

HOW TO RESPOND TO HATE SPEECH AND FAKE NEWS AT THE LOCAL LEVEL IN SLOVAKIA

Ondrej Mital¹ - Miroslav Fečko² - Silvia Ručinská³

Abstract

As new media continue to shape the digital information space, the occurrence of disinformation, hate speech, and fake news is evolving rapidly. These issues arise as a consequence of information overload and information disorder catalysed by digital transformation. This phenomenon is recognised globally as a significant challenge for public institutions, democracy, and public policy. In general we address disinformation with two approaches: hard and soft response activities and tools. Both approaches are complementary to each other and together create a multi-response approach to disinformation. This study aims to examine the viability of the hard response approach alone. This approach typically serves as a reactive measure by targeting the source or the creator of disinformation. We explored the dimensions of hate speech, fake news and disinformation, highlighting the growing pressures that necessitate a multi-response approach. A survey was conducted in 2023 to investigate fake news and hate speech, with a particular focus on the perceived viability of response activities and tools among local politicians and communal practitioners in Slovakia. The primary result of this research lies in enhancing the ongoing discourse about the multi-response approach, emphasising the significance of hard response activities and tools, supported by both theoretical desk research and empirical evidence. A further benefit of the study is that it confirmed the feasibility and success of the examined hard responses. These differ for disinformation, as well as for fake news and hate speech.

Keywords: *Disinformation, Hate speech, Fake news, Information disorder, Local politicians, Response activities and tools, Slovakia*

INTRODUCTION

The current digital society is shaping the transformation of the information environment. While there are significant benefits, such as the

¹ Faculty of Public Administration, Pavol Jozef Šafárik University in Košice, 04011 Košice, Slovak Republic, e-mail: ondrej.mital@upjs.sk, ORCID-ID 0000-0003-3773-8714..

² Faculty of Public Administration, Pavol Jozef Šafárik University in Košice, 04011 Košice, Slovak Republic, e-mail: miroslav.fecko@upjs.sk, ORCID-ID 0000-0003-2628-0007.

³ Faculty of Public Administration, Pavol Jozef Šafárik University in Košice, 04011 Košice, Slovak Republic, e-mail: silvia.rucinska@upjs.sk, ORCID-ID 0000-0001-8576-0874.

immediate dissemination and access of information, it also gives rise to problems like information overload and information disorder. These issues manifest as false and harmful information, including disinformation, fake news, hate speech, and various forms of inaccurate content (Farte, 2024; Maati et al., 2024; Schirch, 2021). It is recognised globally as a significant challenge for public institutions, democracy, and public policy (Denniss and Lindberg, 2025; Morse and Pratt, 2025; Ručinská et al., 2025). According to Flash Eurobarometer 522 Democracy data, exposure to disinformation and fake news among respondents in the EU 27 is notable with 13% reporting they are very often exposed, 22% often, and 33% sometimes, and contrarily, only 16% declare they are rarely exposed, and only 7% report never being affected (European Commission, 2023). In Slovakia specifically, 17% of respondents consider their exposure to disinformation and fake news as very frequent, while 22% consider it frequent, placing Slovakia ninth worst among EU countries in this regard (European Commission, 2023). The spread of disinformation, hate speech, and fake news indeed highlights the need for developing complex response policies.

We categorise the approaches used to respond to disinformation, based on our previous research (Ručinská, 2022), as targeted activities rather than random actions. In this context, we differentiate between two types of responses: soft approach and hard approach. Soft response activities and tools involve preventive measures aimed at enhancing the skills of recipients, thereby preparing them for potential disinformation threats. In contrast, hard response activities and tools are reactive in nature. It is usually manifested by blocking, limiting, legally regulating, sanctioning, banning, or deleting.

As digital transformation increasingly shapes the information environment, researchers are intensifying their studies of these topics, uncovering important theoretical insights, while policymakers are confronting the need to deal with the effects of information disorder on public policy and democratic systems (Aïmeur et al., 2023; Bennett and Livingston, 2021; Tandoc, 2019; Wardle and Derakhshan, 2017). As a result, response policies have become a crucial focus for governments worldwide. The complex nature of disinformation, fake news, and hate speech necessitates a multi-response approach to effectively confront and mitigate their negative effects.

The study aims to answer the research question: What is the viability of hard response activities and tools in responding to hate speech, fake news and disinformation? Existing empirical evidence was used to outline

dimensions of disinformation, hate speech, and fake news, highlighting the growing pressures that have led to a need for a multi-response approach to these issues. The theoretical section of this study focuses on the concepts of information overload and information disorder, particularly in the context of disinformation, fake news, and hate speech in the digital age. It emphasises the need for a multi-response approach, highlighting the practical application of hard response strategies and tools. The findings reveal the extent of disinformation, particularly concerning fake news and hate speech, and evaluate the effectiveness of hard response strategies as perceived by local politicians in Slovakia, based on a survey conducted in 2023. The discussion integrates contemporary knowledge and practices with the survey results, demonstrating the varying effectiveness of hard response activities and tools in combating disinformation, hate speech, and fake news. The findings from this examination contribute to understanding how such techniques can be effectively implemented in confronting these prevailing challenges.

1 INACCURATE INFORMATION IN THE DIGITAL AGE

The expressions of hate speech, fake news, and disinformation are numerous and complex. Following nearly a decade of intensive research into these extents of information disorder, a significant number of studies and analyses have been published (Broda, Strömbäck, 2024; Salaverría, Cardoso, 2023; Jerónimo et al., 2023). While there is an opportunity for growth, universally accepted definitions of these terms are still developing (French et al., 2024; Aïmeur et al., 2023; Pérez-Escobar et al., 2023). Exploring the interrelationships of these topics, such as their similarities and differences, could significantly enhance our understanding and provide a more comprehensive framework for discussion and response policies. For this study, the term “disinformation” will be utilised as a comprehensive concept, which reflects any type of information created and shared with the intention to harm and with false content (Hameleers, 2023; Simion, 2023; Ručinská et al., 2022). Theoretical analysis and empirical research will primarily concentrate on hate speech and fake news.

Hate speech and fake news appear not to be new issues because similar forms of human behaviour have been present in our civilisations from ancient times (UN, 2025; Posetti & Matthews, 2018). In recent decades, there has been a notable increase in interest in both scholarly analysis and boosted attention from policymakers (Hoes et al., 2025; Liu, 2025;

Prike and Ecker, 2023). This growing focus reflects a critical recognition of the complexities surrounding issues such as hate speech, fake news, and disinformation within the contemporary information overload (Pleios and Tastsoglou, 2025; Senft and Greenfield, 2023).

1.1 Hate speech, fake news and disinformation

As new media continue to shape the digital information space, the occurrence of disinformation, hate speech, and fake news is evolving rapidly. In this regard, disinformation, hate speech, and fake news can be observed in various forms.

Hate speech has been examined in academic circles since the last decade of the 20th century (Paz, 2020; McConnell, 1992). Hate speech is the most common form of destructive content on social media, and it can come in the form of text, photographs, or video (Saleh et al., 2023; Palani et al., 2022). The most complex view of hate speech encompasses the following characteristics: it provokes violence or hate, it attacks or diminishes, it targets specific groups, and it can also be connected with humour (Fortuna et al., 2018; Gagliardone et al., 2015). The most widely accepted idea in the context of hate speech is that this form is essentially tied to an emotion of hate (Marques, 2022). Hate speech must also be discussed in the context of factors that influence different users' perceptions, such as personal characteristics, presentation form, and content-related characteristics (Schmid et al., 2022). These characteristics only confirm that hate speech shall not be perceived as a sort of expression that the right to freedom of expression has to protect (Howard, 2019; Waldron, 2012). In this sense, hate speech refers to any form of communication that threatens or offends with aggression based on racial, ethnic, gender, religious, or other characteristics, thereby questioning human dignity (Tontodimamma, 2021; Charitidis et al., 2020; Geschke et al., 2019).

Hate speech is focused not only on traditional topics like religion or human rights, but also aims to harm individuals or groups in the light of new societal problems, such as gender questions, terrorism or emotions (Tontodimamma et al., 2021). Put differently, hate speech refers to the use of offensive language and wording, focused on a specific group of people with the same characteristics (Anderson and Barnes, 2022; Castaño-Pulgarín et al., 2021). Hate speech attacks the group's identity and belongings of members in a group and also belonging to society (Kovács et al., 2021; Gibert et al., 2018). As a consequence, hate speech is often increasingly concerned with

racism (Matamoros-Fernández and Farkas, 2021; Lamerichs et al, 2018). Language used in hate speech can be perceived as unethical, particularly due to its destructive content and consequences (Gorenc, 2022; Hietanen & Eddebo, 2022).

The most common negative impacts of hate speech include prejudice, violence, or, in the worst case scenario, even genocide (Gandhi et al., 2024; Vasist et al., 2024). Hate speech affects communities and public discourse (Hangartner et al., 2021). Hate speech establishes, confirms and also underlines social hierarchies based on inequality and domination (Paz, 2020).

Fake news, similar to hate speech, often utilises multimedia information or content to mislead readers, thereby spreading and expanding its influence, as well as causing severe social panic (Comito et al., 2023; Xue et al., 2021). There is a universal acceptance that fake news influences our societies and the decisions of individuals, particularly following the 2016 US presidential election, Brexit, or the COVID-19 pandemic (Mihálik et al., 2022; Brashier et al., 2021; Xue et al., 2021). These events can be perceived as a catalyst for governments worldwide to pay attention to policy initiatives (Cipers et al., 2023). Digital transformation and technology can be perceived as an accelerator of fake news (Grimes & Gorski, 2025; Pherson et al., 2021).

The gradual questioning of the truth or falsity of information, along with its ever-increasing volume, contributed to the deliberate spread of fake news. Fake news can be defined as false, inaccurate or misleading information created, published and disseminated for intentionally or unintentionally causing damage in society, or for gaining some advantage (Tandoc and Seet, 2024; Bader and Bender, 2022). Fake news may refer to the manipulation of information that can be carried out through the production of false information or the distortion of accurate information (Amri et al., 2023). Fake news attempts to distort the truth with diverse linguistic styles while simultaneously parodying trustworthy news (Shu, 2022). Fake news can also be considered a new form of information poverty, where individuals fail to seek truthful information (Cooke, 2021).

Fake news must also be discussed in the context of factors that influence different users' perceptions, dissemination processes, spreading channel features, outcomes, fabricated legitimacy, and attitudes (Di Domenico et al., 2021). Together with emotionally evocative content, other psychological factors, personality, literacy, cognitive factors, and a combination of internal and external sociological factors can influence the tendency to believe fake news (Beauvais, 2022).

Fake news stands in contrast to real news and information published for a long time by verified sources, whose socially beneficial task was to separate fact from falsehood (Gradoń et al., 2021; Tandoc et al., 2021). Fake news naturally provides supplemental stories to a large number of stories included in mainstream media discussions (Galletta Horner et al., 2023). A similar phenomenon, evident in the last few years, is deepfake. Deepfakes, created using deep learning techniques, are artificially generated audio-visual interpretations that can be used, for example, to damage an individual's reputation, disseminate disinformation or hate speech, or simply spread chaos in society (Salman et al., 2023).

A key characteristic of fake news is its tendency to be emotionally evocative to the general public or a specific group of individuals (Pennycook and Rand, 2021). Fake news can determine how governments, organisations, and people respond to everyday events in society (Olan et al., 2024). The ability of fake news to influence public policy discourse and politics is increasing in our democracies (Lewandowsky & Van Der Linden, 2021). In this regard, fake news can pollute the public sphere, damage democracy, and compromise democratic processes and public welfare (D'Andrea et al., 2025; Adjin-Tettey, 2022). Fake news is shared to make a profit or influence the behaviour of individuals or groups in our society (Bastick, 2021; Szczepański et al., 2021). In the digital age, fake news and hate speech are disseminated at a rapid pace in an almost unrestricted environment (Di Domenico et al., 2021; Mathew et al., 2020).

Disinformation is a concept that refers to false, inaccurate or misleading information created, presented and disseminated intentionally to cause harm in the public space, or to gain profit (Kapantai et al., 2020; Colomina et al., 2021). Disinformation is problematic since its goal is not to provide valuable information to the receiver, because it inherently creates ignorance. Therefore, in this sense, we can wonder whether disinformation is information in the true meaning of the word at all (Simion, 2023). Disinformation is the intentional dissemination of falsehoods or distortions, often spread as news, to achieve goals such as discrediting opponents, disrupting policy debates, influencing voters, inflaming existing social conflicts, or creating a general backdrop of confusion and informational paralysis (Bennett and Livingston, 2021; Marwick and Lewis, 2017).

Disinformation, as well as hate speech and fake news, have evolved into the unwanted but stable part of our digital society. Based on the current state of knowledge mentioned, it is worth noting that disinformation affects our society, democracies, governments, businesses, and individuals differently.

These differences reflect the life cycle of disinformation, which consists of creation, dissemination, impact, and response to disinformation (Ručinská et al., 2025).

To understand the essence of the examined phenomena, we have primarily focused on the phase related to creation, or just before it. Therefore, in the next section, we will mainly focus on responding to disinformation. The study focuses on the reaction to disinformation because, in contemporary democracies, there is a growing demand for solutions aimed at creating complex response policies.

1.2 Responding to disinformation

The complexity of disinformation, hate speech, and fake news leads to the assumption that a comprehensive solution is necessary and there is no one, or no simple and easy solution. It is necessary to establish clarity regarding the terminology associated with reactions and solutions to disinformation, hate speech, and fake news. Various authors use differing terminologies, which results in a lack of consensus regarding the most accurate or appropriate term to use. One prevalent term is “tackling,” which represents efforts to deal with the mentioned issues (Sádaba and Salaverría, 2023; Nenadić, 2019). At the same time, “combating” is primarily associated with strategies aimed at reducing or preventing the dissemination of harmful content (Rød et al., 2025; Chen, 2024). The term “countering” is predominantly utilised in contexts that involve specific actions taken against some instances of disinformation, hate speech, or fake news (McBride et al., 2023; Gradoń et al., 2021). Considering these distinctions, the term “response” can be regarded as a more comprehensive concept. It signifies a targeted, systematic, and sophisticated response and solutions that encompass a variety of activities and tools utilised to address and mitigate the challenges caused by disinformation, hate speech, and fake news (Tandoc & Seet, 2024; Ručinská et al., 2022; Vafeiadis et al., 2022). The complexity of the problem necessitates multi-response activities and tools (Cho et al., 2024; Gradoń et al., 2021; de Cock Buning, 2018; European Union, 2018).

The multi-response approach is examined from various angles, depending on whether the focus is on policy areas, the scope of responses, or the types of responses. In contemporary democracies, responses shall integrate a wide range of policy fields related to civil security in both the public and private spheres, including legal and regulatory, educational, political and

governance, psychological and social psychological, and technological domains (Rød et al., 2025). Responding can be applied from different perspectives and approaches, such as the planning approach, truth theory, systems thinking, and military approach (Dowse & Bachmann, 2022). As part of a multi-response approach, interactions among government actions, private sector involvement, and citizen engagement are necessary (D'Andrea et al., 2025).

Multi-response approach by governments is growing, focusing mainly on fact-checking and monitoring responses, investigative responses, countercampaigns, election-specific responses, curational responses, technical and algorithmic responses, demonetising and economic responses, ethical and normative responses, educational responses, empowerment responses and COVID-19-specific responses (Cipers et al., 2023; Bontcheva et al., 2020). Response strategies are imposed by the government or created by the private sector, and both initiatives may include content-focused solutions or user-focused solutions (Pherson et al., 2021).

As shown in Table 1, the Disinformation response activities and tools reveal two main types of response activities. We have specified a classification of approaches used to effectively respond to disinformation, emphasizing that these actions are intentional and strategically formulated rather than occurring randomly (Ručinská et al., 2022). This crucial classification highlights the need for a structured framework that guides the implementation of responses to disinformation. By differentiating responses based on their nature and objectives, we can better understand how to respond disinformation effectively in various contexts and meanings. This targeted approach allows for more tailored interventions that align with the specific types of the disinformation encountered.

Table 1: *Disinformation response activities and tools*

SOFT		
Open data	Open government	Critical thinking
Soft skills for the digital age	Ethical codex for social networks	Digital ethical code
Opinion plurality	Influence of public authorities	Policies of technological companies
HARD		
Regulation	National legislation acts	EU binding norms
Sanctions	Technical regulation	Deleting posts
Banning accounts	Demonetising accounts	Web search restrictions

Source: Ručinská et al., 2022

The table summarises hard and soft responses to disinformation, which we perceive as a typology rather than a specific, exhaustive list of all possible response activities and tools. We see hard activities and tools as reactive solutions to disinformation, primarily aimed at limiting, prohibiting and blocking. Soft activities and tools can also be preventive, aiming to upskill recipients and prepare them for potential disinformation targeting. Hard solutions can, on the contrary, be seen as reactive measures, which primarily target the source of disinformation or the creator of a particular piece of disinformation. Consequently, we view the highlighted response activities as complementary to one another. By utilising a variety of strategies, such as public awareness campaigns, fact-checking efforts, and initiatives aimed at improving media literacy, alongside website blocking, social media filters, and identifying creators, democracies can more effectively respond to disinformation. Each strategy targets different aspects of the issue, thereby enhancing overall ability to manage and mitigate the impact of information disorder in our society.

Key hard response tools and activities are primarily understood as legal norms, legislation, and regulations established at the national level. While legal frameworks at the national level are essential, it is important to recognise that such laws may, in certain instances, restrict freedom of expression and democratic principles. Therefore, the implementation of

these regulations and sanctions must be approached carefully (Nagasako, 2020; Claussen, 2018). National legal issues and challenges regarding normative responses to disinformation, hate speech, and fake news should be considered within the framework of international law. Regulations often target both local and foreign groups, making it more challenging to respond effectively. Therefore, governments need to find a balance between addressing disinformation and respecting international standards (Vese, 2021; Tenove, 2020). In this regard, the responsibilities of states from the perspective of international human rights law lie in taking necessary and appropriate measures that, at the same time, protect freedom of expression and prevent the spread of disinformation, hate news, and fake news (Balcerzak, Kapelańska-Pręgowska, 2023).

This legislative framework is notably shaped by the directives and tendencies articulated by the European Union. The formulation and implementation of standard European response policies is gradually becoming, along with traditional economic and political goals, an essential part of strategies within the European Union (Guedes et al., 2025; Bayer, 2024). However, member states of the European Union continue to adopt varying regulations and platforms, demonstrating that they reflect the general legal framework established by the European Union (Caruso, 2024; Cavaliere, 2022).

Hard response activities and tools derive from this legal basis and may encompass a range of technological regulations or sanctions. These interventions are often aimed at mapping and regulating the functioning of social networks (Tan, 2022; Nagasako, 2020) and websites (Sokol, Bachňáková Rózenfeldová, 2025; Ververis et al., 2024), potentially involving measures such as demonetisation of content, removal of posts, website blocking or restrictions on internet searches. Public service media are regulated from a normative perspective through media policy intervention (Horowitz et al., 2022). The establishment of special governmental bodies and special working groups can also be perceived as part of countermeasures policies (Hanzelka & Pavlíková, 2021).

2 METHODOLOGY

A comprehensive European research project, endorsed by the Congress of Local and Regional Authorities of the Council of Europe, brought together a diverse group of European academic institutions to investigate the significant impact of hate speech, dis- and misinformation on local and

regional governance structures. This initiative launched with the distribution of a detailed questionnaire to congress delegates, aiming to capture a wide array of insights. The project explored the trends and harm of fake news and hate speech, and also the way they affect the local political decisions (Chlapák, 2021). The results were evaluated as important to ensure a meaningful and democratic political life to serve the needs of all citizens, by the Congress of Local and Regional Authorities (Kiefer, 2022). To further deepen the research, the team conducted in-depth surveys of politicians and local administrators across Germany, Austria, Italy, Romania, Moldova, Hungary, and Slovakia, seeking to understand their perspectives and first-hand experiences in addressing the challenges posed by the disinformation, fake news and hate speech. The surveys were carried out during the second and third quarters of 2023. Partial results from surveys conducted in seven EU countries were published by colleagues from the Eurac Research Institute for Public Management in Italy (Maffei and Moscon, 2024). In a pilot study, the results were also published for German local and regional politicians and administrative leaders (Bagnato et al., 2024). The empirical research presented in this article is a partial result of the abovementioned initiative undertaken by a consortium of European research institutions. At the same time, we are building on and continuing our research, which conceptually and schematically describes the life cycle of disinformation and the necessity for a multi-response approach at all its stages and classification system for soft and hard response activities (Ručinská et al., 2025; Ručinská et al., 2022).

The study aims to answer the research question: What is the viability of hard response activities and tools in responding to disinformation, hate speech and fake news among local politicians and communal practitioners in Slovakia? The findings from this investigation contribute to understanding how such methods can be effectively implemented in confronting Information disorder.

The survey design has been developed primarily to analyse the extent and responses related to hate speech and fake news, particularly in their associations with disinformation. This effort is grounded in the collaborative expertise of collaborating European academic institutions and previous collective research. The survey questions permitted respondents to indicate their preferences using a three-point scale (hardly ever, at times, frequently). Additionally, the respondents had the opportunity to express their agreement or disagreement with various statements through a different set of questions.

The research sample was intentionally chosen to focus on municipalities and cities located in the Košice and Prešov regions of Slovakia, reflecting the fragmentation of local self-government in these areas. A total of 1,103 municipalities and cities were invited to participate in the survey, which was distributed through the Evasys survey and examination software. The survey generated a response rate of 12.06%, as detailed in Table 2. Data were collected between July 2023 and October 2023, with two reminder notifications sent during this period.

Table 2: Municipalities and cities outline

Respondent position	
local politician	3
mayor	59
member of the administration	69
Not stated	2
Cities and municipalities' size categories	
0-200	7
201 - 500	25
500 - 1000	30
1001 - 2000	30
2001 - 5000	12
5001 - 20 000	9
20 001 - 50 000	11
50 000 - 500 000	0
Not stated	9
Age category	
under 30 years of age	1
between 30 and 50 years of age	71
over 50 years of age	60
Not stated	1

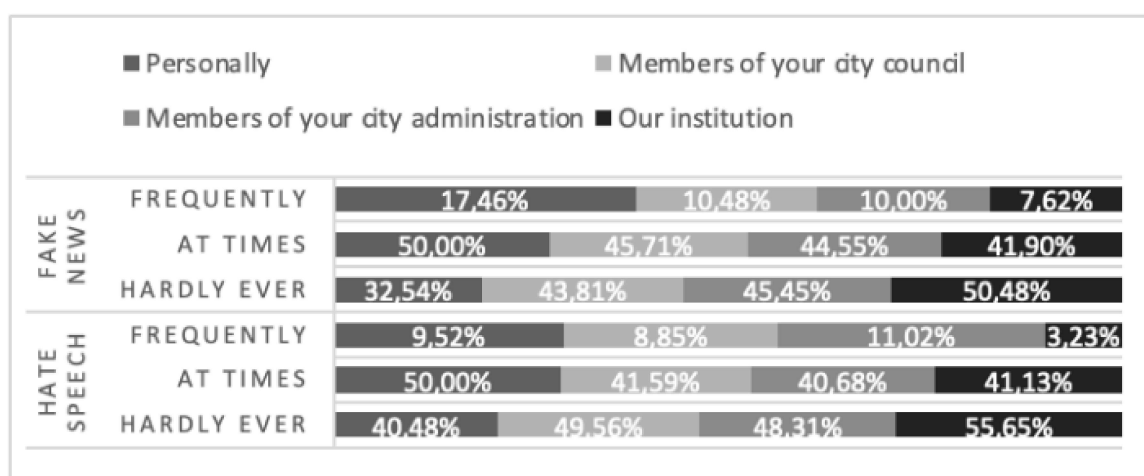
Source: authors

3 RESULTS

Focusing on the occurrence rate of fake news and hate speech, respondents were asked if they had experienced hate speech or fake news. Only 9.52% of respondents in the case of hate speech and 17.46% of respondents in the case of fake news often experienced the examined terms in person. The respondents identified that they most often encounter fake news personally. On the other hand, according to the respondents (11.02%), the term 'hate speech' is most often encountered by employees of administrative city offices. Respondents state that, regardless of the target, they experience fake news more often than hate speech.

Considering fake news and hate speech, the least frequently targeted entity perceived by respondents is the organisation of the municipality or city as a whole. Respondents perceive that the target of fake news and hate speech is themselves rather than other members of the council or other members of the administration of the office in the city or municipality. In this context, the results indicate that both fake news and hate speech are primarily perceived concerning the respondents themselves (Graph 1).

Graph 1: *Perceived occurrence of fake news and hate speech*



Source: authors

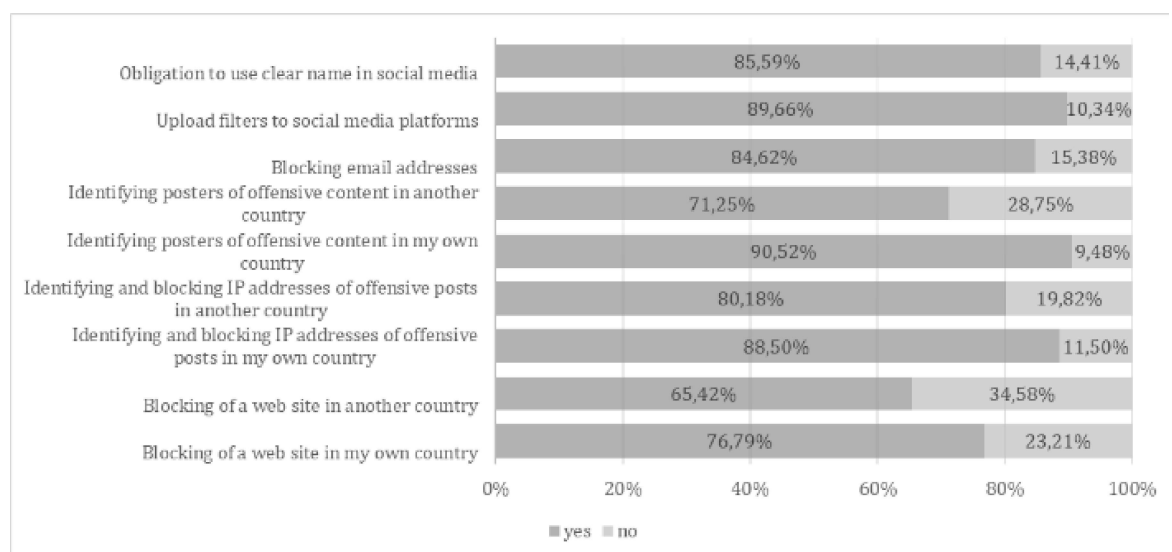
The extent of hate speech faced by respondents differs. Respondents were hardly ever confronted with hate speech associated with physical violence against the respondent's family (85 respondents) or physical violence against the respondent's person (83 respondents). Respondents least often

described the extent of hate speech as rare in the case of slander in real life, personal insults in real life, personal insults in the media and slander in the media. Subsequently, the extent of hate speech was perceived by respondents as frequent in the case of slander in real life (42 respondents), personal insults in real life (15), slander in the media (13) and personal insults in the media (7).

While respondents identified the issue of hate speech in the shown contexts, they encountered fake news hardly ever and at times more frequently. Respondents encountered fake news hardly ever in connection with influencing the decision-making process in a municipality or city (56 respondents) and as part of hate speech (58 respondents). On the other hand, respondents most often identified fake news in the context of influencing municipal elections, specifically in 22% of responses, and also in the context of controlling the decision-making process in a municipality or city, as noted in 14 responses.

Respondents were also asked which measures they would consider a technically and legally viable option against fake news and hate speech, with a total of 9 options to choose from (Graph 2).

Graph 1: *Legal and technical viability of selected hard response activities and tools*



Source: authors

More than two-thirds of respondents agree that the measures are technically and legally viable. Respondents expressed the least agreement

with the viability of the measure blocking a website in another country (65.42%) and the measure identifying posters of offensive content in another country (71.25%). In other words, these measures are considered difficult in terms of their successful implementation. The measures that are technically and legally most feasible from the respondents' point of view are identifying posters of offensive content in their own country, with 89.66% of respondents agreeing with this tool. Respondents also consider uploading filters to social media platforms and identifying and blocking IP addresses of offensive posts in their own country to be feasible, as only slightly more than 10% of respondents perceive these as difficulties. Measures that can be considered moderately difficult to implement, based on the respondents' answers, include the obligation to use a clear name in social media (101 respondents) and blocking email addresses (99 respondents). Based on the respondents' answers, it can be said that they perceive measures in their own country as technically and legally more viable. Respondents consider measures that should be implemented in a country other than their home country to be more challenging to implement.

4 DISCUSSION

The results suggest that both fake news and hate speech are primarily perceived by the respondents themselves, rather than by an institution. In this regard, research confirms the trend that disinformation is mainly targeted at individuals rather than organisations, cities or governments (Brashier, 2024; CISA, 2022; Woolley, 2022).

Respondents state also, regardless of the target, they experience fake news more often than hate speech. The extent of perceived hate speech and fake news also reflects some key trends. Within the context of hate speech, the most common contexts are slander and personal insults, both in real life and in the media. Fake news was identified by respondents primarily in connection with influencing the results of the electoral process. This part of democratic processes is most often associated with the targeted provision of false information.

Respondents from our research were hardly ever confronted with physical violence against the respondent's family or physical violence against the respondent's person. This result is slightly different from cases in other states, where physical violence is also being discussed before electoral campaigns. Cases when politicians were attacked while they were performing political agitation and political campaign in Germany

(Tagesschau, 2025; Neubert, 2024; Stroud, 2024), in Slovakia (Mihalková, 2023), in Czechia (Karban, 2021) or Great Britain (BBC, 2016). Violence is also linked to the activities of politicians during their terms of office in Denmark (Henley, 2024) and France (Gijs, 2022).

In our theoretical research, we focused on three areas based on desk research: hate speech, fake news and disinformation. Subsequently, our empirical research focused on hate speech and fake news. Based on theoretical discourse and empirical results, we present general findings on the applicability of robust responses to disinformation, hate speech and fake news. In various forms and to varying extents, disinformation and related terms are being increasingly perceived as an unwanted aspect of political life (Nai et al., 2025; Berntzen et al., 2024). Given the extent and negative impact of disinformation confirmed by theory and empirical research, it is necessary to discuss the need for a response. Disinformation theory emphasises the importance of a multi-faceted response approach. From a legal implementation and technical readiness perspective, the research confirmed the importance of measures taken within social networks to identify individuals sharing offensive content, as well as blocking IP addresses. Research confirmed the need for a multi-response approach, especially in the context of a single trajectory. Empirical results confirmed that respondents perceive hard response activities and tools as viable. Respondents agreed that specific hard tools and activities are technically and legally feasible in practice.

In response to the tools, respondents identified activities and tools that should be implemented abroad as more challenging to implement. Respondents, knowing the legal environment, are aware of the limits of cross-border interventions in website providers, and rationally perceive these tools and activities as difficult. Despite intensive research and discussions, there is no unified framework to manage response policies concerning disinformation. Although theoretical discourse advocates for a multi-response approach, implementing such a strategy necessitates careful formulation by governmental entities and accountable public institutions to ensure its effectiveness. A practical framework aimed at strengthening societal resilience against disinformation should thoroughly address multiple dimensions, including legal and regulatory measures, educational initiatives, political and governance structures, psychological factors and technological components (Rød et al., 2025). A multi-response framework must focus on effective prevention, resilience building, observation, disclosure, evaluation, and impact management (Sommarberg et al., 2024).

International experiences, especially from globally engaged organisations, provide a robust foundation for understanding the multi-response approach to combat disinformation effectively. In the European Union context, the Code of Conduct on Disinformation was integrated into the framework of the Digital Service Act (European Union, 2025a). The Code of Conduct on Disinformation consolidates the content of the multi-response policies into six key areas: demonetisation, transparent political advertising, reducing manipulative behaviour, user empowerment, fact-checking coverage throughout the EU, and data access for research (European Union, 2025b). Considering the worldwide context, states should protect, respect and promote freedom of expression and ensure access to information, avoid regulating based on vague definitions, refrain from internet shutdowns, provide accurate information sharing, hold public officials accountable, and involve civil society in the design of response policies (United Nations, 2022). Under the OECD initiatives, a framework for response policies should include communication strategies, plans, and coordination mechanisms, as well as identifying and debunking disinformation, regulatory and legal measures, and civic and media initiatives (Matasick et al., 2020).

Based on our research, the order of feasibility and success of individual hard responses differs in general for disinformation, as well as for fake news and hate speech. In examining the effectiveness of hard responses to disinformation, hate speech and fake news, it is evident that there are no fundamental differences among certain types. For instance, blocking email addresses is widely considered by both respondents and theoretical frameworks as very challenging to execute, often delivering minimal impact. This difficulty is also reflected in the context of implementing hard responses, such as working groups, the significance of national legislation and the requirement to use a clear name in social media interactions. Resulting from the literature review and empirical research, we have identified the viability of hard responses (Table 3).

Table 3: Viability of hard responses and activities

	Disinformation	Hate speech	Fake news	
h i g h	Blocking email addresses	Blocking email addresses	Blocking email addresses	l o w
	Working groups	Blocking of a website	Working groups	
	Obligation to use a clear name on social media	Working groups	Obligation to use a clear name on social media	
	National legislation	Obligation to use a clear name on social media	National legislation	
V i a b i l i t y	Identifying and blocking IP addresses	National legislation	Identifying posters of offensive content	V i a b i l i t y
	Filters on social media platforms	Identifying and blocking IP addresses	Identifying and blocking IP addresses	
	Blocking of a website	Filters on social media platforms	Filters on social media platforms	
	Identifying posters of offensive content	Identifying posters of offensive content	Blocking of a website	

Source: authors

Responses considered to have a higher degree of viability, such as filters on social media platforms and identifying and blocking IP addresses, are assessed to be generally similar in their level of effectiveness. Considering the complexity of the information environment in the digital age, implementing these measures often struggles to achieve the desired outcomes.

Our research indicates that the viability of blocking websites as a hard response tool varies significantly depending on the context. When it comes to hate speech, this approach is often considered ineffective due to the immediate and dynamic nature of such content. In contrast, blocking websites proves to be a much more effective response to fake news. Since these websites serve as the primary platforms for disseminating misleading information, their removal can significantly decrease the spread of false narratives and disinformation.

The process of identifying posters of offensive content, while having a high applicability in addressing both hate speech and disinformation, raises a key challenge when responding to fake news. In the case of fake news, linking the spread of fake news back to its original creators is particularly complex. This difficulty arises from the often anonymous or pseudonymous environments in which fake news increases.

CONCLUSION

A systematic approach to disinformation knowledge, supported and verified by empirical data in this study, contributes to the scientific and policy discourse. The study supports the current discussion on the multi-response approach, highlighting the importance of hard response activities and tools based on theoretical desk research and empirical research. The survey revealed that respondents perceive mainly hard response activities and tools as technically and legally viable. Respondents are also aware of the limits associated with the need to implement response activities and tools outside the borders of their home country. In this regard, the response policies of the European Union shall be perceived as necessary. We have discussed our research within the current trends related to disinformation, hate speech, and fake news in European countries. We underlined hard responses and tools as being a feasible part of the comprehensive strategies that effectively address disinformation challenges.

The research does show certain limitations, primarily concerning the empirical findings, which are based solely on the conditions of a single country with a relatively limited sample size. The evolving topic of hard responses to disinformation also poses a limitation to this research. As legislative frameworks shift and the sophistication of response tools and policies progresses, the effectiveness and challenges associated with these responses can change dramatically over time. Nevertheless, it is essential to interpret these limitations considering the broader context in which these results will be presented in the future, precisely as part of a comprehensive initiative undertaken by European academic institutions from seven European countries. Although the findings may reflect a degree of partiality, they nonetheless contribute meaningfully to the overall understanding of the issues being examined and offer valuable insights when contextualised within the larger research framework.

When considering future research trajectories, the study confirms present perspectives and introduces new questions within the disinformation.

Further examination into the perception of examined concepts across various demographic groups, particularly among young individuals, senior citizens, policymakers, and public servants, remains a crucial need. It is essential for research efforts to focus on developing practical response activities and tools that will function as complementary components within a multi-response approach to addressing disinformation. Such initiatives are essential for navigating the complexities of the contemporary information overload and information disorder.

To foster informed decision-making, individuals should actively cultivate an awareness of misleading and potentially harmful content. Individuals should critically evaluate the authenticity of information, especially in the face of information overload, to enhance their understanding and make more informed choices. Fundamental prerequisites include the knowledge and literacy required to identify various forms of disinformation and the sophistication of responses to it.

REFERENCES

- Adjin-Tettey, T. D. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education, *Cogent Arts & Humanities*, 9(1), article n. 2037229. <https://doi.org/10.1080/23311983.2022.2037229>
- Aïmeur, E., Amri, S., Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13, article number 30. DOI: <https://doi.org/10.1007/s13278-023-01028-5>
- Anderson, L., & Barnes, M. (2022). Hate Speech. [online]. Available at: <https://plato.stanford.edu/ENTRIES/hate-speech/>. [Accessed 17.04. 2025].
- Bader, L., & Bender, B. (2022). What Is “fake news” and ‘hate speech’ and How Do They Work in Practice?. In: Brüggemann, S., Kutlu, N., Müller-Török, R., Prosser, A., Ručinská, S., Szádeczky, T., Vrabie, C. (eds.) *Central and Eastern European EDem and EGov Days 342*. pp. 17-36. DOI: <https://doi.org/10.24989/ocg.v.342.1>
- Bagnato, D., Hemker, T., Müller-Török, R., Prosser, A., Sroka, S. (2024). Fake News and Hate Speech - Evidence from Germany. In: *CEEeGov 2024: Central and Eastern European eDem and eGov Days 2024*, Budapest September 2024. New York: Association for Computing Machinery, pp. 282 - 287. <https://doi.org/10.1145/3670243.3677383>
- Balcerzak, M., & Kapelańska-Pręgowska, J. (2023). International human rights law in the era of digital disinformation and propaganda: case studies from Myanmar and Ukraine. *Studia Prawnicze KUL*, 4(96), pp. 7-26. <https://doi.org/10.31743/sp.16760>
- Bastick, Z. (2021). Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*, 116, article ID 106633. <https://doi.org/10.1016/j.chb.2020.106633>
- Bayer, J. (2024). The EU policy on disinformation: aims and legal basis. *Journal of Media Law*, 16(1), pp. 18–27. <https://doi.org/10.1080/17577632.2024.2362478>
- BBC. (2016). Labour MP Jo Cox ‘murdered for political cause’. [online]. Available at: <https://www.bbc.com/news/uk-37978582>. [Accessed 17.04. 2025].
- Beauvais, C. (2022). Fake news: Why do we believe it?. *Joint Bone Spine*, 89(4), article ID 105371. <https://doi.org/10.1016/j.jbspin.2022.105371>

- Bennett, W. L., & Livingston, S. (2021). The Coordinated Attack on Authoritative Institutions - Defending Democracy in the Disinformation Age. In: Bennett, W. L., & Livingston, S. (eds.) *The Disinformation Age - Politics, Technology, and Disruptive Communication in the United States*. Cambridge: Cambridge University Press, pp. 261-293. <http://dx.doi.org/10.1017/9781108914628>
- Berntzen, L. E., Kelsall, H., Harteveld, E. (2024). Consequences of affective polarization: Avoidance, intolerance and support for violence in the United Kingdom and Norway. *European Journal of Political Research*, 63, pp. 927-949. <https://doi.org/10.1111/1475-6765.12623>
- Brashier, N. M. (2024). Fighting misinformation among the most vulnerable users. *Current Opinion in Psychology*, 57, article ID 101813. <https://doi.org/10.1016/j.copsyc.2024.101813>
- Broda, E., Strömbäck, J. (2024). Misinformation, Disinformation, and Fake News: Lessons from an Interdisciplinary, Systematic Literature Review. *Annals of the International Communication Association*, 48(2), pp. 139-166, <https://doi.org/10.1080/23808985.2024.2323736>
- Caruso C. (2025). Towards the Institutions of Freedom: The European Public Discourse in the Digital Era. *German Law Journal*. 26(1), pp. 114-137. <https://doi.org/doi:10.1017/glj.2024.68>
- Castañó-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T., López, H. M. H. (2021). Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior*, 58, article ID 101608. <https://doi.org/10.1016/j.avb.2021.101608>
- Cavaliere, P. (2022). The Truth in Fake News: How Disinformation Laws Are Reframing the Concepts of Truth and Accuracy on Digital Platforms. *European Convention on Human Rights Law Review*, 3(4), pp. 481-523. <https://doi.org/10.1163/26663236-bja10044>
- Cipers, S., Meyer, T., Lefevere, J. (2023). Government responses to online disinformation unpacked. *Internet Policy Review*, 12(4), pp. 1-19. <https://doi.org/10.14763/2023.4.1736>
- CISA. (2022). *Tactics of Disinformation*. [online]. Available at: https://www.cisa.gov/sites/default/files/publications/tactics-of-disinformation_508.pdf. [Accessed 17.04. 2025].
- Claussen, V. (2018). Fighting Hate Speech and Fake News. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation. *Rivista di Diritto dei Media*, 2(3), pp. 110-136.

- Colomina, C., Sánchez Margalef, H., Youngs, R. (2021). The impact of disinformation on democratic processes and human rights in the world. Brussels: European Union. [online]. Available at: https://consellodacultura.gal/mediateca/extras/CCG_ac_2023_Estudio_ParlamentoEuropeo_reducido.pdf. [Accessed 17.04. 2025].
- Comito, C., Caroprese, L., Zumpano, E. (2023). Multimodal fake news detection on social media: a survey of deep learning techniques. *Social Network Analysis and Mining*. 13, article number 101, pp. 1-22. <https://doi.org/10.1007/s13278-023-01104-w>
- Cooke, N. A. (2021). A Right to be Misinformed? Considering Fake News as a Form of Information Poverty. In: Taylor, N. G., Kettlich, K., Gorham, U., Jaeger, P. T. (eds.) *Libraries and the Global Retreat of Democracy: Confronting Polarization, Misinformation, and Suppression (Advances in Librarianship vol. 50)*. Leeds: Emerald Publishing Limited, pp. 45-60. <https://doi.org/10.1108/S0065-283020210000050002>
- D'Andrea, A., Fusacchia, G., D'Ulizia, A. (2025). Policy Review: Countering Disinformation in the Digital Age - Policies and Initiatives to Safeguard Democracy in Europe. *Information Polity*, 30(1), pp. 82-91. <https://doi.org/10.1177/15701255251318900>
- de Cock Buning, M. (2018). A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation. Luxembourg: Publications Office of the European Union, 39 p. ISBN 978-92-79-80419-9.
- de Gibert, O., Perez, N., García-Pablos, A., Cuadros, M. (2018). Hate Speech Dataset from a White Supremacy Forum. In: *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, Belgium, October 31 2018. Brussels: Association for Computational Linguistics. pp. 11–20. <https://doi.org/10.18653/v1/W18-5102>
- Denniss, E., Lindberg, R. (2025). Social media and the spread of misinformation: infectious and a threat to public health. *Health Promotion International*, 40(2), article ID: daaf023. <https://doi.org/10.1093/heapro/daaf023>
- Di Domenico, G., Sit, J., Ishizaka, A., Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, 124, pp. 329-341. <https://doi.org/10.1016/j.jbusres.2020.11.037>
- Dowse, A., & Bachmann, S. D. (2022). Information warfare: methods to counter disinformation. *Defense & Security Analysis*, 38(4), pp. 453-469. <https://doi.org/10.1080/14751798.2022.2117285>
- European Commission. (2023). Flash Eurobarometer 522 Democracy Report. [online]. Available at: <https://www.ipsos.com/sites/default/>

- files/ct/news/documents/2023-12/Democracy_fl_522_report_en.pdf. [Accessed 17.04.2025].
- European Union. (2018). A multi-dimensional approach to disinformation – Report of the independent High level Group on fake news and online disinformation. [online]. Available at: <https://data.europa.eu/doi/10.2759/739290>. [Accessed 17.04.2025].
- European Union. (2025a). The 2022 Code of Practice on Disinformation. [online]. Available at: <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>. [Accessed 17.04.2025].
- European Union. (2025b). The Code of Conduct on Disinformation. [online]. Available at: <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>. [Accessed 17.04.2025].
- Farte, G. (2024). The Digital Transformation of the Democratic Public Sphere: Opportunities and Challenges. *Meta: Research in Hermeneutics, Phenomenology, and Practical Philosophy*, 16(2), pp. 484-513.
- Fortuna, P., & Nunes, S. (2018). A Survey on Automatic Detection of Hate Speech in Text. *ACM Computing Surveys*, 51(4), pp. 1-30. <https://doi.org/10.1145/3232676>
- French, A., Storey, V. C., Wallace, L. (2024). A typology of disinformation intentionality and impact. *Information Systems Journal*, 34(4), pp. 1324-1354. <https://doi.org/10.1111/isj.12495>
- Gagliardone, I., Gal, D., Alves, T., Martinez, G. (2015). Countering online hate speech. Paris: United Nations Educational, Scientific and Cultural Organization, 71 p. ISBN 978-92-3-1001105-5.
- Galletta Horner , Ch., Galletta, D., Crawford, J., Shirsat, A. (2023). Emotions: The Unexplored Fuel of Fake News on Social Media. In: Dennis, A. R., Galletta, D. F., Webster, J. (eds.) *Fake News on the Internet*. London: Routledge, pp. 147-174. <https://doi.org/10.4324/9781003433934>
- Gandhi, A., Ahir, P., Adhvaryu, K., Shah, P., Lohiya, R., Cambria, E., Poria, S., Hussain, A. (2024). Hate speech detection: A comprehensive review of recent works. *Expert Systems*, 41(8), article ID e13562. <https://doi.org/10.1111/exsy.13562>
- Geschke, D., Klaßen, Quent, M., Richter, Ch. (2019). #Hass im Netz: Der Schleichende Angriff auf Unsere Demokratie, Eine Bundesweite Repräsentative Untersuchung. Berlin: Amadeu Antonio Stiftung. [online]. Available at: https://blog.campact.de/wp-content/uploads/2019/07/Hass_im_Netz-Der-schleichende-Angriff.pdf. [Accessed 17.04.2025].
- Gijs, C. (2022). Tomatoes, a slap and lots of eggs: 5 times French protesters attacked Emmanuel Macron. [online]. Available at: <https://www.politico>.

- eu/article/5-times-emmanuel-macron-attack-food/. Accessed [17.04. 2025].
- Gorenc, N. (2022). Hate speech or free speech: an ethical dilemma? *International Review of Sociology*, 32(3), pp. 413–425. <https://doi.org/10.1080/03906701.2022.2133406>
- Gradoń, K. T., Holyst, J. A., Moy, W. R., Sienkiewicz, J., Suchecki, K. (2021). Countering misinformation: A multidisciplinary approach. *Big Data & Society*, 8(1), article 20539517211013848. <https://doi.org/10.1177/20539517211013848>
- Grimes, D. R., Gorski, D. H. (2025). Quantifying Public Engagement With Science and Malinformation on COVID-19 Vaccines: Cross-Sectional Study. *Journal of Medical Internet Research*, 27, article ID e64679. <https://doi.org/10.2196/64679>
- Guedes, I. S., Vigo, B., Santos, M., Moreira, S., Cardoso, C. S., Castro, J. (2025). Disinformation Dynamics and Regulation in Portugal: Insights from a Qualitative Study. *European Journal on Criminal Policy and Research*, 2025. <https://doi.org/10.1007/s10610-025-09617-0>
- Hameleers, M. (2023). Disinformation as a context-bound phenomenon: toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory*, 33(1), pp. 1-10. <https://doi.org/10.1093/ct/qtac021>
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Buse Demirci, B., Derksen, L., Hall, A., Jochum, M., Murias Munoz, M., Richter, M., Vogel, F., Wittwer, S., Wuthrich, F., Gilardi, F., Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50), article e2116310118. <https://doi.org/10.1073/pnas.2116310118>
- Hanzelka, J., Pavlíková, M. (2021). Institutional Responses of European Countries. In: Gregor, M., & Mlejnková, P. (eds.) *Challenging Online Propaganda and Disinformation in the 21st Century. Political Campaigning and Communication*. Palgrave Macmillan, Cham, pp. 195-224. https://doi.org/10.1007/978-3-030-58624-9_7
- Henley, J. (2024). Danish PM 'not doing great' four days after assault in Copenhagen. [online]. Available at: <https://www.theguardian.com/world/article/2024/jun/11/danish-pm-mette-frederiksen-not-doing-great-four-days-after-suspected-assault>. [Accessed 17.04. 2025].

- Hietanen, M., & Eddebo, J. (2022). Towards a Definition of Hate Speech— With a Focus on Online Contexts. *Journal of Communication Inquiry*, 47(4), pp. 440-458. <https://doi.org/10.1177/01968599221124309>
- Hoes, E., Clemm, B., Gessler, T., Qian, S., Wojcieszak, M. (2025). (Media Attention to) Misinformation Can Undermine Trust in Scientists. *Political Behavior*. <https://doi.org/10.1007/s11109-025-10090-y>
- Horowitz, N., Cushion, S., Dragomir, M., Gutiérrez Manjón, S., Pantti, M. (2022). A Framework for Assessing the Role of Public Service Media Organizations in Countering Disinformation, *Digital Journalism*, 10(5), pp. 843-865. <https://doi.org/10.1080/21670811.2021.1987948>
- Howard, J. W. (2019). Free Speech and Hate Speech. *Annual Review of Political Science*, 22, pp. 93-109. <https://doi.org/10.1146/annurev-polisci-051517-012343>
- Charitidis, P., Doropoulos, S., Vologianidis, S., Papastergiou, I., Karakeva, S. (2020). Towards countering hate speech against journalists on social media. *Online Social Networks and Media*, 17, article ID 100071. DOI: <https://doi.org/10.1016/j.osnem.2020.100071>
- Chen, L. (2024). Combatting Climate Change Misinformation: Current Strategies and Future Directions. *Environmental Communication*, 18(1-2), pp. 184-190. <https://doi.org/10.1080/17524032.2023.2299756>
- Chlapak, R. (2021). A scientific basis for a policy fighting fake news and hate speech [online]. Available at: <https://www.coe.int/en/web/congress/-/counterfake-a-scientific-basis-for-a-policy-fighting-fake-news-and-hate-speech>. [Accessed 17.04. 2025].
- Cho, H., Li, W., Lopez, R. (2024). A Multidimensional Approach for Evaluating Reality in Social Media: Mixed Methods Study. *Journal of medical Internet research*, 6, article ID 26:e52058. <https://doi.org/10.2196/52058>
- Jerónimo, P., Amaral, I., Correia, J. C. (2023). Disinformation Studies: Global Perspectives. *Journalism Practice*, 17(10), pp. 2079–2083. <https://doi.org/10.1080/17512786.2023.2253246>
- Kapantai, E., Christopoulou, A., Berberidis, Ch., Peristeras, V. (2020). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, 23(5), pp. 1301-1326. <https://doi.org/10.1177/1461444820959296>
- Karban, F. (2021). Útok na Babiše se stále prošetřuje. Policie akci podcenila, míní expert. [online]. Available at: <https://www.seznamzpravy.cz/clanek/utok-na-babise-se-stale-prosetruje-policie-nebyla-pripravena-mini-expert-171782>. [Accessed 17.04. 2025].

- Kiefer, A. (2022). Foreword by Dr Andreas Kiefer, General Secretary of the Congress. In: Brüggemann, S., Kutlu, N., Müller-Török, R., Prosser, A., Ručinská, S., Szádeczky, T., Vrabie, C. (eds.) *Counterfake: A scientific basis for a policy fighting fake news and hate speech*. Wien: Facultas Verlags- und Buchhandels AG, pp. 5-6. <https://doi.org/10.24989/ocg.v.342.0>
- Kovács, G., Alonso, P., Saini, R. (2021). Challenges of Hate Speech Detection in Social Media. *SN Computer Science*, 2(95), pp. 1-15. <https://doi.org/10.1007/s42979-021-00457-3>
- Lamerichs, N., Nguyen, D., Puerta Melguizo, M. C., Radojevic, R. (2018). Elite male bodies: The circulation of alt-Right memes and the framing of politicians on Social Media. *Participations: Journal of Audience & Reception Studies*, 15(1), pp. 180-206.
- Lewandowsky, S., & Van Der Linden, S. (2021). Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology*, 32(2), pp. 348-384. <https://doi.org/10.1080/10463283.2021.1876983>
- Liu, Z. (2025). Can shifting attention to accuracy reduce misinformation on social media? A replication and extension in China. *Humanities and Social Sciences Communications*, 12, article number 833. <https://doi.org/10.1057/s41599-025-05233-9>
- Maati, A., Edel, M., Saglam, K., Schlumberger, O., Sirikupt, C. (2024). Information, doubt, and democracy: how digitization spurs democratic decay. *Democratization*, 31(5), pp. 922-942. <https://doi.org/10.1080/13510347.2023.2234831>
- Maffei, D., Moscon, V. (2024). Hate speech and fake news in Trentino-South Tyrol: an underestimated phenomenon?. [online]. Available at: <https://doi.org/10.57708/bvla9xe5brjmobqvsgkvcmw>. [Accessed 17.04. 2025].
- Marques, T. (2023), The Expression of Hate in Hate Speech. *J Appl Philos*, 40(5), pp. 769-787. <https://doi.org/10.1111/japp.12608>
- Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online*. [online]. Available at: <https://www.posiel.com/wp-content/uploads/2016/08/Media-Manipulation-and-Disinformation-Online-1.pdf>. [Accessed 17.04. 2025].
- Matamoros-Fernández, A., & Farkas, J. (2021). Racism, Hate Speech, and Social Media: A Systematic Review and Critique. *Television & New Media*, 22(2), pp. 205-224. <https://doi.org/10.1177/1527476420982230>

- Matasick, C., Alfonsi, C., Bellantoni, A. (2020). Governance responses to disinformation: How open government principles can inform policy options. OECD Working Papers on Public Governance No. 39. Paris: OECD Publishing, 46 p, <https://doi.org/10.1787/d6237c85-en>
- Mathew, B., Illendula, A., Saha, P., Sarkar, S., Goyal, P., Mukherjee, A. (2020). Hate begets Hate: A Temporal Study of Hate Speech, In: Nichols, J. (ed.) Proceedings of the ACM on Human-Computer Interaction. New York: Association for Computing Machinery, pp. 1-24. <https://doi.org/10.1145/3415163>
- McBride, M. K., Faber, P. G., Haney, K., Kannapel, P. J., Plapinger, S. (2023). Evidence-Based Techniques for Countering Misinformation, Disinformation, and Malinformation, Center for Naval Analyses. [online]. Available at: <https://coilink.org/20.500.12592/4b8h0px>. [Accessed 17.04. 2025].
- McConnell, M. W. (1992). America's First "Hate Speech" Regulation. *Constitutional Commentary*, 739(9), pp. 17-23.
- Mihálik, J., Garaj, M., Bardovič, J. (2022). Social Media and Political Communication of Youth Political Organisations in Slovakia, Czechia and European Level: A Cross-Case Analysis. *Social Sciences*, 11(2), article ID 69. <https://doi.org/10.3390/socsci11020069>
- Mihalková, H. (2023). Nvotovej poškodili auto, kolega zase dostal päťou. Ľuďom sa zdá, že môžu všetko, hovorí o útokoch v kampani politológ. [online]. Available at: <https://spravy.pravda.sk/parlamentne-volby-2023/clanok/679333-utoky-na-politikov-predvolebna-kampan-vyhrazky/>. [Accessed 17.04. 2025].
- Morse, J. C., Pratt, T. (2025). Information Disorder and Global Politics. *International Organization*, 79(S1), pp. 26-S43. <https://doi.org/10.1017/S0020818325101069>
- Nagasako, T. Global disinformation campaigns and legal challenges. *International Cybersecurity Law Review*, 1, pp. 125-136. <https://doi.org/10.1365/s43439-020-00010-7>
- Nai, A., van Erkel, P. F. A., Bos, L. (2025). Violence Against Politicians Drives Support for Political Violence Among (Some) Voters: Evidence from a Natural Experiment. *Public Opinion Quarterly*, 2025, article ID nfaf010. <https://doi.org/10.1093/poq/nfaf010>
- Nenadić, I. (2019). Unpacking the "European approach" to tackling challenges of disinformation and political manipulation. *Internet Policy Review*, 8(4), pp. 1-22. <https://hdl.handle.net/10419/214095>

- Neubert, K. (2024). Political violence in Germany on the rise. [online]. Available at: <https://www.euractiv.com/section/politics/news/political-violence-in-germany-on-the-rise/>. [Accessed 17.04. 2025].
- Olan, F., Jayawickrama, U., Arakpogun, E. O., Suklan, J., Liu, S. (2024). Fake news on Social Media: the Impact on Society. *Information Systems Frontiers*, 26, pp. 443-458. <https://doi.org/10.1007/s10796-022-10242-z>
- Palani, B., Elango, Viswanathan, V. K. (2022). CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. *Multimedia Tools and Applications*, 81, pp. 5587-5620. <https://doi.org/10.1007/s11042-021-11782-3>
- Paz, M. A., Montero-Díaz, J., Moreno-Delgado, A. (2020). Hate Speech: A Systematized Review. *SAGE Open*, 10(4), pp. 1-12. <https://doi.org/10.1177/2158244020973022>
- Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, 25(5), pp. 388-402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Pérez-Escolar, M., Lilleker, D., Tapia-Frade, A. (2023). A Systematic Literature Review of the Phenomenon of Disinformation and Misinformation. *Media and Communication*, 11(2), pp. 76-87. <https://doi.org/10.17645/mac.v11i2.6453>
- Pherson, R. H., Mort Ranta, P., Cannon, C. (2021). Strategies for Combating the Scourge of Digital Disinformation. *International Journal of Intelligence and Counter Intelligence*, 34(2), pp. 316–341. <https://doi.org/10.1080/08850607.2020.1789425>
- Pleios, G., Tastsoglou, M. (2025). From Fake News to Hate Speech and Vice Versa: The Logic of Polarization in Information Society. In: Çoban, S., İnceoğlu, Y. G. (eds.) *Media, Populism and Hate Speech*. Leiden: Brill, pp. 229-247. https://doi.org/10.1163/9789004721326_013
- Posetti, J., & Matthews, A. (2018). A short guide to the history of 'fake news' and disinformation. [online]. Available at: https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf [Accessed 17.04. 2025].
- Prike, T., Ecker, U. K. H. (2023). Effective correction of misinformation. *Current Opinion in Psychology*, 54, article ID 101712. <https://doi.org/10.1016/j.copsyc.2023.101712>

- Rød, B., Pursiainen, Ch., Eklund, N. (2025). Combatting Disinformation – How Do We Create Resilient Societies? Literature Review and Analytical Framework. *European Journal for Security Research*, 2025. <https://doi.org/10.1007/s41125-025-00105-4>
- Ručinská, S., Fečko, M., Mital', O. (2022). The role of public authorities in responding disinformation. In *Proceedings of the central and eastern European eDem and eGov days*. New York: Association for Computing Machinery, pp. 111-117. <https://doi.org/10.1145/3551504.3552297>
- Ručinská, S., Fečko, M., Mital', O. (2023). Trust in public institutions in the age of disinformation. In: *Proceedings of the central and eastern European eDem and eGov days 2023*. New York: Association for Computing Machinery, pp. 12-17. <https://doi.org/10.1145/3603304.3604075>
- Ručinská, S., Mital', O., Fečko, M. (2025). Disinformation in Public Policy: The Case of Slovakia. In: Urs, N., Špaček, D., Nõmmik, S. (eds) *Digital Transformation in European Public Services. Governance and Public Management*. Palgrave Macmillan, Cham, pp. 155-186. https://doi.org/10.1007/978-3-031-81425-9_8
- Sádaba, Ch., & Salaverría, R. (2023). Tackling disinformation with media literacy: analysis of trends in the European Union. *Revista Latina de Comunicación Social*, 82, pp. 17-33. <https://www.doi.org/10.4185/RLCS-2023-1552>
- Salaverría, R., Cardoso, G. (2023). Future of disinformation studies: emerging research fields. *Profesional de la información*, 32(5), article ID e320525. <https://doi.org/10.3145/epi.2023.sep.25>
- Saleh, H., Alhothali, A., Moria K. (2023). Detection of Hate Speech using BERT and Hate Speech Word Embedding with Deep Model. *Applied Artificial Intelligence*, 37(1), article ID 2166719. <https://doi.org/10.1080/08839514.2023.2166719>
- Salman, S., Shamsi, J. A., Qureshi, R. (2023). Deep Fake Generation and Detection: Issues, Challenges, and Solutions. *IT Professional*, 25(1), pp. 52-59. <https://doi.org/10.1109/MITP.2022.3230353>
- Senft, T.M., Greenfield, S. (2023). People's Experience of Information Overload and Its Impact on Infodemic Harms. In: Purnat, T.D., Nguyen, T., Briand, S. (eds.) *Managing Infodemics in the 21st Century*. Cham: Springer, pp. 27-40. https://doi.org/10.1007/978-3-031-27789-4_3
- Shu, K. (2022). Combating disinformation on social media: A computational perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2(1), article ID 100035. <https://doi.org/10.1016/j.tbench.2022.100035>

- Schirch, L. (2021). Digital information, conflict and democracy. In: Schirch, L. (ed.) *Social Media Impacts on Conflict and Democracy: The Tectonic Shift*. London: Routledge, pp. 21-42. <https://doi.org/10.4324/9781003087649>
- Schmid, U. K., Kümpel, A. S., Rieger, D. (2022). How social media users perceive different forms of online hate speech: A qualitative multi-method study. *New Media & Society*, 26(5), pp. 2614-2632. <https://doi.org/10.1177/14614448221091185>
- Simion, M. (2023). Knowledge and Disinformation. *Episteme*, 21(4), pp. 1-12. DOI: <https://doi.org/10.1017/epi.2023.25>
- Sokol, P., & Bachňáková Rózenfeldová, L. (2025). Content blocking mechanism in cybersecurity: Slovakia case study. *EURASIP Journal on Information Security*, 2025, article n. 4. <https://doi.org/10.1186/s13635-025-00190-x>
- Sommarberg, S., Ketola, M., Huhtinen, A., Sederholm, T. (2024). Counterforces and Detection of Disinformation. In: Uusikylä, P., Jalonen, H., Jokipii, A. (eds.) *Information Resilience and Comprehensive Security. Information Technology and Global Governance*. Cham: Palgrave Macmillan, pp. 181-204. https://doi.org/10.1007/978-3-031-66196-9_9
- Stroud, L. (2024). Violence against German politicians almost doubles in five years, data shows. [online]. Available at: <https://www.euronews.com/2024/05/17/violence-against-german-politicians-almost-doubles-in-5-years>. [Accessed 17.04. 2025].
- Szczepański, M., Pawlicki, M., Kozik, R., Choraś, M. (2021). New explainability method for BERT-based model in fake news detection. *Scientific Reports*, 11, article n. 23705, pp. 1-13. <https://doi.org/10.1038/s41598-021-03100-6>
- Tagesschau. (2025). Rekordwert bei politisch motivierter Kriminalität. [online]. Available at: <https://www.tagesschau.de/inland/innenpolitik/bilanz-politisch-motivierte-straftaten-dobrindt-innenministerium-100.html>. [Accessed 17.04. 2025].
- Tan, C. (2022). Regulating disinformation on Twitter and Facebook. *Griffith Law Review*, 31(4), pp. 513–536. <https://doi.org/10.1080/10383441.2022.2138140>
- Tandoc, E. C., Thomas, R. J., Bishop, L. (2021). What Is (Fake) News? Analyzing News Values (and More) in Fake Stories. *Media and Communication*, 9(1), pp. 110-119. DOI: <https://doi.org/10.17645/mac.v9i1.3331>

- Tandoc, E.C., Seet, S. K. (2024). War of the Words: How Individuals Respond to “Fake News,” “Misinformation,” “Disinformation,” and “Online Falsehoods.” *Journalism Practice*, 18(6), pp. 1503-1519. <https://doi.org/10.1080/17512786.2022.2110929>
- Tenove, C. (2020). Protecting Democracy from Disinformation: Normative Threats and Policy Responses. *The International Journal of Press/Politics*, 25(3), pp. 517-537. <https://doi.org/10.1177/1940161220918740>
- Tontodimamma, A., Nissi, E., Sarra, A., Fontanella, L. (2021). Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics*, 126, pp. 157-179. DOI: <https://doi.org/10.1007/s11192-020-03737-6>
- United Nations. (2025). Countering Disinformation. [online]. Available at: <https://www.un.org/en/countering-disinformation>. [Accessed 17.04.2025].
- United Nations. (2025). Hate speech is rising around the world. [online]. Available at: <https://www.un.org/en/hate-speech> [Accessed 17.04.2025].
- Vafeiadis, M., Bortree, D. S., Buckley, Ch., Diddi, P., Xiao, A. (2020), Refuting fake news on social media: nonprofits, crisis response strategies and issue involvement. *Journal of Product & Brand Management*, 29(2), pp. 209-222. <https://doi.org/10.1108/JPBM-12-2018-2146>
- Vasist, P. N., Chatterjee, D., Krishnan, S. (2024). The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-country Configurational Narrative. *Information Systems Frontiers*, 26, pp. 663-688. <https://doi.org/10.1007/s10796-023-10390-w>
- Ververis, V., Lasota, L., Ermakova, T., Fabian, B. (2024). Website blocking in the European Union: Network interference from the perspective of Open Internet. *Policy & Internet*, 16, pp. 121–148. <https://doi.org/10.1002/poi3.367>
- Vese D. (2022). Governing Fake News: The Regulation of Social Media and the Right to Freedom of Expression in the Era of Emergency. *European Journal of Risk Regulation*, 13(3), pp. 477-513. <https://doi.org/10.1017/err.2021.48>
- Waldron, J. (2012). *The Harm in Hate Speech*. Cambridge: Harvard University Press. 264 p. ISBN 978-0-674-065-086. <https://doi.org/10.4159/harvard.9780674065086>