notably, BG and HU report zero positive animations, suggesting a lack of positive animated content in these regions.

Visual Type-Specific Positive Trends:

Photos (319 instances) and **Videos** (371 instances) are the most frequently associated with positive sentiment across all countries, suggesting their utility in conveying positive messages.

Memes/Gifs (129 instances) and **Multimedia materials** (108 instances) also contribute to positive visual content, though to a lesser extent.

Animations register the lowest positive count (93 instances), with particularly low numbers in BG (0), HU (1), and DE (18), suggesting that animated content is less commonly used for positive framing.

Analysis of Neutral Sentiment in Visual Content

Neutral sentiment instances totaled 2,618, representing the second largest sentiment category. This suggests a significant tendency towards factual or objective presentation through visual media.

Czech Republic (CZ) exhibits an overwhelmingly high volume of neutral visual content (1008 instances), disproportionately driven by 'Memes/Gifs' (226), 'Multimedia materials' (212), 'Animations' (228), and 'Photos' (189). This indicates a strong inclination towards factual or non-emotive visual reporting in CZ. Slovakia (SK) also shows a substantial amount of neutral visual content (744 instances), particularly with 'Photos' (230), 'Memes/Gifs' (156), 'Multimedia materials' (153), and 'Animations' (161). In contrast, Germany (DE) records the lowest total neutral visual instances (115), with 'Photos' (68) being the most prominent. This suggests a comparatively less critical visual discourse in DE.

'Memes/Gifs' (498 instances) and 'Multimedia materials' (434 instances) registered the highest number of neutral mentions globally, primarily driven by the high counts in CZ and SK. Following this, 'Photos' (466 instances) also exhibited significant neutral associations. These findings indicate that these visual types are often presented objectively. 'Animations' recorded the lowest neutral count (398), suggesting that animated content is less frequently presented without sentiment compared to other visual types.

Analysis of Negative Sentiment in Visual Content

Negative sentiment is the most dominant category (43.64%), suggesting a strong emphasis on emotional presentation through visual media.

Slovakia (SK) (616 instances) and Germany (DE) (612 instances) exhibit the highest volumes of negative visual content. In SK, 'Videos' (251) and 'Photos' (108) are key. In DE, 'Photos' (145), 'Memes/Gifs' (125), 'Multimedia materials' (140), and 'Animations' (101) are prominent. Czech Republic (CZ) (570 instances) and Romania (RO) (525 instances) also contribute significantly to negative visual content, with 'Videos' (202 for CZ, 195 for RO) being particularly strong. Bulgaria (BG) reports the lowest total number of neutral visual content (156 instances), with 'Videos' (81) and 'Photos' (52) being the most frequent, indicating a lower propensity for neutral visual reporting.

Videos (885 instances) and Photos (757 instances) are predominantly used for negative content, reinforcing their role in factual representation. Memes/Gifs (425 instances), Multimedia materials (436 instances), and Animations (315 instances) also have substantial negative counts, suggesting that these formats are exclusively used for emotional or opinion-based content.

The distribution of negative content across visual types is relatively balanced compared to neutral sentiments, indicating that most visual forms can be used in a more negative manner rather than a positive one in terms of their visual impact on consumers.

Negative sentiment generally leads across all visual types and countries, followed closely by neutral sentiment. Positive sentiment is consistently the least frequent. This highlights a prevailing media landscape in which visuals either present information objectively or highlight issues and problems.

The Czech Republic exhibits a notably high proportion of neutral visual content (55.4% of its total instances), alongside a substantial negative component (31.3%) and a lower positive share (13.3%). This suggests a predominantly factual visual discourse with significant critical undertones. Slovakia shows a high proportion of negative sentiment (35.9% of its total instances), a considerable neutral share (43.4%), and a moderate positive share (10.0%) while Bulgaria displays the highest proportion of neutral sentiment (66.5% of its total instances), with a substantial negative component (49.4%) and the lowest positive share (4.0%). This indicates a highly critical visual discourse in BG. Germany and Hungary show a significant negative sentiment share, 64.1% for DE and 37.8% for HU. Their profiles suggest a more polarised visual discourse with strong negative and low positive framing. Romania also has a comparatively negative sentiment profile (58.07%), with a strong positive share (25.8% of its total instances) and a low neutral distribution (16%).

Photos and Videos are versatile and appear frequently across all sentiment categories. They are particularly strong in negative (757 and 885 instances) and neutral contexts (724 and 389 instances). Memes/Gifs, Multimedia materials, and Animations demonstrate a notable neutral bias, especially in CZ and SK, suggesting their use in satirical, critical, or problem-highlighting contexts. Their neutral usage is also significant, indicating their capacity for factual or descriptive communication.

The analysis of visual content types by sentiment reveals a nuanced picture of visual communication across the six European countries. Negative framing is significantly more prevalent than positive framing, especially in dynamic, easily shareable formats such as memes, multimedia, and animations, in certain countries (e.g., DE, SK). The dominance of negative sentiment in 'Videos' and 'Photos' suggests their frequent leveraging to convey criticism or highlight problems, while 'Memes/Gifs' and 'Multimedia materials' also play significant roles in both neutral and negative contexts. The data suggests that visual media are actively employed not only for factual reporting but also as potent tools for expressing critical perspectives and racist messages.

3.2 intensity and frequency of hate – impact

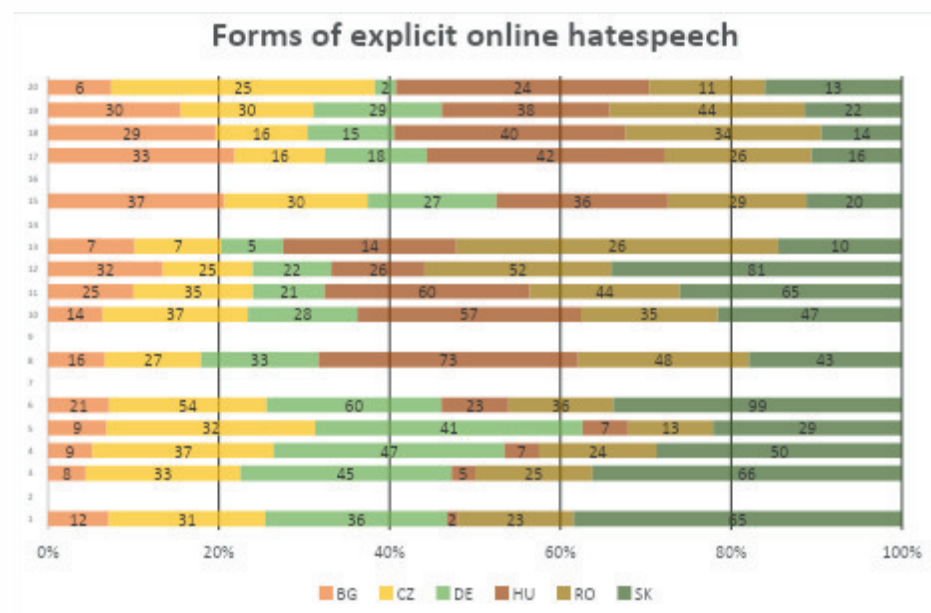
The second component of the monitoring instrument focused on identifying various forms of hate speech, using an adapted analytical model originally developed by Siapera, Moreo, and Zhou (2018). This model was specifically designed for the comprehensive monitoring of racist discourse within online environments. However, due to specific project constraints, including collaborative efforts with young Roma operators and methodological considerations pertaining to data collection and the development of country-specific databases, a simplified three-point scale was adopted to quantify the perceived impact of reported cases. This scale comprised '1 – acceptable,' '2 – strong,' and '3 – extremely strong,' with assessments derived from the personal perspectives of the participating young individuals. The unique contribution of this methodological approach lies in incorporating the subjective appraisal of Roma youth, who are direct recipients of such hate speech manifestations and experience their varied intensity. This approach further extends to the collection of specific contextual examples illustrating the encountered situations.

3.2.1. Explicit online hate speech (A)

The initial phase of the analysis pertaining to hate speech manifestations focused on documenting explicit discursive elements. These elements were directly extracted from the online content under examination, irrespective of its modality (visual or textual) or dissemination platform. This analysis (Figure 8) revealed that the identified discursive forms were primarily characterised by indicators demonstrating a high-intensity aggressive tenor across the following categories:

- Crude epithets
- Racist slurs
- Dehumanizing utterances
- Expressions of intense negative affect (e.g., anger, indignation, hostility)
- Mockery and sarcasm
- Incitement to violence (inclusive of lethal threats)

Figure 8. Forms of explicit online antigypsyist hate speech



Source: Author's own contribution

This dimension investigates the prevalence and perceived aggressiveness of various forms of hate speech in online content across Bulgaria (BG), the Czech Republic (CZ), Germany (DE), Hungary (HU), Romania (RO), and Slovakia (SK). The data is structured around a three-point aggressivity scale: '1 no hate' (interpreted as low aggressivity or acceptable), '2 weak' (moderate aggressivity), and '3 extremely strong' (high aggressivity). Hate speech forms examined include 'Crude epithets,' 'Dehumanising utterances,' 'Saturated emotions (anger, indignation, hostility),' 'Mockery and sarcasm,' and 'Appeal to violence (including killing).' 'Racist slurs' and 'Others' categories reported zero instances across all aggressivity levels and countries, precluding their analysis.

The dataset comprises 2,716 instances of hate speech categorised by aggressiveness level. The overall distribution indicates that 'weak' (moderate aggressivity) is the most frequent category with 1,015 instances (37.4%), followed by 'no hate' (low aggressivity) with 949 instances (34.9%), and 'extremely strong' (high aggressivity) with 752 instances (27.7%).

Level 1: 'no hate' (Low Aggressivity / Acceptable)

This category, representing the lowest perceived level of aggressivity, accounts for 949 instances.

Dominant Hate Speech Forms: 'Appeal to violence (including killing)' (293 instances) is remarkably the most prevalent form within this low aggressivity category. This finding is counterintuitive and suggests that, from the personal perspective of the young Roma operators, certain instances involving 'Appeal to violence' are perceived as 'acceptable' or possessing a very low level of hate/aggressivity. This could indicate nuances in context, subtle framing, desensitisation, or a specific cultural interpretation of aggression. Following this, 'Dehumanising utterances' (182 instances) and 'Saturated emotions' (174 instances) are also frequently classified at this low aggressivity level.

Country-Specific Patterns: Slovakia (SK) registers the highest number of 'no hate' instances (309), largely driven by 'Appeal to violence' (99) and 'Dehumanising utterances' (66). Germany (DE) also shows a high prevalence (229 instances), with 'Appeal to violence' (60), 'Saturated emotions' (47), and 'Dehumanising utterances' (45) being prominent. Conversely, Hungary (HU) records the fewest 'no hate' instances (44), suggesting that most hate speech in HU is perceived at higher aggressivity levels.



Level 2: 'weak' (Moderate Aggressivity)

This category represents moderate aggressivity and accounts for 1,015 instances, making it the most frequent overall.

Dominant Hate Speech Forms: 'Saturated emotions' (250 instances), 'Crude epithets' (240 instances), and 'Mockery and sarcasm' (238 instances) are the most common forms observed at this moderate aggressivity level. This suggests that expressions of strong negative feelings, direct derogatory terms, and disparaging humour are frequently perceived as moderately aggressive. 'Appeal to violence' (69 instances) significantly drops in this category compared to 'no hate,' implying that when direct calls for violence are unequivocally perceived as aggressive, they tend to escalate to higher levels.

Country-Specific Patterns: Hungary (HU) leads this category (230 instances), predominantly driven by 'Crude epithets' (73), 'Saturated emotions' (60), and 'Dehumanising utterances' (57). Slovakia (SK) (246 instances) also shows a high presence, with 'Mockery and sarcasm' (81) and 'Saturated emotions' (65) being notable. Conversely, Germany (DE) reports the fewest 'weak' aggressivity instances (109), particularly low for 'Appeal to violence' (5).

Level 3: 'extremely strong' (High Aggressivity)

This category denotes the highest level of perceived aggressivity, accounting for 752 instances.

Dominant Hate Speech Forms: 'Mockery and sarcasm' (193 instances) surprisingly emerges as the most frequent form at this highest aggressivity level, indicating that certain forms of disparaging humour are considered highly impactful. This is followed by 'Crude epithets' (179 instances) and 'Dehumanising utterances' (151 instances). 'Saturated emotions' (148 instances) also contribute significantly. 'Appeal to violence' (81 instances), while higher than in the 'weak' category, is still lower than other forms in this 'extremely strong' category, which might suggest that truly explicit and undeniable 'Appeal to violence' may be less frequent, or that the majority of 'Appeal to violence' is classified as lower aggressivity.

Country-Specific Patterns: Romania (RO) registers the highest number of 'extremely strong' instances (144), with 'Mockery and sarcasm' (44) and 'Crude epithets' (29) being prominent. Bulgaria (BG) also shows a substantial proportion (135 instances), particularly for 'Crude epithets' (37) and 'Dehumanising utterances' (33). Germany (DE) reports the fewest 'extremely strong' instances (91), with 'Appeal to violence' (2) being notably low.

The distribution across the three aggressivity levels highlights country-specific sensitivities and contextual interpretations of hate speech:

- **Bulgaria (BG):** Exhibits a clear dominance of 'extremely strong' aggressivity (46.9% of its total instances), suggesting that when hate speech is identified in BG, it is often perceived as highly aggressive.
- **Czech Republic (CZ) and Germany (DE):** Both show a majority of hate speech forms categorised as 'no hate' (low aggressivity) (43.0% and 53.4% respectively). This could imply either a greater prevalence of milder forms of hate speech or a different threshold for what constitutes higher aggressivity.
- **Hungary (HU) and Romania (RO):** Both predominantly classify hate speech as 'weak' (moderate aggressivity), accounting for 50.7% and 43.6% of their total instances, respectively. This suggests that the prevailing perception in these countries leans towards moderate forms of aggression.
- **Slovakia (SK):** Displays the highest proportion of 'no hate' (low aggressivity) instances (48.3%), similar to CZ and DE.

A critical observation is the high number of 'Appeal to violence' classified as 'no hate' (Level 1). This phenomenon requires further investigation, as it could indicate desensitisation to such rhetoric among observers, a contextual understanding that diminishes its perceived threat, or a reporting bias in which subtler forms of incitement are more frequently encountered and classified at this lower tier of aggressivity. Conversely, explicit and undeniable 'Appeal to violence' appears less frequent in the highest aggressivity categories. 'Mockery and sarcasm' demonstrate a significant shift from low to high aggressivity, suggesting that, while often mild, they can also be perceived as extremely hostile depending on their intensity and context.

The analysis reveals substantial variation in the manifestation and perceived aggressiveness of hate speech forms across the surveyed countries. While moderate aggressivity is the most common overall, countries like Bulgaria show a disproportionate presence of 'extremely strong' hate speech, whereas the Czech Republic, Germany, and Slovakia frequently report 'no hate' (low aggressivity) instances. The most striking finding is the high number of 'Appeal to violence' cases classified under 'no hate,' which warrants further qualitative research to understand the underlying contextual factors and subjective interpretations by the young Roma operators. This study underscores the complex and context-dependent nature of hate speech identification and aggression assessment, highlighting the need for nuanced, locally informed approaches to monitoring and intervention.

“HATE SPEECH REACHES MILLIONS, ENABLING PERPETRATORS TO INCITE HATRED AND VIOLENCE.”

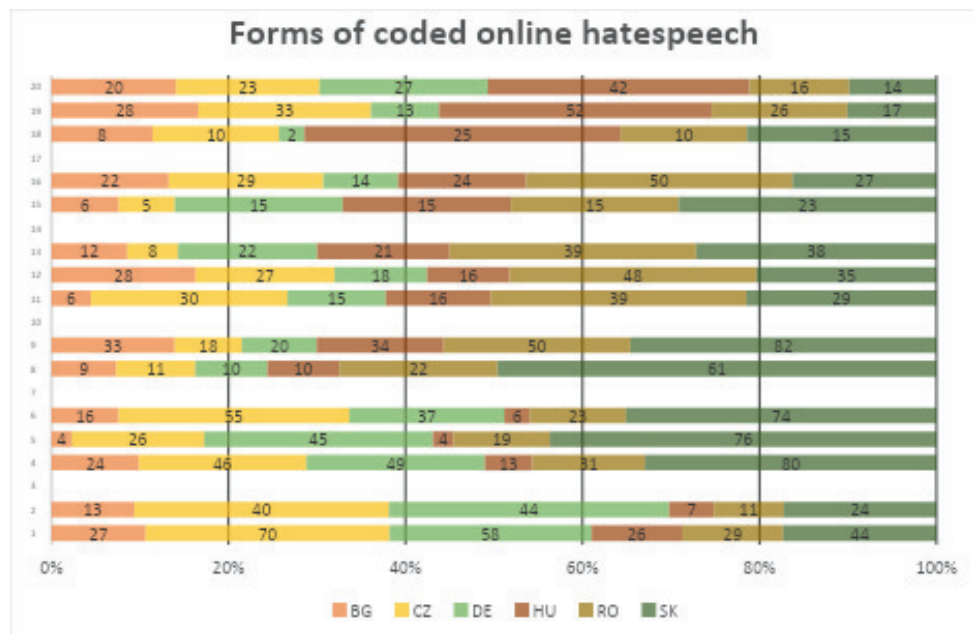
3.2.2. Coded online hate speech (B)

The enduring presence of rhetorical aggression within online environments necessitated the development of methodologies capable of detecting subtle, implicit manifestations of hate speech. These implicit forms are often strategically employed to circumvent content moderation policies, bypass algorithmic detection, and prevent temporary suspensions of online activity. Such tactics facilitate the connotative penetration of public digital spaces, allowing racist and xenophobic discourse to persist and proliferate without immediate repercussions.

To address this, a three-point aggressivity scale was applied and adapted from the aforementioned analytical framework [Siapera, Moreo, and Zhou, 2018]. This scale categorises implicit rhetorical figures by perceived intensity: '1' indicating low aggressivity (interpreted as acceptable or negligible hate speech), '2' denoting moderate aggressivity, and '3' representing extremely strong aggressivity. This refined methodology specifically targeted implicit rhetorical figures operating at the highest level of aggressivity ('extremely strong'), as depicted in relevant data presentations (Figure 9). Indicators for this most intense category included: slang, circumlocutions, irony, ambiguity, pseudo-scientific references, references to white supremacist ideologies, whataboutery, narratives of elsewhere, bogus statistics and metonyms.

This approach introduces a novel element by integrating personal assessments from Roma youth, who are directly affected by such subtle forms of hate speech and experience their variable impact, alongside providing concrete examples of the observed situations.

Figure 9. Forms of coded online antigypsyist hate speech



Source: Author's own contribution

This analysis systematically examines the prevalence and perceived aggressiveness of implicit forms of hate speech in online content across the six European countries: Bulgaria (BG), the Czech Republic (CZ), Germany (DE), Hungary (HU), Romania (RO), and Slovakia (SK). The assessment utilises a three-point aggressivity scale, categorising instances as '1 no hate' (low aggressivity/acceptable), '2 weak' (moderate aggressivity), and '3 extremely strong' (high aggressivity). Implicit hate speech forms under investigation include 'Using slang,' 'Irony,' 'Pseudo-scientific references (genetics),' 'Race (White) supremacies,' and 'Specific rhetorics (metonyms, circumlocutions, ambiguity).' It is important to note that the 'Ambiguity' and 'Others' categories contained no data, thereby precluding their analysis within this framework.

The dataset comprises 2,504 instances of implicit hate speech across all categories and countries. The overall distribution across the aggressivity scale is as follows:

- '1 no hate' (Low Aggressivity): 1,111 instances (44.37%)
- '2 weak' (Moderate Aggressivity): 757 instances (30.23%)
- '3 extremely strong' (High Aggressivity): 636 instances (25.40%)

This distribution indicates that implicit hate speech is most frequently perceived at a low level of aggressivity, suggesting either its widespread and often subtle presence or a higher tolerance/lower detection threshold for these forms. Moderate aggressivity constitutes the second largest category, while highly aggressive implicit hate speech is the least frequent but still a substantial concern.

Analysis by Aggressivity Level

Level 1: '1 no hate' (Low Aggressivity / Acceptable)

This category, representing the lowest perceived level of aggressivity, accounts for the highest proportion of implicit hate speech instances (44.37%).

Dominant Hate Speech Forms: 'Using slang' (254 instances) and 'Pseudo-scientific references (genetics)' (243 instances) are the most prevalent forms classified as 'no hate.' This suggests that colloquial language and ostensibly scientific but biased claims are often perceived as carrying minimal malicious intent or impact. 'Specific rhetorics' (211 instances) also contribute significantly.

Country-Specific Patterns: Slovakia (SK) registers the highest number of 'no hate' instances (298), primarily driven by 'Pseudo-scientific references' (80) and 'Race (White) supremacies' (76). Germany (DE) (233 instances) and the Czech Republic (CZ) (237 instances) also exhibit high frequencies in this category, with 'Using slang' (58 in DE, 70 in CZ) being prominent. Conversely, Hungary (HU) records the fewest 'no hate' instances (62), indicating that implicit hate speech from HU is generally perceived at higher aggressivity levels.

Level 2: '2 weak' (Moderate Aggressivity)

This category, representing moderate aggressivity, accounts for 30.23% of the total instances.

Dominant Hate Speech Forms: 'Irony' (237 instances) and 'Race (White) supremacies' (172 instances) are the most common forms at this moderate aggressivity level. This suggests that while not always overtly aggressive, ironic statements and assertions of racial superiority are frequently recognised as having a discernible, albeit not extreme, negative impact. 'Using slang' (123 instances) and 'Specific rhetorics' (140 instances) also show significant presence.

Country-Specific Patterns: Slovakia (SK) leads this category with 245 instances, predominantly driven by 'Irony' (82) and 'Using slang' (61). Romania (RO) (198 instances) and Hungary (HU) (135 instances) also show high frequencies, with 'Irony' (50 in RO, 34 in HU) and 'Race (White) supremacies' (48 in RO, 16 in HU) being key contributors. Germany (DE) reports the fewest 'weak' aggressivity instances (103).

Level 3: '3 extremely strong' (High Aggressivity)

This category, denoting the highest level of perceived aggressivity, accounts for 25.40% of the total instances.

Dominant Hate Speech Forms: 'Race (White) supremacies' (169 instances) and 'Irony' (166 instances) are the most frequently perceived as 'extremely strong.' This indicates that overt or subtly conveyed notions of racial superiority, and certain forms of irony, are recognized as highly impactful and harmful. 'Specific rhetorics' (142 instances) also demonstrate a substantial presence in this category.

Country-Specific Patterns: Hungary (HU) records the highest number of 'extremely strong' instances (158), largely driven by 'Race (White) supremacies' (52) and 'Specific rhetorics' (42). Romania (RO) (117 instances) and Bulgaria (BG) (114 instances) also contribute significantly, with 'Irony' (50 in RO) and 'Race (White) supremacies' (28 in BG) being prominent. Germany (DE) reports the fewest 'extremely strong' instances (81).

The aggressivity profiles vary significantly across the six nations:

- **Slovakia (SK):** Exhibits the highest proportion of implicit hate speech in the 'no hate' (low aggressivity) category (47.7%), suggesting a higher volume of subtly aggressive content or a tendency for a more lenient classification. However, SK also has a notable presence in the 'weak' category (32.4%).
- **Czech Republic (CZ) and Germany (DE):** Both show a similar pattern to SK, with a majority of implicit hate speech instances falling into the 'no hate' category (48.0% and 56.3%, respectively). This indicates that in these countries, implicit forms are frequently perceived as less aggressive.
- **Hungary (HU):** Demonstrates a distinct profile with a relatively low proportion of 'no hate' (14.2%) and a dominant share in 'weak' (31.1%) and 'extremely strong' (36.5%) categories. This suggests that implicit hate speech in HU is generally perceived as more aggressive than in other countries.
- **Romania (RO):** Displays a relatively balanced distribution, with a significant presence across all three aggressivity levels, though 'weak' (30.7%) and 'no hate' (45.3%) are more prominent than 'extremely strong' (24.0%).
- **Bulgaria (BG):** Shows a considerable proportion of 'extremely strong' aggressivity (36.5%), suggesting that when implicit hate speech is identified, it is often perceived as highly impactful.

Comparative Analysis Across Hate Speech Forms:

Using slang: Primarily classified as 'no hate' (49.7%), suggesting that colloquial language, even when aggressive, is often not perceived as strongly malicious. However, it still contributes to 'weak' (24.1%) and 'extremely strong' (11.7%) aggressivity, indicating context-dependency.

Irony: Exhibits a notable shift from 'no hate' (17.5%) to 'weak' (43.7%) and 'extremely strong' (30.6%) aggressivity. This highlights that while irony can be subtly aggressive, it is frequently perceived as moderately or highly impactful.

Pseudo-scientific references:

Predominantly found in the 'no hate' category (49.8%), implying that these forms, often cloaked in scientific language, are frequently not identified as overtly aggressive despite their underlying harmful intent.

Specific rhetorics:

Displays a balanced distribution across 'no hate' (45.8%), 'weak' (30.4%), and 'extremely strong' (30.8%) aggressivity. This suggests that these complex linguistic tactics are effective in conveying varying levels of aggression.

Race (White) supremacies:

While present in the 'no hate' category (34.7%), this form shows a substantial presence in both 'weak' (34.3%) and 'extremely strong' (33.7%) aggressivity levels. This indicates that expressions of racial superiority are consistently recognised as aggressive, even if their intensity varies.

The analysis of implicit hate speech reveals a complex landscape of manifestations and perceptions across the six European countries. While low aggressivity is the most common classification, significant variation exists, particularly for 'Irony' and 'Race (White) supremacies,' which frequently escalate to higher levels of aggressivity. Countries like Hungary and Bulgaria show a greater tendency to perceive implicit hate speech as highly aggressive, whereas Slovakia, the Czech Republic, and Germany tend to classify a larger proportion as low in aggressivity. The high frequency of 'Pseudo-scientific references' and 'Using slang' in the 'no hate' category suggests a potential gap in the recognition of the insidious nature of these forms, particularly when their aggressive intent is subtle. This underscores the need for enhanced awareness and critical literacy regarding implicit rhetorical figures, as well as the importance of context-sensitive training for individuals involved in identifying and assessing hate speech. The absence of data for 'Ambiguity' remains a limitation, preventing a full understanding of its role in implicit hate speech.

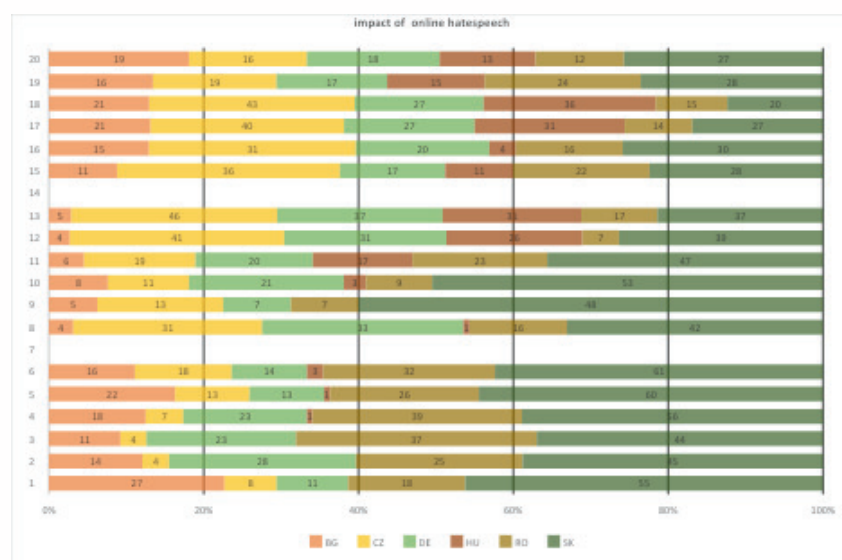


3.2.3. impact of online hate speech

In the context of evaluating the impact of hate speech on victims and online users, the most salient consequences of such manifestations, derived from the monitoring grid analysis, are visually represented (in Figure 10). The observed levels of hate speech are categorised along a three-point aggressivity scale. The initial two levels, designated as 'no hate' and 'weak,' primarily represent what can be termed "frozen hate speech." This refers to relatively static, explicit, or less actively evolving forms of prejudice that, while present, may not be perceived as dynamically shaping or transforming broader discourse.

Conversely, the third and highest level of aggressivity, 'extremely strong,' is indicative of a "motile, transformative antigypsyist repertoire," a concept elaborated by Lentin (2016) in her analysis of racism in public discourse. Lentin's work emphasises how racism adapts and becomes more insidious, often operating through everyday practices and diffuse mechanisms rather than solely overt, extreme expressions. A "motile, transformative antigypsyist repertoire" thus signifies a dynamic and adaptive form of anti-Roma racism that actively reshapes public perception and reinforces prejudice. Such hate speech, while sometimes overtly aggressive, can also manifest subtly, making it difficult to challenge in ostensibly "post-racial" contexts. The "extremely strong" categorisation on our scale specifically captures instances where this dynamic, adaptive, and actively harmful form of anti-Roma racism is at play, directly contributing to adverse effects experienced by victims and fostering an online environment conducive to prejudice and marginalisation. This conceptualisation moves beyond merely cataloguing static expressions of hatred to understanding how such discourse actively influences and perpetuates discrimination.

Figure 10. Transformative antigypsyist online repertoire



Source: Author's own contribution

This data presents a comparative analysis of six key manifestations of hate speech consequences, categorised by a three-point aggressivity scale across the 6 countries. The aggressivity scale comprises: '1 no hate' (interpreted as low aggressivity or acceptable), '2 weak' (moderate aggressivity), and '3 extremely strong' (high aggressivity). The analysed manifestations include: 'Entrench polarisation,' 'Reinforce stereotypes,' 'Spread myths,' 'Spread disinformation,' 'Justify exclusion, stigmatisation and inferiorisation,' and 'Reinforce exclusivist notions of national belonging and identity.' The 'Others' category contained no data and is therefore excluded from this analysis.

The dataset comprises 2,329 recorded instances. The overall distribution across the aggressivity levels is remarkably uniform: 'no hate' accounts for 33.36% (777 instances), 'weak' for 32.85% (765 instances), and 'extremely strong' for 33.79% (787 instances). This even distribution suggests that the consequences of hate speech manifest across a full spectrum of perceived aggressivity, from subtly acceptable to overtly severe.

Analysis by Aggressivity Level

Level 1: '1 no hate' (Low Aggressivity / Acceptable)

This category, representing the lowest perceived aggressivity, comprises 777 instances.

Dominant Manifestations: 'Spread disinformation' and 'Reinforce exclusivist notions of national belonging and identity' are the most frequent manifestations within this category (both 144 instances). This indicates that the dissemination of misinformation and the promotion of exclusionary nationalistic ideas are often perceived as having low or acceptable levels of aggressiveness in the online environment. 'Justify exclusion, stigmatisation and inferiorisation' (135 instances) also features prominently.

Country-Specific Patterns: Slovakia (SK) contributes the most 'no hate' instances (321), notably leading in all manifestations at this level (e.g., 'Reinforce exclusivist notions' with 61, 'Justify exclusion' with 60). Conversely, Hungary (HU) registers an exceptionally low count (5 instances), indicating that the consequences of hate speech are rarely perceived as low aggressivity in this country.

Level 2: '2 weak' (Moderate Aggressivity)

This category, denoting moderate aggressivity, accounts for 765 instances.

Dominant Manifestations: 'Reinforce exclusivist notions of national belonging and identity' (173 instances) and 'Justify exclusion, stigmatisation and inferiorisation' (148 instances) remain highly prevalent. This suggests that while these actions are perceived as moderately aggressive, they are a consistent feature of hate speech consequences. 'Spread disinformation' (132 instances) and 'Entrench polarisation' (127 instances) are also significant.

Country-Specific Patterns: Slovakia (SK) again leads with the highest number of 'weak' instances (266), particularly in 'Reinforce stereotypes' (48) and 'Spread myths' (53). Czech Republic (CZ) (161 instances) and Germany (DE) (149 instances) also show substantial contributions. Bulgaria (BG) reports the lowest count in this category (32 instances), indicating a lower perception of moderate aggressivity.

Level 3: 'extremely strong' (High Aggressivity)

This category, representing the highest perceived aggressivity, comprises 787 instances.

Dominant Manifestations: 'Spread disinformation' (162 instances) and 'Spread myths' (160 instances) are the most frequent manifestations classified as 'extremely strong.' This highlights that the deliberate dissemination of false information and the propagation of harmful narratives are often recognised as highly aggressive and impactful. 'Entrench polarisation' (125 instances) and 'Justify exclusion, stigmatisation and inferiorisation' (119 instances) also contribute significantly.

Country-Specific Patterns: Czech Republic (CZ) records the highest number of 'extremely strong' instances (185), predominantly driven by 'Spread disinformation' (43) and 'Spread myths' (40). Slovakia (SK) and Germany (DE) also show high contributions (160 and 126 instances, respectively). Hungary (HU) (110 instances) shows a significant increase at this level compared to its lower-aggressivity categories, suggesting a tendency to perceive hate speech consequences as severely aggressive when identified.

The distribution of perceived aggressivity varies considerably across the countries:

- **Slovakia (SK):** Consistently exhibits the highest number of instances across all three aggressivity levels. This suggests a high overall volume of hate speech consequences being identified in SK, with a relatively even distribution of perceived aggressivity.
- **Hungary (HU):** Displays a highly skewed profile, with minimal instances in 'no hate' (5) and 'weak' (78), but a substantial increase in 'extremely strong' (110). This implies that when consequences of hate speech are reported in Hungary, they are predominantly perceived as highly aggressive.
- **Czech Republic (CZ):** Shows a similar, though less extreme, pattern to Hungary, with low 'no hate' instances (54) but significant contributions to 'weak' (161) and leading in 'extremely strong' (185) instances. This indicates a general tendency to perceive hate speech consequences as more aggressive in CZ.
- **Bulgaria (BG) and Romania (RO):** Both countries demonstrate a higher proportion of 'no hate' instances (108 and 177, respectively) and 'extremely strong' instances (103 and 103, respectively), but comparatively lower contributions to the 'weak' category (32 and 79, respectively). This suggests a polarised perception, where hate speech consequences are either deemed low aggressivity or extremely strong, with less in the moderate range.
- **Germany (DE):** Presents a relatively balanced distribution across the three aggressivity levels (112 'no hate', 149 'weak', 126 'extremely strong'), suggesting a more nuanced and less polarised assessment of hate speech consequences.

Comparative Analysis Across Manifestations

'Spread disinformation' and 'Reinforce exclusivist notions': These manifestations are consistently among the most frequent across all three aggressivity levels. This underscores their fundamental role in hate speech, operating through both subtle and overt means to shape perceptions and attitudes.

'Justify exclusion, stigmatisation and inferiorisation': Shows a robust presence in 'no hate' and 'weak' categories, but its prominence slightly decreases in the 'extremely strong' category, potentially indicating that overt justification is less frequent in the most aggressive forms, which might instead rely on more direct attacks.

'Reinforce stereotypes,' 'Entrench polarisation,' and 'Spread myths': These manifestations also show a relatively even distribution across aggressivity levels, highlighting their pervasive nature in contributing to both subtle and overt hate speech impacts.

The analysis reveals a nuanced landscape of hate speech consequences, characterised by a near-equal distribution across low, moderate, and high aggressivity levels. While 'Spread disinformation' and 'Reinforce exclusivist notions' are pervasive across all intensities, country-specific patterns highlight distinct perceptual tendencies. Hungary and the Czech Republic demonstrate a propensity to categorise hate speech consequences at higher aggressivity levels, whereas Slovakia frequently identifies consequences across the entire spectrum. Bulgaria and Romania show a more polarised assessment. These findings emphasise the context-dependent nature of aggressivity perception and the complex ways in which hate speech manifests its impact, from subtle reinforcement of prejudice to overtly harmful actions.

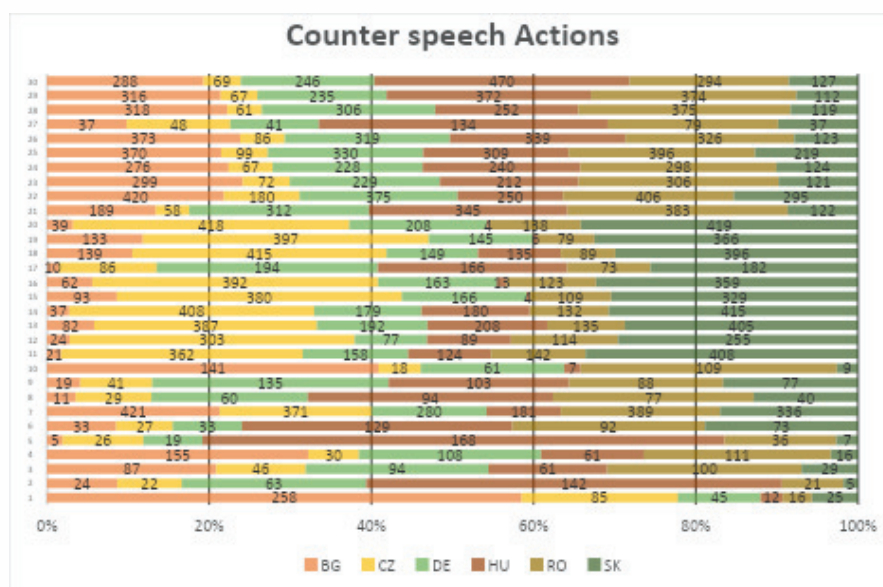
3.3 counteracting

Adapted from Smith's (2018) online counter-speech discursive strategies, this item presents a comparative analysis of individual responses to hate speech in online environments. The data is structured around a three-point scale indicating the reported action or stance: '1 Yes' (active engagement or affirmation of an action), '2 Not sure' (uncertainty or hesitation), and '3 No' (non-engagement or refusal of an action). The response strategies developed include: 'Engaging sincerely,' 'Ignoring it,' 'Exposing the author's tactic/strategy,' 'Critiquing arguments and behaviors,' 'Mocking and making fun of it (or author),' 'Reciprocating in kind,' 'Reporting it to the platform's system,' 'Asking for legal help (civil society),' 'Reporting to the authorities (institutional),' and 'Others.'

The dataset (Figure 11) comprises 28,956 recorded instances. The overall distribution across the response scale reveals a significant prevalence of non-engagement or uncertainty:

- '1 Yes' (Active Engagement): 5,161 instances (17.82%)
- '2 Not sure' (Uncertainty/Hesitation): 11,416 instances (39.43%)
- '3 No' (Non-engagement/Refusal): 12,379 instances (42.75%)

Figure 11. An overview of the counter online antigypsyism actions



Source: Author's own contribution

This distribution indicates a dominant tendency among respondents to either refrain from specific actions or express uncertainty regarding their response when confronted with hate speech. Active engagement, while present, represents a minority of reported instances.



Analysis by Response Scale

'1 Yes' (Active Engagement)

This category reflects affirmative actions taken against hate speech, comprising 5,161 instances.

Dominant Strategies: 'Reporting it to the platform's system' is overwhelmingly the most frequent 'Yes' response, accounting for 1,978 instances (38.3% of all 'Yes' responses). This suggests a strong reliance on platform-level moderation mechanisms. 'Reporting to the authorities (institutional)' (463 instances) and 'Critiquing arguments and behaviours' (481 instances) are also notable active responses.

Country-Specific Patterns: Romania (RO) and Bulgaria (BG) have the highest absolute numbers of 'Yes' responses (1,039 and 1,154, respectively), particularly excelling in 'Reporting to platform' (RO: 389, BG: 421). In contrast, the Czech Republic (CZ) and Slovakia (SK) exhibit the lowest 'Yes' totals (605 and 617 instances, respectively), indicating less frequent active engagement in these regions.

'2 Not sure' (Uncertainty/Hesitation)

This category highlights indecision or a lack of clarity regarding a suitable response, accounting for 11,416 instances, making it the second most common overall.

Dominant Strategies: 'Exposing the author's tactic/strategy' (1,409 instances), 'Critiquing arguments and behaviours' (1,351 instances), and 'Asking for legal help (civil society)' (1,323 instances) are the strategies respondents are most unsure about. This suggests a perceived complexity or potential risks associated with these more involved responses.

Country-Specific Patterns: The Czech Republic (CZ) and Slovakia (SK) report exceptionally high numbers of 'Not sure' responses (3,548 and 3,534 instances, respectively). This high uncertainty is broadly distributed across most strategies in these countries, notably in 'Engaging sincerely' (CZ: 362, SK: 408) and 'Critiquing arguments' (CZ: 408, SK: 415). Hungary (HU) shows the fewest 'Not sure' responses (929), indicating a clearer stance among its respondents.

'3 No' (Non-engagement/Refusal)

This category represents a decision not to undertake a specific action, comprising 12,379 instances, the most frequent overall.

Dominant Strategies: 'Ignoring it' is the most common 'No' response (1,926 instances), closely followed by 'Mocking and making fun of it (or author)' (1,723 instances) and 'Reciprocating in kind' (1,566 instances). This suggests a widespread disinclination to engage with hate speech through direct confrontation or retaliatory methods, or a preference for non-engagement.

Country-Specific Patterns: Romania (RO) and Bulgaria (BG) record the highest total 'No' responses (2,743 and 2,548 instances, respectively), particularly for 'Ignoring it' (RO: 406, BG: 420). Conversely, Slovakia (SK) shows a comparatively lower total (1,272 instances) in this category, indicating a less pronounced tendency to refuse all engagement.

Analysis by Response Strategy

Reporting (Platform/Authorities): These strategies consistently exhibit a polarised distribution. 'Reporting to the platform's system' is the most frequent 'Yes' action (1,978 instances) but has relatively few 'Not sure' (711) or 'No' (376) responses. Similarly, 'Reporting to the authorities' (463 'Yes' instances) is a clear choice, though it also sees substantial 'No' (1,476) and 'Not sure' (1,126) responses, implying barriers or reluctance for formal reporting.

Ignoring it: This is predominantly a 'No' response (1,926 instances), indicating that many respondents actively choose not to ignore hate speech, possibly preferring some form of action or at least not consciously disengaging. However, 862 'Not sure' instances also exist, suggesting ambivalence. Notably, only 277 instances affirmed ignoring it, indicating it's not the primary affirmative action.

Engaging Sincerely: This strategy garners a substantial number of 'Not sure' (1,215 instances) and 'No' (1,409 instances) responses, significantly outweighing 'Yes' (441 instances). This suggests a general hesitation or reluctance to engage directly and constructively with authors of hate speech.

Exposing Tactics/Critiquing Arguments: Both strategies show a high degree of 'Not sure' (1,409 and 1,351 instances respectively) and 'No' (1,239 and 1,233 instances respectively) responses, compared to their 'Yes' totals (417 and 481 respectively). This implies that while these analytical responses are considered, there's significant uncertainty or reluctance to execute them.

Mocking/Reciprocating in Kind: These more aggressive or retaliatory responses are overwhelmingly situated in the 'No' category (1,723 and 1,566 instances respectively), far exceeding 'Yes' (261 and 387 instances) and 'Not sure' (1,081 and 1,112 instances). This suggests a strong dispreference for responding to hate speech with similar negativity.

Country-Specific Response Profiles

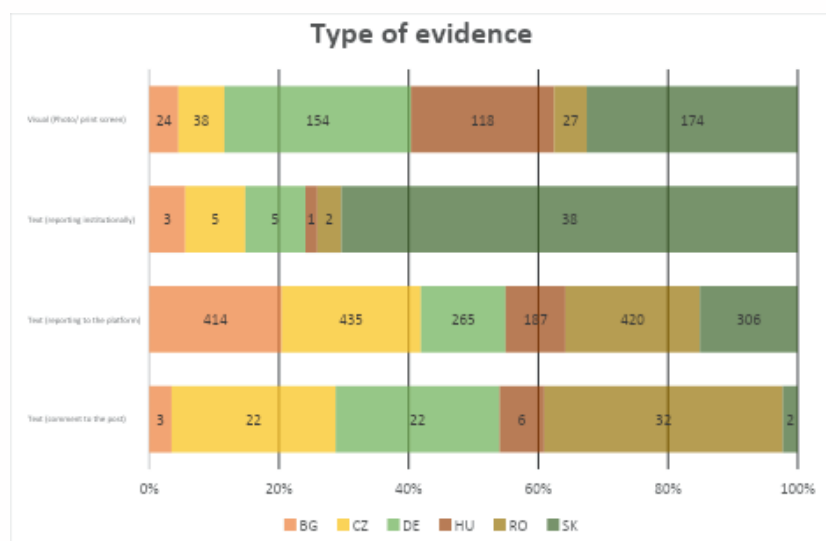
- **Bulgaria (BG) & Romania (RO):** Exhibit a common pattern of high 'Yes' in 'Reporting to platform,' coupled with very high 'No' responses across many strategies, particularly 'Ignoring it,' 'Mocking,' and 'Reciprocating.' This indicates a preference for formal reporting, combined with a general refusal to engage directly or in retaliation.
- **Czech Republic (CZ) & Slovakia (SK):** Are characterised by exceptionally high 'Not sure' responses across nearly all strategies, suggesting significant uncertainty in how to effectively address hate speech. While their 'Yes' responses are low, their 'No' responses are also comparatively lower than BG/RO for many direct engagement strategies, possibly because high 'Not sure' counts absorb potential 'No' responses.
- **Germany (DE):** Shows a balanced but overall high 'No' response across many strategies, with moderate 'Yes' and 'Not sure' counts. Its highest 'Yes' is for 'Reporting to platform' (280) and 'Reporting to authorities' (135), indicating a reliance on institutional mechanisms.
- **Hungary (HU):** Displays a lower overall 'Not sure' count, indicating a clearer stance. It has a high 'Yes' for 'Mocking' (168) and 'Reciprocating' (129) compared to other countries, suggesting a greater willingness to confront or retaliate, alongside substantial 'No' responses to many other strategies.

The analysis reveals a complex landscape of individual responses to online hate speech. Overall, there is a strong inclination towards non-engagement or uncertainty regarding appropriate counter-strategies. Formal reporting to platforms is the most favoured active response, highlighting a reliance on external moderation. Strategies involving direct engagement, critique, or legal action are often met with uncertainty, while retaliatory or mocking responses are largely refused.

Significant country-level differences exist, with the Czech Republic and Slovakia exhibiting widespread uncertainty in their responses, while Bulgaria and Romania tend to combine active reporting with a refusal to engage directly. Hungary presents a unique profile, with a higher willingness to mock and reciprocate. This heterogeneity underscores the influence of local contexts, cultural norms, and perceived efficacy of different strategies. The high proportion of 'Not sure' responses across several countries underscores a critical need for education and clear guidelines to empower individuals to address hate speech effectively, reduce ambivalence, and foster more confident, constructive responses.

The final section of the quantitative component of the Monitoring Tool (MT) documents an inventory of the most frequently utilised forms of evidence used by monitors when initiating a response or counteraction to observed instances of online anti-Gypsyism. This inventory provides empirical insight into the types of data deemed relevant and actionable by field monitors, thereby informing intervention and educational strategies. The systematic compilation of these evidentiary categories is expected to serve as a foundational resource for the development of future training programs, specifically designed to enhance observers' capacity to identify, critically analyse, and mitigate anti-gypsyism in online environments.

Figure 12. Forms of evidence for initiating counter actions against antigypsyism



Source: Author's own contribution

The results (Figure 12) present a comparative analysis of the types of textual evidence employed by monitors when responding to instances of online hate speech. The objective is to identify the most frequently utilised forms of textual evidence for counter-action and to highlight country-specific variations in these practices. The evidence types examined are: 'Text (comment to the post),' 'Text (reporting to the platform),' and 'Text (reporting institutionally).' The 'I no hate' label for this dataset indicates that these evidence types were specifically gathered in contexts where the encountered hate speech was perceived to be of low aggressivity.

The dataset comprises a total of 2,168 instances of textual evidence utilised by monitors.

Overall Distribution of Textual Evidence Types

The analysis reveals a highly skewed distribution among the three categories of textual evidence:

- **Text (reporting to the platform):** 2,027 instances (93.5%)
- **Text (comment to the post):** 87 instances (4.0%)
- **Text (reporting institutionally):** 54 instances (2.5%)

This overwhelming dominance of 'Text (reporting to the platform)' signifies that direct reporting to online platforms is by far the most prevalent counter-action strategy employed by monitors when confronting low-aggressivity hate speech. Other forms of textual evidence, such as commenting or institutional reporting, are utilised to a significantly lesser extent.

Analysis by Evidence Type and Country

Text (comment to the post)

This category, which represents textual comments made directly on the post containing hate speech, accounts for a minority of the evidence.

Prevalence: With only 87 instances, commenting directly on the post is a rarely chosen method of evidence collection for counter-action related to low-aggressivity hate speech.

Country-Specific Patterns: Romania (RO) utilises this method most frequently (32 instances), followed by the Czech Republic (CZ) and Germany (DE) (both 22 instances). Bulgaria (BG) (3 instances) and Slovakia (SK) (2 instances) show minimal engagement with this form of evidence. Hungary (HU) (6 instances) also exhibits low usage. This suggests that while some monitors engage in direct commentary, it is not a widespread or primary strategy for documenting low-aggressivity hate speech.

Text (reporting to the platform)

This category, representing textual evidence submitted when formally reporting to the online platform, is the overwhelmingly dominant form.

Prevalence: Accounting for 93.5% of all textual evidence (2,027 instances), reporting to the platform is the primary mechanism through which monitors engage in counter-action against low-aggressivity hate speech.

Country-Specific Patterns: Czech Republic (CZ) (435 instances), Romania (RO) (420 instances), and Bulgaria (BG) (414 instances) are the leading countries in utilising this form of evidence, demonstrating a strong reliance on platform-based reporting mechanisms. Slovenia (SK) (306 instances) and Germany (DE) (265 instances) also show substantial activity. Hungary (HU) (187 instances) utilises this method least, though it still accounts for the vast majority of its textual evidence collection. The high numbers across all countries underscore the universal role of platform reporting in responding to hate speech.

Text (reporting institutionally)

This category refers to textual evidence submitted for formal reporting to external institutions (e.g., government agencies, NGOs, legal bodies).

Prevalence: With only 54 instances, institutional reporting is the least common form of textual evidence used for counter-action against low-aggressivity hate speech.

Country-Specific Patterns: Slovakia (SK) stands out significantly in this category, with 38 instances, representing 70.4% of all institutional reporting. This is a disproportionately high contribution compared to other countries (BG: 3, CZ: 5, DE: 5, HU: 1, RO: 2), suggesting a distinct emphasis on formal institutional engagement in Slovakia even for hate speech deemed of low aggressivity. Most other countries show negligible use of this method.

Country-Specific Profiles of Evidence Utilisation

- **Bulgaria (BG), Czech Republic (CZ), Germany (DE), Romania (RO), and Hungary (HU):** These countries predominantly rely on 'Text (reporting to the platform)' as their primary method of evidence submission, accounting for 98.6%, 94.2%, 90.8%, 92.5%, and 96.4% of their respective totals. This uniformity highlights a shared strategy centred on direct platform engagement. Institutional reporting and commenting on posts are marginal in these contexts.
- **Slovakia (SK):** While still heavily reliant on 'Text (reporting to the platform)' (88.4% of its total), Slovakia uniquely demonstrates a significantly higher propensity for 'Text (reporting institutionally)' (11.0% of its total), indicating a more diversified approach that integrates formal external institutional channels even for low-aggressivity hate speech.

The analysis of textual evidence utilised for counter-action against low-aggressivity online hate speech reveals a strong and pervasive reliance on direct reporting to online platforms across all six European countries. This suggests that monitors primarily view platforms as the frontline for addressing such content, reflecting either the perceived effectiveness of platform moderation or a lack of viable, accessible alternatives.

While other forms of textual evidence, such as commenting on posts, are infrequently used, Slovakia stands out by incorporating institutional reporting into its counteraction strategy, even for hate speech perceived as less aggressive. This divergence in practice may reflect differing legal frameworks, the perceived responsiveness of national institutions, or varying organisational mandates of the monitors in Slovakia compared to other nations. The overall low utilisation of institutional reporting indicates that, for low-aggressivity hate speech, formal reporting beyond the platform level is generally not prioritised. These findings underscore the need to further investigate the perceived efficacy of various counter-action strategies and the contextual factors influencing their adoption by online hate speech monitors.

3.4. qualitative component (the narrative)

To obtain a more comprehensive understanding of the challenges Roma individuals face in online environments, the second phase of this research involved extensive data collection by trained monitors. This collection documented reported cases of anti-gypsyist hate speech, the monitors' personal experiences with counter-action, and their direct, evidence-based interventions involving both online platforms and traditional media to combat this phenomenon. Adhering to a standardised data collection structure and methodology, each quarterly report incorporated a "Narrative part" of the selected study cases. This section comprises 1-3 study cases/country, carefully selected from the total reported by the Monitoring Tool (MT), based on the monitors' discretion and national group consensus. These selected cases were meticulously monitored and documented (including links, screenshots, or photos), with personal observations (emotional impact), contextual explanations, and a description of the counteractions undertaken by the monitors to mitigate the content's impact.

A database consisted of 54 case studies of online antigypsyism collected from Bulgaria (BG), the Czech Republic (CZ), Germany (DE), Hungary (HU), Romania (RO), and Slovakia (SK). The cases were analysed using a qualitative content analysis, with coding adapted from Molnar (2021). The analysis systematically categorises the observed hate speech manifestations using a predefined coding scheme and evaluates the outcomes of reporting such content to online platforms or authorities. The case studies were qualitatively analysed to identify the presence of six distinct categories of anti-Roma rhetoric, referred to as "codes":

1. **Deviance and criminality (-):** Attribution of criminal behaviour to Roma.
2. **Insults, disgust, violence and punitiveness (-):** Use of derogatory language, dehumanisation, or advocacy for harm against Roma.
3. **Cultural essentialism (-):** Blaming Roma ethnicity or culture for societal problems.
4. **Group differentiation (-):** Emphasising "us vs them" narratives to separate Roma from the majority.
5. **Fear toward the Roma (-):** Expressing constant apprehension or alarm regarding Roma individuals.
6. **Culture and history of the Roma (=):** Discussions focusing on Roma culture or traditions (noted as potentially neutral, but observed in a negative context in this dataset).

Additionally, the reporting outcome for each case was categorised as:

- **0:** No report mentioned.
- **1-0:** Reported, but no action taken by the platform/authority.
- **1-1:** Reported, and an answer was received (though not necessarily action).
- **1-1-1:** Reported, an answer was received, and action was taken (e.g., content removed).

Table 1. An overview of the study cases

Country	Cases (number)	Type					
		website	Facebook	Instagram	X	You Tube	Tik Tok
	Total						
Bulgaria	11	5	6	0	0	0	0
Czech R.	11	0	1	1	1	1	7
Germany	7	0	1	0	4	2	0
Hungary	5	2	1	0	0	0	2
Romania	10	1	7	1	0	0	1
Slovakia	10	3	7	0	0	0	0
	54	11	23	2	5	3	10

Source: Author's own contribution

Analysis by Antigypsyism Code (Themes of Hate Speech)

A total of 75 coded manifestations were identified across the 54 case studies, of which 15 reported cases were from traditional media sources.

Code 4: Group Differentiation (29 instances, 38.7% of all codes)

This code is the most prevalent form of antigypsyism across the dataset, highlighting a pervasive "us vs. them" narrative.

Manifestations: Explicit calls for exclusion ("Gypsies do not call" in job/housing ads), blaming Roma for societal issues by contrasting them with the "Bulgarian people" or "Hungarians," and highlighting ethnic origin in criminal reports to differentiate from the majority population.

Country-Specific Observations: This code is overwhelmingly dominant in Slovakia (SK), featuring in all 10 reported cases. Examples include housing and sales listings explicitly excluding Roma ("Speculators and Gypsies do not write, do not call") and job advertisements with phrases like "Roma do not call." Czech Republic (CZ) (5 cases), Bulgaria (BG) (4 cases), and Romania (RO) (4 cases) also show significant use of this code, often through political rhetoric or online comments emphasizing ethnic origin to create division.

Code 2: Insults, Disgust, Violence, and Punitiveness (15 instances, 20% of all codes)

This category reflects the use of derogatory language, dehumanisation, and direct or implied threats.

Manifestations: Direct slurs such as "mangal" and "Zigeuner," dehumanizing terms like "embryos," "two-legged rats," "locusts," "parasites," and calls for violence (e.g., "50 thousand gypsies/Roma must die," "shot").

Country-Specific Observations: This code is notably present in Bulgaria (BG) (5 cases), with examples like "dirty Gypsies" and explicit calls for murder. Germany (DE) (3 cases) utilises animal metaphors ("locusts," "parasites") for dehumanisation. Czech Republic (CZ) (3 cases) uses mocking language and "brown face" imagery, often with AI-generated content.

Code 1: Deviance and Criminality (14 instances, 18.7% of all codes)

This code attributes criminal behavior to Roma individuals or the entire community.

Manifestations: Claims of Roma stealing, "Roma crime" narratives, linking Roma identity to "lawless" behavior, and implying criminality through AI-generated content.

Country-Specific Observations: This code is prominent in Bulgaria (BG) (5 cases), particularly from the "News24Sofia.eu" pseudo-media outlet, which sensationalises crime stories and links them to Roma ethnicity. Germany (DE) (3 cases) features claims of theft and lawlessness impacting local residents. Czech Republic (CZ) (2 cases) uses AI to portray Roma as thieves (e.g., "this is my Audi that I stole").

Code 3: Cultural Essentialism (9 instances, 12% of all codes)

This code attributes negative characteristics or behaviors to Roma culture or ethnicity itself, implying they are inherent and unchangeable.

Manifestations: Blaming Roma culture for poverty, lack of education, or perceived negative behaviors (e.g., "irresponsibly has children," "practicing customs they learned throughout their lives" that disturb others).

Country-Specific Observations: Romania (RO) (3 cases) attributes negative behaviours (e.g., being "loud, expressive") to Roma customs. Hungary (HU) (2 cases) frames demographic trends as problematic due to Roma "subcultures." Bulgaria (BG) (3 cases) uses "mangal" and "Gypsy" slurs within broader dehumanising narratives.

Code 5: Fear Toward the Roma (7 instances, 9.3% of all codes)

This code expresses or incites apprehension or alarm regarding the presence or actions of Roma.

Manifestations: Housing preferences explicitly excluding Roma due to fear, generalized anxiety about Roma "invading" public spaces, and political rhetoric that aims to evoke fear for electoral gain.

Country-Specific Observations: This code appears in Bulgaria (BG) (3 cases), the Czech Republic (CZ) (2 cases), and Germany (DE) (2 cases). Examples include fears about "too many gypsies" stealing (CZ) and claims of "Zigeuner terrorising" neighbourhoods (DE).

Code 6: Culture and History of the Roma (1 instance, 1.3% of all codes)

This code focuses on discussions of Roma culture or traditions.

Manifestations: Only one instance was coded under this category in the **Czech Republic (CZ)**. It involves a TikTok video using "black/brown face" and Roma slang in a mocking, caricatured manner alongside an older song, implying a distorted or stereotypical engagement with cultural elements under the guise of humour. The "equal sign" in the code description suggests it *could* be neutral, but in this context, it clearly propagates negative stereotypes.

Analysis by Evidence Type and Country

Slovakia (SK) - 10 Cases

- **Dominant Theme:** Exclusively (10/10 cases) features **Group Differentiation (Code 4)**, particularly in market-related contexts (housing, sales, job offers) with explicit exclusionary language ("Gypsies do not call," "Roma do not call").
- **Platforms:** Primarily Facebook (Marketplace, private accounts, local groups) and Instagram.
- **Reporting Outcome:** All 10 cases were reported, but **no action was taken** by platforms (10x 1-0). This highlights a severe lack of platform moderation for blatant discrimination in SK.

Bulgaria (BG) - 11 Cases

- **Dominant Themes:** A mix of **Deviance and Criminality (Code 1)**, **Insults, Disgust, Violence (Code 2)**, **Cultural Essentialism (Code 3)**, and **Fear Toward the Roma (Code 5)**. Pseudo-media outlet "News24Sofia.eu" is heavily implicated in 5 cases.
- **Platforms:** YouTube (News24Sofia.eu channel), Facebook (personal accounts, public groups), TikTok. Political figures (MPs) also contribute to hate speech.
- **Reporting Outcome:** 7 out of 11 cases were reported, and **no action was taken** (7x 1-0). 4 cases had no reporting mentioned (4x 0). This indicates a high prevalence of various hate speech forms, largely unmoderated.

Czech Republic (CZ) - 11 Cases

- **Dominant Themes:** Strong presence of **Group Differentiation (Code 4)** and **Insults, Disgust, Violence (Code 2)**. A notable trend is the widespread use of **AI-generated content** for mocking and stereotyping Roma.
- **Platforms:** Predominantly TikTok and YouTube, but also Instagram, Facebook (political parties), and X (Twitter).
- **Reporting Outcome:** All 11 cases were reported, and **no action was taken** (11x 1-0). This demonstrates a significant failure of platform moderation to address AI-driven and other forms of antigypsyism in CZ.

Romania (RO) - 10 Cases

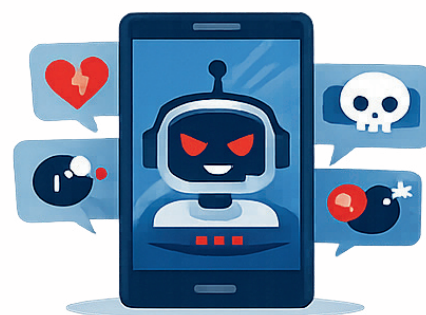
- **Dominant Themes:** Features **Group Differentiation (Code 4)** and **Cultural Essentialism (Code 3)**, often linking Roma behaviour to "customs" or "lack of education." Also includes instances of **Deviance and Criminality (Code 1)** and explicit calls for **Insults, Disgust, Violence (Code 2)**.
- **Platforms:** Facebook (personal accounts, public), TikTok, Reddit, Instagram, news media sites (Adevarul, stiridecluj, PressOne, gandul.ro) and their comments sections.
- **Reporting Outcome:** 6 out of 10 cases were reported. Notably, 2 cases had positive outcomes: one received an answer (1x 1-1), and another led to action (1x 1-1), resulting in a post being deleted. However, 4 cases resulted in no action (4x 1-0), and 4 cases had no reporting mentioned (4x 0). This suggests a slightly more responsive platform environment than other countries, but still with significant challenges.

Germany (DE) - 7 Cases

- **Dominant Themes:** A balanced mix of **Deviance and Criminality (Code 1)**, **Insults, Disgust, Violence (Code 2)**, **Group Differentiation (Code 4)**, and **Fear Toward the Roma (Code 5)**. Cases involve both sensationalist media and coded hate speech.
- **Platforms:** X (Twitter), Facebook, YouTube (public broadcasters, news channels like Bild and WELT).
- **Reporting Outcome:** 5 out of 7 cases were reported with **no action taken** (5x 1-0). 2 cases had no reporting mentioned (2x 0). This highlights issues with content moderation even from established media platforms.

Hungary (HU) - 5 Cases

- **Dominant Themes:** Primarily **Group Differentiation (Code 4)** and **Cultural Essentialism (Code 3)**, often linked to demographic concerns ("Gipsy bomb" myth) or caricaturing Roma. Also includes **Deviance and Criminality (Code 1)**.
- **Platforms:** Public media sites (Kuruc.info), TikTok, Facebook (public figures), Radio/blogs.
- **Reporting Outcome:** 3 out of 5 cases were reported, and **no action was taken** (3x 1-0). 2 cases had no reporting mentioned (2x 0). This indicates persistent unmoderated antigypsyism, especially from political figures.



Cross-Cutting Observations

- **Pervasive Normalisation and Impunity:** The most striking finding is the overwhelming lack of moderation. **Out of 40 reported cases**, only 2 resulted in any positive action (1 deletion, 1 answer received), and the vast majority, **38 cases, saw no action taken by platforms**. This supports the observation that antigypsyism is often normalised, and platforms frequently fail to enforce their own guidelines, fostering a climate of impunity.
- **Social Media as Primary Vector:** Facebook and TikTok are heavily utilised platforms for the dissemination of antigypsyist content, ranging from explicit discriminatory ads to AI-generated mockery. YouTube and X (Twitter) are also significant.
- **Rise of AI-Generated Hate Speech:** Particularly evident in the Czech Republic, AI is leveraged to create highly stereotypical and derogatory content, often dismissed as "parody" but effectively normalising racism.
- **Political Instrumentalisation:** Political parties and public figures across Bulgaria, the Czech Republic, and Hungary are actively involved in generating or amplifying antigypsyist discourse for electoral gain, further legitimising hatred.
- **Diverse Manifestations:** Antigypsyism manifests in a wide array of forms, from explicit slurs and calls for violence to subtle group differentiation, dehumanising metaphors, and the propagation of myths and false narratives.
- **Limited Positive Outcomes from Reporting:** The empirical evidence from these cases strongly suggests that current reporting mechanisms are largely ineffective in combating online antigypsyism.

This content analysis underscores the widespread and multifaceted nature of online antigypsyism across these six European countries. The analysis reveals a disturbing pattern of pervasive hate speech, predominantly characterised by group differentiation and the attribution of criminality and negative stereotypes to Roma communities. A critical finding is the almost universal failure of online platforms to effectively moderate reported antigypsyist content, contributing to its normalisation and reinforcing a climate of impunity. The growing use of AI-generated content for hate speech and the active involvement of political actors further exacerbate this challenge. These findings necessitate urgent and more robust interventions from online platforms, regulatory bodies, and civil society to counteract the pervasive and unmoderated spread of antigypsyism.

4. limits of the study

The study acknowledges several inherent limitations, primarily stemming from its methodological design and the sensitive nature of the subject matter. While analysing a substantial number of quantitative cases (3147), the data collection relied on non-probabilistic sampling and the "subjective judgment" and "personal choice" of young Roma monitors, whose observations were based on their individual news feeds and personal accounts.

This approach, though valuable for capturing lived experiences and emotional impact, limits the generalizability of findings to the entire Roma community and introduces potential selection bias. Furthermore, the assessment of hate speech aggressivity, particularly the perceived impact and intensity, was explicitly based on the "personal perspectives of the participating young individuals."

This deliberate integration of user experience, while ethically acknowledging the monitors' exposure to online hate speech and their "high level of emotions and users' feelings," means that a purely neutral or positivist approach could not be fully applied. Consequently, this introduces potential reliability issues regarding coding consistency, as different monitors might interpret or classify codes and intensity levels differently.

Although a common understanding was sought, language barriers in multi-country data collection, even after aggregation, may have further affected the nuanced interpretation and coding of content. These factors underscore the inherent challenges in quantifying and categorising such a complex and emotionally charged phenomenon.

**“QUANTIFYING
ONLINE
ANTIGYPSYISM
REMAINS
INHERENTLY
CHALLENGING
DUE TO ITS
COMPLEX AND
EMOTIONALLY
CHARGED
NATURE.”**

5. conclusions

Based on a mixed-methods approach involving 30 young Roma monitors and analysing 3147 quantitative cases and 54 qualitative case studies from October 2024 to September 2025, the study "Together against antigypsyism online (TAAO) in Europe," provides clear insights into the prevalence, forms, and impact of online anti-gypsyism, as well as the efficacy of counter-action strategies:

Prevalence and Sentiment of Anti-Gypsyist Content:

- **Dominance of Negative and Neutral Sentiment:** Out of 15,080 topic-related instances, neutral sentiment constituted the largest proportion (44.92%), closely followed by negative mentions (40.58%), with positive mentions being significantly lower (14.49%). This indicates a tendency towards critical or problematic framing of Roma-related topics when sentiment is expressed.
- **Country-Specific Sentiment Biases:** The Czech Republic exhibited the highest proportion of neutral sentiment (64.1% of its total instances), alongside a substantial negative component (27.09%). Conversely, Hungary, Germany and Romania displayed a greater propensity for negative reporting (43.95%, 63.26% and 50%, respectively), with Romania also showing a balanced, albeit neutral and positive-leaning, distribution. Bulgaria registered the lowest positive mentions (2.89%), suggesting a more polarised reporting landscape, as did Slovakia, where 39.03% negative mentions overwhelm the positive ones (8.72%).
- **Content Style Sentiment:** Across 11, 272 instances of various content styles, negative content was most prevalent (48.14%), followed by neutral (36.02%) and positive (15.82%). "Emotional" and 'Funny' content showed the highest negative instances, suggesting that much emotional and humorous content, as well as official communications, are frequently used to convey negative sentiment. "Scientific" content exhibited the highest neutral instances.

Forms and Aggressivity of Online Hate Speech:

- **Explicit Hate Speech:** Of 2716 instances of explicit hate speech, "weak" (moderate aggressivity) was the most frequent category (37.4%), followed by "no hate" (low aggressivity) at 34.9%, and "extremely strong" (high aggressivity) at 27.7%. Notably, "Appeal to violence (including killing)" was surprisingly dominant in the "low aggressivity" category, suggesting potential desensitisation or contextual interpretation by monitors. "Mockery and sarcasm" were prominent in the "extremely strong" category.
- **Implicit Hate Speech:** Among 2504 instances of implicit hate speech, "I no hate" (low aggressivity) accounted for the highest proportion (44.37%), with "Irony" and "Race (White) supremacies" emerging as dominant forms in the "weak" and "extremely strong" categories, respectively. Hungary and Bulgaria exhibited a higher tendency to perceive implicit hate speech as highly aggressive.
- **Prevalent Antigypsyism Codes (Qualitative Analysis):** "Group Differentiation" was the most common theme (38.7% of 75 instances), manifested through "us vs. them" narratives, explicit exclusion, and ethnic blaming in criminal reports, particularly dominant in Slovakia. "Insults, Disgust, Violence, and Punitiveness" (20%) and "Deviance and Criminality" (18.7%) were also significant, often utilising AI-generated content for mockery and stereotyping.

Platform Usage and Visual Content:

- **Primary Online Vectors:** Facebook (Meta) was identified as the most widely utilised social media platform, with TikTok also exhibiting a significant presence. These platforms serve as primary channels for public figures and for the dissemination of online news, particularly in Slovakia, Bulgaria, and Romania.
- **High Visual Content Inclusion:** The study found a general propensity for media articles to include visual elements (77.15% of 709 articles), with Hungary, Slovakia, and Romania showing the highest rates of visual integration.
- **Visual Content Sentiment:** Visuals showed a distribution towards negative sentiment, with negative (43.64%) and neutral (40.55%) predominating over positive (15.79%). Videos (885 instances) and Photos (757 instances) are predominantly used for negative content, reinforcing their role in factual representation, particularly in Germany and Slovakia.

Counter-Action Strategies and Effectiveness:

- **Low Active Engagement:** Out of 28,956 recorded instances of individual responses to hate speech, "Non-engagement/Refusal" (42.75%) and "Uncertainty/Hesitation" (39.43%) were significantly more prevalent than "Active Engagement" (17.82%).
- **Reliance on Formal Reporting:** "Reporting it to the platform's system" was the most frequent active response (38.3% of active engagements), highlighting a strong reliance on platform-level moderation mechanisms.
- **Limited Positive Outcomes:** A critical finding from the qualitative analysis was the overwhelming lack of moderation; only 2 out of 40 reported cases resulted in positive action (one deletion, one answer), with 38 cases seeing no action taken by platforms. This suggests current reporting mechanisms are largely ineffective and platforms frequently fail to enforce their own guidelines, fostering a climate of impunity.

Overarching Cross-Cutting Observations:

- **Pervasive Normalisation and Impunity:** Online antigypsyism is often normalised, 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 normalising racism.
- **Political Instrumentalisation:** Political actors and public figures actively contribute to generating or amplifying antigypsyist discourse for electoral gain.

These findings underscore the complex and multifaceted 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 antigypsyism will require not only robust legal frameworks but also sustained collaboration between regulators, platforms, civil society, and Roma-affected communities.

6. recommendations

6.1 policymakers

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 (only 2 out of 40 reported cases had a positive outcome).

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.

Address AI-Generated Hate Speech:

The study identifies a "rise of AI-generated hate speech" used for mocking and stereotyping, which is often dismissed as "parody" but effectively normalises racism.

Action:

Fund research into and implement regulatory frameworks for AI systems that generate or amplify hate speech, particularly those targeting vulnerable groups such as Roma, ensuring robust bias detection, data governance, and human oversight in line with AI Act principles.

Combat Political Instrumentalization:

Political instrumentalisation of anti-gypsyist discourse for electoral gain was observed across Bulgaria, the Czech Republic, and Hungary, further legitimising 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 favoured active response, its effectiveness is low. Improving and supporting these mechanisms is vital.

Action:

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

6.2 social media platforms

Improve Content Moderation Effectiveness:

The study demonstrates that platforms frequently fail to act on reported antigypsyism 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 necessitate 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.

Enhance AI for Bias Detection and Mitigation:

AI is being leveraged to create highly stereotypical and derogatory content, contributing to the problem. Platforms must ensure their AI systems are part of the solution, not the problem.

Action:

Develop and deploy more sophisticated AI and algorithmic tools to detect and mitigate biased content generation and the algorithmic amplification of anti-Roma narratives. Regularly audit these tools for fairness and accuracy.

Streamline Reporting and Feedback Mechanisms:

"Reporting to the platform's system" is a dominant counter-action, but its perceived ineffectiveness discourages engagement.

Action:

Make reporting tools more user-friendly and provide clear, timely feedback on the status and outcomes of reported instances of anti-Roma hate speech to empower users and build trust.

Collaborate with Roma Organizations:

Roma organisations possess critical insights into the lived experiences and specific manifestations of antigypsyism.

Action:

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

6.3 roma organisations

“ROMA ORGANISATIONS PLAY A CRITICAL ROLE IN DOCUMENTING, CHALLENGING, AND RESISTING ONLINE ANTIGYPSYISM.”

Empower Communities with Digital Literacy:

There is a high prevalence of "not sure" responses to counteractions, and a potential "desensitisation" 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 the methodology captures nuanced forms and country-specific contexts. Collaborate with academic institutions for rigorous analysis.

Intensify Advocacy and Lobbying:

The evidence of pervasive normalisation, ineffective moderation, and political instrumentalisation provides a strong basis for advocacy.

Action:

Utilise 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.

references

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.

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.

European Commission Communication "A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime", 2021, available at: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/extending-eu-crimes-hate-speech-and-hate-crime_en

European Commission Communication (2018). Antigypsyism: increasing its recognition to better understand and address its manifestations, available at: file:///C:/Users/user/Downloads/hlg_conclusions_paper_antigypsyism_final_8F1D1436-B234-0AD8-C4BE97D84A3D9E1A_55652.pdf

European Commission (2022). EU Code of Conduct on Countering Illegal Hate Speech Online. available at: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

European Roma Rights Center (2023). Challenging digital antigypsyism, available at: <https://www.errc.org/reports--submissions/challenging-digital-antigypsyism-albania-serbia-turkey-ukraine>

European Parliament (2022). Digital Services Act, available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065>

European Parliament (2024). AI Act, available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
ERGO Network (2020). The Impact of Covid-19 on Roma Communities in the European Union and the Western Balkans, available at: <https://ergonetwork.org/2020/12/the-impact-of-covid-19-on-roma-communities-in-the-european-union-and-the-western-balkans/>

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>,

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>

Molnar, Lorena (2021) The Law of the Jungle. The Online Hatespeech against the Roma in Romania, *Victims & Offenders*, 16:8, 1108-1129. <https://doi.org/10.1080/15564886.2021.1895391>

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>)

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>

TAAO

TOGETHER AGAINST ANTIGYPSYISM ONLINE

