Central European Journal of Communication

Scientific Journal of the Polish Communication Association

Volume 18 Number 3 (41) Fall 2025

CENTRAL EUROPEAN JOURNAL OF COMMUNICATION (ISSN 1899-5101)
WWW.CEJC.PTKS.PL

EDITORS

Editor-in-Chief:
Agnieszka Stępińska, Adam Mickiewicz University, Poznań, Poland
Executive Editor:
Bartłomiej Łódzki, University of Wrocław, Poland

EDITORIAL TEAM

Kinga Adamczewska, Adam Mickiewicz University, Poznań, Poland
Denis Halagiera, Adam Mickiewicz University, Poznań, Poland
Roksana Gloc, Jagiellonian University, Cracow, Poland
Agnieszka Węglińska, University of Wrocław, Poland

ASSOCIATE EDITORS

Delia C. Balaban, Babes-Bolyai University, Cluj-Napoca, Romania
Carlos Edmundo Arcila Calderón, University of Salamanca, Spain
Renata Matkevičienė, Vilnius University, Lithuania
Cristina Monzer, University of Pennsylvania, United States of America
Jakob Ohme, Weizenbaum Institute, Berlin, Germany
Lada Trifonova Price, University of Sheffield, United Kingdom
Sergio Splendore, University of Milan, Italy.

SCIENTIFIC NETWORK

List of members of Scientific Network is available on cejc.ptks.pl

Central European Journal of Communication

Volume 18 | No 3 (41) | Fall 2025

Scientific Journal of the Polish Communication Association

Table of Contents

304 *Editors' Introduction*
Agnieszka Stępińska, Bartłomiej Łódzki

307 *A 'Safe Space' for Disagreement?: An Experiment on the Effect of Social Media Cross-Cutting Exposure on Internal Political Efficacy*
Márton Bene, Veronika Patkós

329 *Who Shouts the Loudest? Predictors of Conflict-Oriented Behavior on Social Network Sites*
Martina Novotná, Alena Pospíšil Macková, Lucie Čejková

349 *Navigating Narratives: The Dynamics of Disinformation Exposure and Believability to Information in Kosovo*
Dren Gërguri, Darren Lilleker

371 *Mobilizing Distrust: Persuasive Strategies in Czech Disinformation Email Subject Headings*
Todd Nesbitt, Tess Slavíčková, Veronika Zavřelová

391 *Deep Symmetrism? A Populist Strategy for Political Communication and Beyond*
Jacek Kołodziej

410 *Can Disinformation be Regulated? A Comprehensive Overview of the European Union's Pertaining Initiatives*
Gergely Ferenc Lendvai, János Tamás Papp, Tamás Attila Szikora, Krzysztof Wasilewski

427 *Unveiling the Many Faces of Fact-Checking: State of the Art of Academic Research on Information Correction.*
Maia Klaassen, Ianis Bucholtz, Marju Himma, Krista Lepik-Verliin

447 *Towards a Multi-Level Model of Resilience to Information Disorders: A Systematic Literature Review*
Dmytro Iarovyi, Ragne Kõuts-Klemm, Sten Torpan, Kristina Juraitė

478 *Book Reviews*

487 *Events*

490 *Notes on Contributors*

DOI: 10.51480/1899-5101.18.3(41).1011

Editors' Introduction

In 2025, artificial intelligence significantly affected the media industry, automating editorial workflows and introducing hyper-personalized content. Many newsrooms have implemented or streamlined the use of AI to generate news, analyze data, and create personalized feeds. Publisher actions stemmed indirectly from changing audience behavior. The trend of deliberately disconnecting from news has deepened, forcing media outlets to seek new forms of engagement through algorithm-based personalization. These changes have increased efficiency but also raised concerns about disinformation. The year 2025 was also marked by the intensive implementation of the European Artificial Intelligence Act (AI Act) and the Digital Services Act (DSA).

In the *Central European Journal of Communication* FALL 2025 issue, our authors undertake an analysis of current problems and challenges. In particular, the authors focus on the challenges posed by disinformation and misinformation, including how people are exposed to it, why they believe it, and how difficult it is to counter or regulate it. They analyze the difficulty of maintaining healthy democratic communication, especially in environments shaped by conflict-oriented behavior, polarization, and political disagreement on social media. The research concentrates on the limited public capacity to critically evaluate information, due to low media literacy, low verification practices, and cognitive biases that reinforce belief in false or preferred narratives. The authors focus on the struggle to design effective institutional, regulatory, and societal responses to information disorders, where existing mechanisms (e.g., EU policies, fact-checking, and literacy programs) are often insufficient or lack enforcement. Finally, they research ethical and societal challenges emerging from new communication technologies, including issues of privacy, consent, trust, and shifts in social norms – especially in mediated or digital forms of interaction.

Márton Bene and Veronika Patkós' question whether the „dark side" of political disagreement – making people uncertain about their views – also diminishes their sense of political competence in a social media environment. The findings suggest it does not, positioning social media as a potentially „safe space" for such discourse.

The research of Martina Novotná, Alena Pospíšil Macková, and Lucie Čejková addresses the need to understand what drives online conflict, given its significant potential to shape public opinion, increase societal polarization, and impact deliberative democracy. It finds that younger, less educated men with low political interest are highly active in these conflicts.

Dren Gërguri and Darren Lilleker, based on the Kosovo case, analyze individuals' exposure to disinformation and the factors influencing their belief in it. The study finds that repeated exposure, along with cognitive and situational factors, is a more significant predictor of belief than trust in traditional news media.

Todd Nesbitt, Tess Slavíčková, and Veronika Zavřelová examined the persuasive function of subject headings in Czech-language disinformation chain emails. Their study focuses on the first six months of the full-scale Russian invasion of Ukraine in 2022 and analyzes a corpus of email subject headings collected by the civic initiative Čeští Elfové (Czech Elves). Drawing on critical discourse analysis, relevance theory, and presupposition theory, the article explores how subject headings establish relevance, mobilize shared assumptions, and encourage recipients to access and disseminate disinformation content.

Jacek Kołodziej focuses on the issue of „deep symmetrism." The article addresses a gap in academic literature by defining this phenomenon as a mechanism of political struggle. Deep symmetrism is described as a system of duplicating social institutions („doppelgangering") to orchestrate political influence, linking it to modern propaganda and disinformation, with the 2023 Polish election serving as a case study.

A critical analysis of the European Union's regulatory initiatives to combat disinformation by Gergely Ferenc Lendvai, János Tamás Papp, Tamás Attila Szikora, and Krzysztof Wasilewski sheds new light on these issues. While the EU has created frameworks like the Digital Services Act to improve platform accountability, these measures are identified as being largely voluntary and lacking sufficient enforcement mechanisms. The paper concludes that regulation alone is insufficient to solve the problem of disinformation.

The final research by Maia Klaassen, Ianis Bucholtz, Marju Himma, and Krista Lepik-Verliin presents a comprehensive review of academic research on fact-checking from 2010 to 2023. The rapid expansion of research on fact-checking has made it difficult to get a complete overview of the field. This review of 675 articles maps the research landscape and identifies key gaps.

The study in Methods & Concepts section addresses the need for a systematic, evidence-based framework to understand and enhance societal and individual resilience to information manipulation. Dmytro Iarovyi, Ragne Kõuts-Klemm, Sten Torpan, and Kristina Juraitė present a three-dimensional model (macro, meso, and micro levels) to serve as a foundation for future research and practical interventions.

We are also returning to a significant event: the Media and Democracy Karol Jakubowicz Award 2025. Marius Dragomir – the editor of the *Media Capture Monitoring Report* project – received the Award for the *Media Capture Monitoring Report*, a research report by the Media and Journalism Research Center and the International Press Institute (co-authored with Zsuzsa Detreköi). A special Media and Democracy Karol Jakubowicz Award 2025 was awarded to Basil Kerski for his book *Europejczycy z kantonu Polska* (*Europeans from the Canton Poland*; Cracow; Budapest; Syracuse: Austeria Press, 2024).

In this issue, we are introducing a new editorial team to the *Central European Journal of Communication* for the term 2025–2028, as well as a new group of associate editors.

We would like to express our gratitude to the former editor-in-chief, Michał Głowacki, for his clear vision of the scientific journal's role in fostering international collaboration within the academic community. We would also like to thank Damara Sidyk-Furman and Jacek Mikucki for their tremendous work in the last 6 years. Finally, we thank all the associate editors who supported the journal during the previous term (2022–2025): Catherine Johnson, Kristina Juraitė, Marcus Kreutler, Päivi Maijanen-Kyläheiko, Marco Mazzoni, Dariya Orlova, Gabriella Szabó, Dren Gërguri, Dina Vozab, and Bissera Zankova.

We welcome new editorial team members, Denis Halagiera and Roksana Gloc. We are looking forward to our cooperation with scholars who accepted our invitation to join the CEJC as associate editors for the next three years: Delia Cristina Balaban, Carlos Edmundo Arcila Calderón, Renata Matkevičienė, Cristina Monzer, Jakob Ohme, Lada Trifonova Price, and Sergio Splendore.

Agnieszka Stępińska
ADAM MICKIEWICZ UNIVERSITY, POZNAŃ, POLAND

Bartłomiej Łódzki
UNIVERSITY OF WROCŁAW, POLAND

ORIGINAL ARTICLE

DOI: 10.51480/1899-5101.18.3(41).673

A 'Safe Space' for Disagreement?: An Experiment on the Effect of Social Media Cross-Cutting Exposure on Internal Political Efficacy

Márton Bene

0000-0003-0177-9717

ELTE Centre for Social Sciences, Budapest, Hungary

ELTE Eötvös Loránd University, Budapest, Hungary

Veronika Patkós

0000-0003-3636-6357

ELTE Centre for Social Sciences, Budapest, Hungary

ELTE Eötvös Loránd University, Budapest, Hungary

Abstract: We investigate whether cross-cutting and mixed exposure to political information on social media is associated with users' internal political efficacy using an online survey experiment. Existing research has demonstrated several benefits of cross-cutting exposure, but it has also shown its dark side. Mutz (2006) found that political ambivalence is significantly associated with cross-cutting talk as disagreement makes people more uncertain about their political views. However, it has not been investigated if cross-cutting exposure makes people more uncertain about their own political capacities to understand politics and meaningfully participate within it. Our research question is whether this detrimental effect on internal political efficacy can be detected in the social media context. The findings show that participants' internal political efficacy is not significantly shaped by the type of exposure. Consequently, cross-cutting exposure on social media can be seen as a 'safe space' for political disagreement where the 'dark side' of cross-cutting exposure cannot prevail.

Keywords: cross-cutting exposure; internal political efficacy; social media; experimental research; health care system

INTRODUCTION

People have natural tendencies to avoid cognitive dissonance and intentionally seek attitude-consistent information to reinforce that their political views are right and valid (Festinger, 1957). They usually socialize with people who are like themselves (McPherson et al., 2001) and consume media content that echoes their pre-existing views (Iyengar & Hahn, 2009). This is reasonable as studies showed that cross-cutting exposure has the potential to make people more uncertain about their own opinions (Mutz, 2006), and ambivalence can lead to the feeling of psychological discomfort (van Harreveld et al. 2009). The emergence of social media as a political information resource, however, has resulted in a rather diverse information environment. Research demonstrated that people are frequently exposed to cross-cutting political content on these platforms (Barnidge, 2017; Beam et al, 2018), which fact may induce a feeling of political uncertainty.

While existing studies focused on uncertainty about specific political attitudes (political ambivalence), we generalize this theory and test the proposition that cross-cutting exposure can make people uncertain in their own political capacities to comprehend and act meaningfully in the political sphere, a feeling that is commonly conceptualized as internal political efficacy (Niemi et al., 1991). This is a truly important question since internal political efficacy is a key antecedent variable of political behavior (Pollock III, 1983): if the heightened cross-cutting exposure experienced in social media results in less efficacious people, it would have wider detrimental consequences on political participation. Although existing research showed that cross-cutting exposure can increase political ambivalence by making people more uncertain about specific political objects such as political actors and issues (Huckfeldt et al., 2004; Mutz, 2006), and they also demonstrated that political ambivalence and internal political efficacy are closely related concepts as people who are more uncertain about their political views are usually less confident in their own political capabilities (Chen & Lin, 2021; Hmielowski et al., 2018), no research tested if cross-cutting exposure is able to affect people's political self-conception. In line with these findings, we expect that people's level of internal political efficacy will be lowered when they are exposed to cross-cutting content as this will make them more uncertain about their capabilities to comprehend politics and act within it in a meaningful way.

We tested this hypothesis with an experimental method in which participants (N = 365) divided into three groups were exposed to a Facebook discussion consisting of four comments about the state of the health care system in Hungary which is a highly important and controversial topic in this specific context. Our manipulation tests the effects of two different forms of cross-cutting exposure compared to full like-minded exposure: we differentiated between 'full cross-cutting' and 'mixed cross-cutting' exposure. In the case of the former, only disagreeable

comments are shown, while in the latter both like-minded and conflicting opinions are listed. Based on the results, our hypothesis is rejected: participants' internal political efficacy is not affected by the type of Facebook discussion they are exposed to. They do not become more skeptical about their own political capacities when their views are challenged, and similarly, the like-minded opinions do not make them more confident either. These findings imply that, from this aspect, social media may be a 'safe space' for disagreement as the uncertainty-inducing effect of cross-cutting exposure does not prevail on these platforms.

THEORETICAL BACKGROUND

INTERNAL POLITICAL EFFICACY IN A MEDIATIZED INFORMATION CONTEXT

The concept of political efficacy is a construct to capture voters' subjective attitudes towards their role within the overall political system (Campbell, 1954). It is usually broken into two components where external political efficacy is related to perceptions about the responsiveness of the political system, while internal political efficacy is more about political self-conceptions and refers to "beliefs about one's own competence to understand, and to participate effectively in politics" (Niemi et al., 1991: 1407). Since the development of the concept, numerous research have demonstrated that internal political efficacy is a key variable in political behavior, and the way people think of their own personal political capacities can shape their willingness to participate in political activities. Lower political self-confidence is a barrier to voting (Pollock III, 1983), protest (Christensen, 2016) and political discussion activity (Gil de Zúñiga et al., 2017).

Although political efficacy is traditionally treated as a static and enduring political attitude that is deeply ingrained in individuals' personality and social-economic status (SES) (see Campbell, 1954; Easton & Dennis, 1967), subsequent research demonstrated that mediatized experiences also shape its level. In his classic study, Robinson (1976) found that television-based news consumption decreased the level of political efficacy, but later studies mostly found positive effects (Ardèvol-Abreu, 2019; Moeller et al., 2014; Newhagen, 1994). Internet use is also demonstrated to have a beneficial impact on internal political efficacy (Lee, 2006; Boulianne et al., 2023), and there is evidence that political activities on social network sites can also increase political self-confidence (Chan et al., 2017; Halpern et al., 2017). While these relationships are not always direct (Ardèvol-Abreu, 2019, Chan et al., 2017) and uniform across media (Moeller et al., 2014) and news type (Lee, 2006; Kenski & Stroud, 2006), these findings suggest that the mediatized information people are exposed to can shape their political self-conception.

SOCIAL MEDIA AS A SPACE OF CROSS-CUTTING EXPOSURE

Social media platforms are now key information resources for many voters (Newman et al., 2021). However, the nature of political information exposure on these platforms is different from what is dominant in the case of news media. While media consumption in the high-choice media environment is largely driven by selective exposure (see, Iyengar & Hahn, 2009), it is demonstrated that social media platforms, especially Facebook, are actually more heterogeneous environments than other political information contexts: people are frequently exposed to political content which are against their political preferences (e.g., Barnidge, 2017; Beam et al, 2018; Cardenal et al, 2019). Studies showed that accidental exposure is a common experience on social media platforms as people frequently see political content without wanting to be exposed to it (Goyanes & Demeter, 2020; Valeriani & Vaccari, 2016). The inadvertent nature of this exposure, however, does not mean that accidently seeing political content would be ineffective. Research demonstrated that people still often pay attention to this content (de Zúñiga et al, 2021; Karnowski et al, 2017, Nanz et al., 2025), and they are also able to shape recipients' political behavior and views (Kim et al, 2013; Valeriani & Vaccari, 2016).

Nonetheless, the heterogeneous political environment of social media sites seems to be counter-intuitive as both algorithms and affordances support the idea of an information-context which is in line with users' true preferences. Carefully-crafted algorithms curate the News Feeds, i.e., the central interface where people are exposed to content, to adjust the supply of posts to the users' true interests and needs as much as possible based on their continuously monitored activities. Also, social media platforms offer several affordances for users to customize and personalize their News Feed: they can freely add and delete content-providers (friends and pages) and make different filtration actions in relation to specific content to indicate if they do not prefer them.

Regarding the content-filtering algorithms, it seems that they give larger weight to personal ties and interactions, and it shows content posted or shared by friends even if it is confronted with the particular users' preferences and interests (Bucher, 2012; DeVito, 2017). Also, studies found that peer-cues in the form of social endorsement or personal mediation can overwrite political and interest-based selectivity in news selection, and people click on cross-cutting items if they are highly recommended by peers or shared by personal friends (Anspach, 2017; Messing & Westwood, 2014). For these reasons, people may often read or click cross-cutting content that restricts the ability of the algorithms to find out the users' true preferences. Indeed, studies showed that on Facebook – whose platform applies the strongest algorithmic filtering (Bossetta, 2018) – the algorithm's contribution to the politically more homogenous information environment was marginal (Bakshy, 2015; Guess et al., 2023).

The relative heterogeneity of social media platforms can be also attributed to the underutilization of the affordances offered by these sites to personally curate the News Feed. People intensively and proactively create connections based on their social needs and interests, but the ease of connection building entails the proliferation of weak ties (Kim & Chen, 2015; Raine & Wellman, 2012) which is a major driver of heterogeneity (Granovetter, 1973). At the same time, studies showed that it is rather uncommon to use the connection-breaking or content-filtering affordances for political reasons, and it is only the politically active and partisan people who occasionally apply them to homogenize their own information environment (John & Dvir-Gvirsman, 2015; Zhu et al, 2017). Overall, for these reasons, it seems that cross-cutting exposure is an everyday experience on social media platforms as they create a rather diverse information environment for people.

THE EFFECT OF CROSS-CUTTING EXPOSURE ON INTERNAL POLITICAL EFFICACY

Heightened cross-cutting exposure experienced in the social media context, however, may have implications for individuals' political self-conception. While democracy theory has long argued for the importance and democratic benefits of cross-cutting exposure (Fishkin, 1995; Gutmann & Thompson, 1998), Diana Mutz (2006) in her seminal study raised attention to its 'dark side', too. She argued that cross-cutting exposure can negatively affect people's willingness to participate in politics because it may make them more uncertain about their own political views. She demonstrated that people who are more frequently exposed to cross-cutting political information are more ambivalent about their political attitudes because conflicting views can increase the inner uncertainty about specific political issues or actors. While the dilemma of 'deliberative versus participatory democracy' raised in her book has been mitigated or even refuted by several subsequent research showing a null (Eveland & Hively, 2009; Nir, 2011, Matthes et al., 2019) or even positive effect (Huckfeldt et al., 2004; Kwak et al., 2005; Scheufele et al., 2006) of cross-cutting exposure on political participation, the related ambivalence-argument remained largely unchallenged (see, Matthes et al., 2019). In fact, the small number of studies investigating the links between cross-cutting exposure and political ambivalence reinforced Mutz's findings (Huckfeldt et al., 2004; Lee, 2012, Chen & Lin, 2021).

Internal political efficacy is strongly related to political ambivalence both conceptually and empirically. While political ambivalence refers to the level of subjective uncertainty about specific political objects (actors or issue), internal political efficacy is also related to subjective political uncertainty but in a more general sense, as it is the (un)certainty about one's own political capabilities of meaningful political activities. It is unsurprising, therefore, that research

found strong association between the two constructs: people who are more ambivalent about specific political objects have generally lower internal political efficacy as well (Chen & Lin, 2021; Hmielowski et al., 2018; McGraw & Bartels, 2005; Lee & Chan, 2009).

For these reasons, we may expect that cross-cutting exposure affects internal political efficacy in a manner similar to political ambivalence. We argue that contrasting views make people more uncertain not only about specific subjects, but also about their own capacities to comprehend and orient themselves in politics. Interestingly, to our knowledge no research has investigated the effect of cross-cutting exposure on internal political efficacy. A closely related finding (Zhang & Chang, 2014) tested how perceived disagreement in everyday talk situations affects internal political efficacy in Singapore, but everyday talk is different from pure exposure to cross-cutting content. Further, another study focused on the effect of political context heterogeneity on efficacy among the minority population in Finland, but context was measured by macro-level factors such as the strength of the minority party in the municipal council and the proportion of the minority population in the municipality (Karv et al., 2021). Anyway, these two studies yield conflicting findings: while the former study found no effect, the latter showed that the level of internal political efficacy is lower for people who live in a more heterogenous context. Nonetheless, the distinct effect of cross-cutting exposure remained unexplored. This is a surprising gap given the fact that Robinson in his early study has already raised the possibility that the negative effect of television viewing on political efficacy can be the product of exposure to information that opposes existing beliefs which leads to people 'questioning their own ability to cope with or comprehend politics' (Robinson, 1976: 417). While this idea was ignored by subsequent research, since cross-cutting exposure is now an everyday experience for people on social media, it is reasonable to revive it. Therefore, in this research we test the following hypothesis:

- *H1. Cross-cutting exposure will decrease individuals' internal political efficacy*

We differentiate between two forms of cross-cutting exposure. First, the ideal type of cross-cutting exposure is when people are exposed only to information that oppose to their views. This can be labelled as 'full cross-cutting exposure'. However, given the heterogeneity of the Facebook information environment, it is more realistic that people are exposed to both like-minded and cross-cutting content at the same time. While in line with our theoretical reasoning, we expect that both type of cross-cutting exposure will decrease internal political efficacy by making people more insecure in their own political capacities, it is reasonable to test the effect of the two conditions separately.

DATA AND METHOD

DESIGN OF THE STUDY

We tested our hypothesis in an online survey experiment conducted in Hungary with 423 respondents; of these, *N* = 365 completed all items in the political efficacy battery used for the main analysis. The experiment was conducted by the Kutatócentrum polling company in 2018, approximately one month after the last national elections (between 25 May and 22 June). Participants were recruited using quotas to represent the Hungarian population with internet access in terms of gender, education, region, and type of settlement. The final sample, however, was somewhat more educated (53% had at least upper-secondary education) and included slightly more women (56%) than the general population. Our experimental condition focused on the topic of the state of the healthcare system; respondents were exposed to different opinions on this issue. At the time, this topic was perceived as the most important national issue in Hungary (Eurobarometer, 2018), and was regularly discussed in both mass and social media in the months preceding the experiment (Bene, 2021; Dobos et al., 2021). Research shows that pro-government and opposition voters were highly polarized in their views on the healthcare system: pro-government voters were generally satisfied with its functioning, while opposition voters were more likely to express dissatisfaction (Political Capital, 2017).

Respondents first completed a short pre-test survey, and then read a Facebook thread consisting of four comments expressing opinions on the state of the healthcare system in Hungary. In order to enhance the ecological validity (i.e., generalizability to real-life settings) of the experiment (see Osborne-Crowley, 2020), the posts used as experimental stimuli were taken from real online discussions. This approach not only authentically reflects the language of online discourse, but also presents the different positions within their actual discursive context, including the discursive markers associated with each stance. This makes it easier for participants to situate their own views in relation to the comments. The manipulation check (see below) confirmed that this approach was successful. Afterward, they answered a post-test survey that included questions about political efficacy, political news consumption habits, and other political attitudes. Participants were randomly assigned to one of three groups, each receiving a different treatment in the form of a Facebook thread. These threads were artificially constructed by the research team to reflect the tone and language of real online debates on the topic (see Appendix for the original texts and their English translations). To avoid potential confounding effects related to gender, age, or attractiveness, all names, profile photos, and comment dates

were blurred. The arguments presented in the threads varied by treatment group, as described below.

- *ProGov Group:* Four comments praised the government's handling of the healthcare system, including some criticism of the performance of previous administrations.
- *AntiGov Group:* Four comments criticized the state of the healthcare system and the performance of the current government in managing it.
- *Mixed Group:* A mix of four comments: two comments from the Pro-Gov group and two from the Anti-Gov group.

We conducted one-way ANOVAs to assess whether randomization produced balanced groups across key pre-treatment variables. No significant differences were found for age, $F(3, 438) = 0.49$, $p = .692$; education level, $F(3, 439) = 0.85$, $p = .465$; or left-right self-placement, $F(2, 336) = 0.52$, $p = .597$. These results suggest that the randomization procedure successfully created comparable groups across key demographic and political variables.

To assess the effectiveness of our manipulation, we asked respondents to describe their overall feelings in an open-ended question and to indicate their level of agreement with the content on a 0 to 10 scale. Both opposition and government partisans responded in line with expectations. Opposition partisans reported a high mean level of agreement with the negative stimuli ($M = 8.41$) and a low level of agreement with the positive stimuli ($M = 2.53$). Conversely, government partisans showed stronger agreement with the positive stimuli ($M = 6.16$) and lower agreement with the negative stimuli ($M = 4.84$). As anticipated, the mixed treatment elicited relatively balanced agreement from both groups (opposition: $M = 5.76$; government: $M = 5.53$), suggesting it have been perceived as representing opinions from both sides.

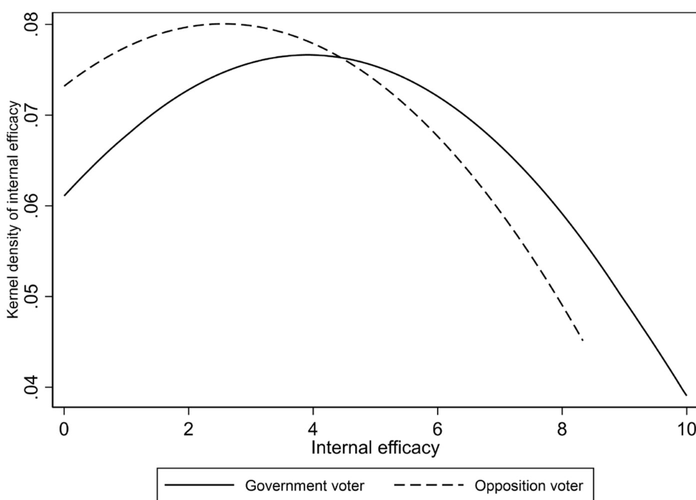
Therefore, we believe the stimuli were sufficiently strong, particularly given the intensity of emotional and cognitive engagement expressed in participants' open-ended responses. Many respondents reacted strongly—both positively and negatively—indicating that the content was not only noticed but also subjectively impactful. Given their personal involvement and prior experiences with the healthcare system, as well as the emotionally charged nature of the comments included in the stimuli, the vast majority of participants reported strong negative feelings. Most commonly, those exposed to the negative stimuli expressed sadness and anger, often stating that the comments resonated with them because they were "true." Among those who received the mixed stimuli, 23% explicitly noted that the information or their emotional reactions were mixed. These responses included expressions of ambivalence and conflicted feelings, references to specific comments they agreed with versus those they

rejected, or identification of the political affiliations behind different statements. In contrast, only 1.05% of respondents in the groups exposed to uniformly negative or positive content reported mixed feelings.

Dependent variable

Our dependent variable is composed from three items that are used for measuring internal political efficacy in the 7th round of the European Social Survey (ESS): "*How able do you think you are to take an active role in a group involved with political issues?; How confident are you in your own ability to participate in politics?; How easy do you personally find it to take part in politics?*". Responses ranged from 0 to 10 on an 11-point scale. The three items were merged into an index using their mean value. We assessed the internal consistency of the political efficacy scale using Cronbach's alpha. The scale, which included five items, demonstrated acceptable reliability (α = .792), indicating that the items measured a coherent construct. The descriptive statistics of the variable show that most people were not very confident about their political efficacy (Mean=2.96; St. D.=2.23; N=36 5). However, in line with the literature (Bene, 2020), perceptions about internal efficacy were strongly influenced by respondents' government-opposition voter status. Figure 1 shows the density of internal efficacy scores for government and opposition voters, separately. While the internal efficacy of opposition voters is mostly low, government voters' internal efficacy is much closer to a normal distribution.

Figure 1. The density of internal political efficacy scores by voters' government-opposition preference



INDEPENDENT AND CONTROL VARIABLES

We constructed our main independent variable from respondents' vote choice and the three treatment groups (ProGov, AntiGov, and Mixed). Respondents were asked who they voted for in the last election. The short time period which passed after the last elections ensures that respondents' vote choice mirrored their actual party preferences. Their answers were recoded into a dummy variable (government voters – 1; opposition voters – 0), while we excluded non-voters from the most part of the analysis.

We constructed our independent variable about the *type of exposure* combining vote choice and treatment group. *Type of exposure* is a nominal variable, with the following values: 0 – *like-minded exposure* (government voters who were exposed to ProGov arguments about health-care policy and opposition voters who were exposed to AntiGov arguments), 1 – *full cross-cutting exposure* (opposition voters who were exposed to ProGov arguments about health-care policy and government voters who were exposed to AntiGov arguments), 2 – *mixed cross-cutting exposure* (both opposition and government voters who were exposed to two AntiGov and two ProGov arguments).

In some of the regressions we controlled for important sociodemographic variables and variables related to political news consumption, such as gender (reference category is men), age (from 19 to 85), education (0 if at most secondary, 1 if at least upper-secondary), political interest (ranging from 1 to 4 with higher values indicating higher interest), the importance of Facebook as a political information source for the respondent (0 to 10 with higher values indicating more importance), and left-right ideological position (0 to 10 with higher values indicating a more rightist position). Table 1 shows the descriptive statistics of all variables.

Table 1. The descriptive statistics of the variables analyzed

Variable	Obs	Mean	Std. Dev.	Min	Max
Internal efficacy	365	2.96	2.24	0	10
Age	423	46.90	15.10	19	85
Gender	423	0.56	0.50	0	1
Government voter	306	0.36	0.48	0	1
Education level	423	0.53	0.50	0	1
Interest in politics	423	2.72	0.85	1	4
Role of Facebook in gaining political information	421	3.92	3.32	0	10
Ideological position	339	5.27	2.59	0	10

FINDINGS

First, we assessed the effect of different types of exposure on internal efficacy with OLS regressions. As reported in Table 2, the type of exposure does not significantly affect the internal efficacy of respondents: the level of internal political efficacy does not lower for those respondents who were exposed to either full cross-cutting (Mean=3.11; St. D.=2.17) or mixed cross-cutting information (Mean=2.94; St. D.=2.43) compared to those who could read like-minded information (Mean=3.09; St. D.=2.3).

Model 2 shows that the difference between the groups remains insignificant when controlling for a range of important sociodemographic variables, relevant political attitudes, and the relevance of Facebook as a source of political information for the respondent.

Table 2. OLS regressions estimating the effect of the type of exposure on internal political efficacy

	Model 1		Model 2	
	B	St. Err.	B	St. Err.
Cross-cutting exposure (Ref: Likeminded)	.021	(.333)	-.048	(.314)
Mixed exposure (Ref: Likeminded)	-.156	(.345)	-.319	(.33)
Government voter (Ref: Opposition)			1.154***	(.339)
Age			.002	(.009)
Gender			-.482*	(.273)
Education level			.218	(.273)
Interest in politics			.535***	(.187)
Role of Facebook in gaining political information			.124***	(.042)
Ideological position (Left to right)			.104*	(.061)
Constant	3.092***	(.237)	.724	(.936)
Observations	275		253	
R-squared	.001		.197	

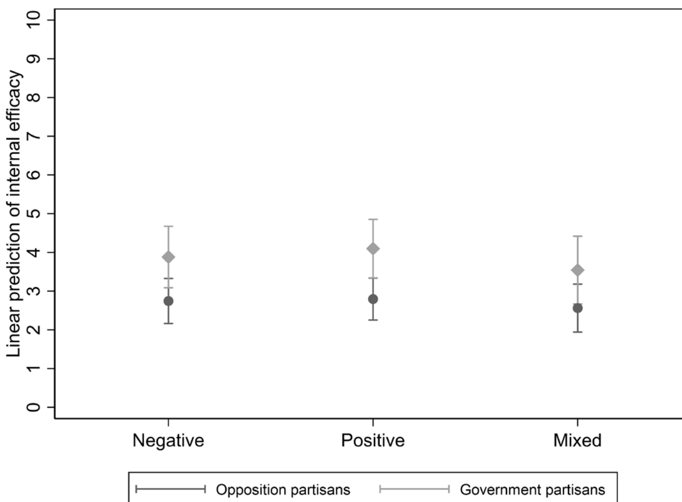
Entries are regression coefficients with standard errors in parentheses
**** p<.01, ** p<.05, * p<.1*

To further explore the question, we examined whether partisan affiliation moderates the effect of the treatments, investigating if the three types of threads—Pro-Government, Anti-Government, and Mixed—have differential effects on the internal political efficacy of government and opposition voters. This analysis builds on our earlier finding that internal efficacy is strongly influenced by respondents' partisan alignment (see Model 2 in Table 2). Regardless

of whether we consider raw differences or those adjusted for relevant control variables, the results indicate no statistically significant variation in efficacy levels among individuals exposed to likeminded, fully cross-cutting, or mixed content (see Figure 2). In all models, men, government voters, those holding a more rightist ideological position, those who considered Facebook as a more important information source for themselves and those more interested in politics proved to be more efficacious, while the type of information they received left their internal political efficacy intact.

To explore potential moderation effects, we interacted all control variables with the type of exposure respondents received, aiming to identify whether socio-economic or political characteristics moderated the impact of exposure type. The analysis revealed no substantial or statistically significant moderation effects[1]. Therefore, we conclude that no meaningful moderator variables were identified. The results imply that there is no significant relationship between the type of exposure and one's internal political efficacy.

Figure 2. The marginal effects of the interaction of partisan group affiliation and the experimental treatment on internal political efficacy, with 95% confidence intervals.



[1] The only notable finding was among respondents who reported no interest in politics and were exposed to cross-cutting content; this group exhibited significantly higher internal political efficacy compared to the other two exposure groups. However, this subgroup consisted of only six individuals, making it impossible to draw reliable generalizations.

CONCLUSION

In our contemporary social media-dominated political communication environment cross-cutting exposure to political information, which has been rather marginal due to network homophily (Mutz, 2006) and selective media-consumption patterns (Iyengar & Hahn, 2009) in the past, is an everyday experience for many citizens. While it is usually the benefits of cross-cutting exposure that is emphasized by political science literature, its 'dark side' is also demonstrated: research showed that people became more uncertain in their own views when they are exposed to political information that challenge their pre-existing beliefs (Mutz, 2006; Huckfeldt et al., 2004; Lee, 2012). Existing studies, however, focused on uncertainty only about specific political attitudes related to political actors or particular issues, namely political ambivalence. In this study, we argued that uncertainty due to exposure to conflicting views can spill over from specific political objects to the self: uncertain people will be less confident in their own political capacities to understand politics and act meaningfully in the political sphere, in other words, the level of their internal political efficacy will be lower. This thesis is also supported by the fact that studies found a strong relationship between political ambivalence and internal political efficacy (Chen & Lin, 2021; Hmielowski et al., 2018).

However, our experimental research could not confirm this hypothesis: in this specific experimental setting people's political self-conception is not changed by the type of information they are exposed to on social media. This is good news from a democratic viewpoint: social media seems to be a 'safe space' for political disagreement, the predicted detrimental effect does not occur on this platform in this specific context. Nonetheless, it is important to emphasize that we obtained this result in a single context and on a single topic, so our observation is far from conclusive. At the same time, this finding may spark interest in the topic, and future studies can investigate whether this observation can be confirmed in other contexts and on other issues as well.

Nonetheless, this tentative conclusion is in sharp contrast with findings about the ambivalence-inducing effect of cross-cutting exposure. It seems that even if cross-cutting exposure makes people more uncertain about specific political subjects, their political self-confidence remains intact on social media context. However, as our research focused only on the links between internal political efficacy and cross-cutting exposure on social media, we can only speculate if the lack of relationship is due to the differentiated effect of cross-cutting exposure on ambivalence and internal efficacy or it is because of the specific social media context. The former explanation would be that even if ambivalence and internal political efficacy are related concepts, ambivalence may be more situational and thereby more affected by the immediate information context, while internal

political efficacy is a more static attitude and formed by long-term influences rather than prompt impressions. Future research should investigate how cross-cutting exposure affects both political ambivalence and internal political efficacy to test this explanation. Another possible reason behind the contradicting findings can be related to the specific communication context. People may process cross-cutting content on social media differently than in offline context. Since exposure to conflicting views is an everyday experience here, they may be more accustomed to it, and develop a sort of immunity against it. Further, people may relate to cross-cutting exposure differently when the sources of this content are not strong ties as is the common case in offline context, but unfamiliar users as in our experimental condition. In these contexts, the well-known political psychological mechanisms of motivated reasoning (Nir, 2011) and confirmation bias (Knobloch-Westerwick et al., 2020) may work more effectively and defend users from being uncertain about their own points. These facts can make social media a 'safe space' for political disagreement, where its dark side cannot prevail. However, to confirm this thesis, future studies should investigate the effects of cross-cutting exposure on internal political efficacy in different communication context. Nonetheless, the calming lesson from our research, how the heightened level of cross-cutting exposure experienced in the social media context may not result in less efficacious people, at least one-off exposure does not make people more uncertain in their own political capacities.

Naturally, our study has several limitations beyond the aforementioned constraint of having been conducted in only one context and on a single topic. In designing the experiment, we prioritized ecological validity by selecting real-life comments as stimuli. However, this may have weakened the internal validity of the stimuli, as the differences between them are not strictly minimal. Balancing between ecological and internal validity is a common trade-off in experimental research, and in our case, it seemed justified to prioritize ecological validity in order to help respondents interpret the stimuli in relation to their own views. Nevertheless, future research would benefit from testing the same question using designs that place greater emphasis on internal validity. At the same time, it is worth noting that had the stimuli been even more similar to one another, this might actually have reduced the size of any potential effect. Yet, since we did not find a significant effect even with stimuli that may have somewhat overestimated the potential effect, this is unlikely to pose a practical problem. On the contrary, if we had found a null effect with very similar stimuli, one could argue that more realistic, real-life-like stimuli might have produced stronger effects—which is a scenario we can now rule out.

Additionally, we must also mention a general limitation of cross-sectional experimental designs: in real life, exposure is continuous—voters encounter cross-cutting content repeatedly over time—whereas in our study, we were

only able to test the immediate effect of a single exposure. It is possible that an effect does exist, but only manifests over a longer period through repeated exposure. This question could be explored using a longitudinal experimental design or self-reported survey data.

Lastly, the null findings in our study contribute to an ongoing and necessary conversation about the role of non-significant results in experimental research. Such outcomes are crucial for mitigating publication bias and building a more accurate and comprehensive scientific record. As positive findings are disproportionately published, reporting null effects helps prevent overestimation of intervention efficacy and supports the development of more robust theoretical frameworks and research designs. In the context of this experiment, the null result does not necessarily indicate that cross-cutting content is ineffective, but rather that a one-time exposure, however strong the stimuli is, may be insufficient to alter entrenched beliefs. By reporting these findings transparently, we aim to encourage a more nuanced understanding of experimental outcomes and to support future research in refining both theoretical models and methodological approaches.

Funding: This work was supported by the Momentum program of the Hungarian Academy of Sciences (MTA-HUN-REN TK Lendület "Momentum" PRiSMa research project [LP2024-2/2024]), the ÚNKP-17-3-IV-BCE-27 New National Excellence Program of the Ministry of Human Capacities and the National Research, Development and Innovation Fund (FK-13775).

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

Anspach, N. M. (2017). The New Personal Influence: How Our Facebook Friends Influence the News We Read. *Political Communication*, *34*(4), 590–606. https://doi.org/10.1080/10584609.2017.1316329

Ardèvol-Abreu, A., Diehl, T., & Gil de Zúñiga, H. (2019). Antecedents of internal political efficacy incidental news exposure online and the mediating role of political discussion. *Politics*, *39*(1), 82–100. https://doi.org/10.1177/0263395717693251

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Barnidge, M. (2017). Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings. *Political Communication*, *34*(2), 302–321. https://doi.org/10.1080/10584609.2016.1235639

Beam, M. A., Child, J. T., Hutchens, M. J., & Hmielowski, J. D. (2018). Context collapse and privacy management: Diversity in Facebook friends increases online news reading and sharing. *New Media & Society*, *20*(7), 2296–2314. https://doi.org/10.1177/1461444817714790

Bene, M. (2021). Topics to talk about. The effects of political topics and issue ownership on user engagement with politicians' Facebook posts during the 2018 Hungarian general election. *Journal of Information Technology & Politics*, *18*(3), 338–354. https://doi.org/10.1080/19331681.2021.1881015

Bennett, W. L., & Iyengar, S. (2008). A New Era of Minimal Effects? The Changing Foundations of Political Communication. *Journal of Communication*, *58*(4), 707–731. https://doi.org/10.1111/j.1460-2466.2008.00410.x

Bossetta, M. (2018). The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. *Journalism & Mass Communication Quarterly*, *95*(2), 471–496. https://doi.org/10.1177/1077699018763307

Boulianne, S., Oser, J., & Hoffmann, C. P. (2023). Powerless in the digital age? A systematic review and meta-analysis of political efficacy and digital media use. *New Media & Society*, 25(9), 2512–2536.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, *14*(7), 1164–1180. https://doi.org/10.1177/1461444812440159

Campbell, A. (1971). *The Voter Decides* (First edition). Praeger.

Cardenal, A. S., Aguilar-Paredes, C., Galais, C., & Pérez-Montoro, M. (2019). Digital Technologies and Selective Exposure: How Choice and Filter Bubbles Shape News Media Exposure. *The International Journal of Press/Politics*, *24*(4), 465–486. https://doi.org/10.1177/1940161219862988

Chan, M., Chen, H.-T., & Lee, F. L. F. (2017). Examining the roles of mobile and social media in political participation: A cross-national analysis of three Asian societies using a communication mediation approach. *New Media & Society*, *19*(12), 2003–2021. https://doi.org/10.1177/1461444816653190

Chen, H.-T., & Lin, J.-S. (2021). Cross-cutting and Like-minded Discussion on Social Media: The Moderating Role of Issue Importance in the (De)mobilizing Effect of Political Discussion on Political Participation. *Journal of Broadcasting & Electronic Media*, *65*(1), 135–156. https://doi.org/10.1080/08838151.2021.1897822

Christensen, H. S. (2016). All the same? Examining the link between three kinds of political dissatisfaction and protest. *Comparative European Politics*, *14*(6), 781–801. https://doi.org/10.1057/cep.2014.52

DeVito, M. A. (2017). From Editors to Algorithms. *Digital Journalism*, *5*(6), 753–773. https://doi.org/10.1080/21670811.2016.1178592

Dobos, G., Gyulai, A., & Horváth, A. (n.d.). Felesleges erőfeszítések? Választási programok és ígéretek 2018-ban. In B. Böcskei & A. Szabó (Eds.), *Várakozások és valóságok. Parlamenti választás 2018* (pp. 317–339). Napvilág Kiadó, MTA TK PTI.

Easton, D., & Dennis, J. (1967). The Child's Acquisition of Regime Norms: Political Efficacy. *American Political Science Review*, *61*(1), 25–38. https://doi.org/10.2307/1953873

Eurobarometer. (2018, August 6). *Public opinion in the European Union—Report. The Standard Eurobarometer survey of spring 2018 (EB89), Factsheet on Hungary.* https://doi.org/10.2775/172445

Eveland, W. P., Jr, & Hively, M. H. (2009). Political Discussion Frequency, Network Size, and "Heterogeneity" of Discussion as Predictors of Political Knowledge and Participation. *Journal of Communication*, *59*(2), 205–224. https://doi.org/10.1111/j.1460-2466.2009.01412.x

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Fishkin, J. S. (1997). *The Voice of the People: Public Opinion and Democracy* (Reprint edition). Yale University Press.

Gil de Zúñiga, H., Borah, P., & Goyanes, M. (2021). How do people learn about politics when inadvertently exposed to news? Incidental news paradoxical Direct and indirect effects on political knowledge. *Computers in Human Behavior*, *121*, 106803. https://doi.org/10.1016/j.chb.2021.106803

Goyanes, M., & Demeter, M. (2020). Beyond positive or negative: Understanding the phenomenology, typologies and impact of incidental news exposure on citizens' daily lives. *New Media & Society*, 1461444820967679. https://doi.org/10.1177/1461444820967679

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380. https://doi.org/10.1086/225469

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., ... & Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign?. *Science*, 381(6656), 398–404.

Gutmann, A., & Thompson, D. F. (1998). *Democracy and Disagreement*. Belknap Press: An Imprint of Harvard University Press.

Halpern, D., Valenzuela, S., & Katz, J. E. (2017). We Face, I Tweet: How Different Social Media Influence Political Participation through Collective and Internal Efficacy. *Journal of Computer-Mediated Communication*, *22*(6), 320–336. https://doi.org/10.1111/jcc4.12198

Hmielowski, J. D., Kim, S., Hutchens, M. J., & Beam, M. A. (2018). Engaged or Disengaged? Examining the Relationship Between Electoral Ambivalence and Indicators of Political Engagement in the 2012 U.S. Election. *Atlantic Journal of Communication*, *26*(1), 32–45. https://doi.org/10.1080/15456870.2018.1398164

Huckfeldt, R., Mendez, J. M., & Osborn, T. (2004). Disagreement, Ambivalence, and Engagement: The Political Consequences of Heterogeneous Networks. *Political Psychology*, *25*(1), 65–95. https://doi.org/10.1111/j.1467-9221.2004.00357.x

Iyengar, S., & Hahn, K. S. (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

John, N. A., & Dvir-Gvirsman, S. (2015). "I Don't like You Any More": Facebook Unfriending by Israelis during the Israel–Gaza Conflict of 2014. *Journal of Communication*, *65*(6), 953–974. https://doi.org/10.1111/jcom.12188

Karnowski, V., Kümpel, A. S., Leonhard, L., & Leiner, D. J. (2017). From incidental news exposure to news engagement. How perceptions of the news post and news usage patterns influence engagement with news articles encountered on Facebook. *Computers in Human Behavior*, *76*, 42–50. https://doi.org/10.1016/j.chb.2017.06.041

Karv, T., Lindell, M., & Rapeli, L. (n.d.). How Context Matters: The Significance of Political Homogeneity and Language for Political Efficacy. *Scandinavian Political Studies*, *n/a*(n/a). https://doi.org/10.1111/1467-9477.12215

Kenski, K., & Stroud, N. J. (2006). Connections Between Internet Use and Political Efficacy, Knowledge, and Participation. *Journal of Broadcasting & Electronic Media*, *50*(2), 173–192. https://doi.org/10.1207/s15506878jobem5002_1

Kim, Y., & Chen, H.-T. (2015). Discussion Network Heterogeneity Matters: Examining a Moderated Mediation Model of Social Media Use and Civic Engagement. *International Journal of Communication*, *9*(0), 22.

Kim, Y., Chen, H.-T., & Gil de Zúñiga, H. (2013). Stumbling upon news on the Internet: Effects of incidental news exposure and relative entertainment use on political engagement. *Computers in Human Behavior*, *29*(6), 2607–2614. https://doi.org/10.1016/j.chb.2013.06.005

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2020). Confirmation Bias, Ingroup Bias, and Negativity Bias in Selective Exposure to Political Information. *Communication Research*, *47*(1), 104–124. https://doi.org/10.1177/0093650217719596

Kwak, N., Williams, A. E., Wang, X., & Lee, H. (2005). Talking Politics and Engaging Politics: An Examination of the Interactive Relationships Between Structural Features of Political Talk and Discussion Engagement. *Communication Research*, *32*(1), 87–111. https://doi.org/10.1177/0093650204271400

Lee, F. L. F. (2012). Does Discussion With Disagreement Discourage All Types of Political Participation? Survey Evidence From Hong Kong. *Communication Research*, *39*(4), 543–562. https://doi.org/10.1177/0093650211398356

Lee, F. L. F., & Chan, J. M. (2009). The Political Consequences of Ambivalence: The Case of Democratic Reform in Hong Kong. *International Journal of Public Opinion Research*, *21*(1), 47–64. https://doi.org/10.1093/ijpor/edn053

Lee, K. M. (2006). Effects of Internet Use on College Students' Political Efficacy. *CyberPsychology & Behavior*, *9*(4), 415–422. https://doi.org/10.1089/cpb.2006.9.415

Matthes, J., Knoll, J., Valenzuela, S., Hopmann, D. N., & Von Sikorski, C. (2019). A Meta-Analysis of the Effects of Cross-Cutting Exposure on Political Participation. *Political Communication*, *36*(4), 523–542. https://doi.org/10.1080/10584609.2019.1619638

McGraw, K. M., & Bartels, B. (2005). Ambivalence Toward American Political Institutions. In S. C. Craig & M. D. Martinez (Eds.), *Ambivalence and the Structure of Political Opinion* (pp. 105–126). Palgrave Macmillan US. https://doi.org/10.1057/9781403979094_6

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Messing, S., & Westwood, S. J. (2014). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, *41*(8), 1042–1063. https://doi.org/10.1177/0093650212466406

Moeller, J., de Vreese, C., Esser, F., & Kunz, R. (2014). Pathway to Political Participation: The Influence of Online and Offline News Media on Internal Efficacy and Turnout of First-Time Voters. *American Behavioral Scientist*, *58*(5), 689–700. https://doi.org/10.1177/0002764213515220

Mutz, D. C. (2006). *Hearing the Other Side: Deliberative versus Participatory Democracy* (Illustrated edition). Cambridge University Press.

Nanz, A., Kaskeleviciute, R., Stubenvoll, M., & Matthes, J. (2025). Scanning vs. Thorough Processing the News: Antecedents of First-and Second-Level Incidental Exposure and the Role of the Relevance Appraisal. *Digital Journalism*, Online first, 1–21.

Newhagen, J. E. (1994). Self-Efficacy and Call-in Political Television Show Use. *Communication Research*, *21*(3), 366–379. https://doi.org/10.1177/009365094021003007

Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T., & Nielsen, R. K. (2021). *Reuters Institute Digital News Report 2021.* University of Oxford. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf

Niemi, R. G., Craig, S. C., & Mattei, F. (1991). Measuring Internal Political Efficacy in the 1988 National Election Study. *American Political Science Review*, *85*(4), 1407–1413. https://doi.org/10.2307/1963953

Nir, L. (2011a). Motivated Reasoning and Public Opinion Perception. *Public Opinion Quarterly*, *75*(3), 504–532. https://doi.org/10.1093/poq/nfq076

Nir, L. (2011b). Disagreement and Opposition in Social Networks: Does Disagreement Discourage Turnout? *Political Studies*, *59*(3), 674–692. https://doi.org/10.1111/j.1467-9248.2010.00873.x

Osborne-Crowley, K. (2020). Social cognition in the real world: Reconnecting the study of social cognition with social reality. *Review of General Psychology*, 24(2), 144–158

Political Capital. (2017). *Gyógyítható-e az egészségügy?* Friedrich Ebert Stiftung. https://www.politicalcapital.hu/pc-admin/source/documents/PC_FES_egeszsegugy_2017_tanulmany_20170613.pdf

Pollock, P. H. (1983). The Participatory Consequences of Internal and External Political Efficacy: A Research Note. *Western Political Quarterly*, *36*(3), 400–409. https://doi.org/10.1177/106591298303600306

Rainie, L., & Wellman, B. (2014). *Networked: The New Social Operating System* (Illustrated edition). The MIT Press.

Robinson, M. J. (1976). Public Affairs Television and the Growth of Political Malaise: The Case of "The Selling of the Pentagon"*. *American Political Science Review*, *70*(2), 409–432. https://doi.org/10.2307/1959647

Scheufele, D. A., Hardy, B. W., Brossard, D., Waismel-Manor, I. S., & Nisbet, E. (2006). Democracy Based on Difference: Examining the Links Between Structural Heterogeneity, Heterogeneity of Discussion Networks, and Democratic Citizenship. *Journal of Communication*, *56*(4), 728–753. https://doi.org/10.1111/j.1460-2466.2006.00317.x

Valeriani, A., & Vaccari, C. (2016). Accidental exposure to politics on social media as online participation equalizer in Germany, Italy, and the United Kingdom. *New Media & Society*, *18*(9), 1857–1874. https://doi.org/10.1177/1461444815616223

van Harreveld, F., Rutjens, B. T., Rotteveel, M., Nordgren, L. F., & van der Pligt, J. (2009). Ambivalence and decisional conflict as a cause of psychological discomfort: Feeling tense before jumping off the fence. *Journal of Experimental Social Psychology*, *45*(1), 167–173. https://doi.org/10.1016/j.jesp.2008.08.015

Zhang, W., & Chang, L. (2014). Perceived Speech Conditions and Disagreement of Everyday Talk: A Proceduralist Perspective of Citizen Deliberation. *Communication Theory*, *24*(2), 124–145. https://doi.org/10.1111/comt.12034

Zhu, Q., Skoric, M., & Shen, F. (2017). I Shield Myself From Thee: Selective Avoidance on Social Media During Political Protests. *Political Communication*, *34*(1), 112–131. https://doi.org/10.1080/10584609.2016.1222471

APPENDIX

ProGov thread (see English translation below)

AntiGov thread (see English translation below)

The *Mixed thread* combined the following comments, AntiGov (1), ProGov (1), AntiGov (4) and ProGov (2).

ProGov thread (English translation)

1) It is a lie that the health service is dying! 1200 billion has been spent on health care on developments and on wages. And another 700 billion to Pest County and Budapest. There has never been such progress! If the opposition media had the time to do its job instead of spreading scare stories, hunting down shabby toilets and producing disinformation, then maybe everyone would know what developments have been made all over the country.

2) Under Ferenc Gyurcsány's government, he has withdrawn billions of forints from the health sector, he has chased thousands of health workers out of the profession and left hundreds of municipalities without a family doctor. The civic [Fidesz] government has been replacing the money taken from the health sector since 2010. It is renovating hospitals and clinics across the country and giving health workers and GPs a pay rise.
3) Among the health measures of recent years are the pay rises and the resettlement programs for residents, and general practitioners. The Fidesz--KDNP, which raises wages in the health sector, is the guarantee for better quality in healthcare, as opposed to the MSZP, which charges a visit fee, or the Jobbik, which is exporting doctors.
4) During the MSZP-SZDSZ thievery between 2002 and 2010, hospitals were closed down or privatized instead of being renovated and expanded. Since 2010, the opposite has been happening: the government has spent a lot of money on health care and the results are visible.

AntiGov thread (English translation)

1) The government, which is doing so well with hundreds of people freezing to death in winter and tens of thousands in the health system because they can't get care, does not give a shit to bring home professionals and keep them home . While Orbán's obsessions with stadiums, railways and theft have been well funded in recent years, the health service is dying, there is a constant shortage of resources and some of the professionals prefer to go abroad.
2) In the light of recent events, I think we have every right to be angry! There is no money for basic hygiene in hospitals, for the care of patients. But billions of taxpayers' money can be wasted on unnecessary stadiums – the passion of a lunatic mind!!!!!
3) In Hungary today, people go to hospitals not to get better, but to die. I do not really cry for Gyurcsány's people, but even then we were not that far behind European standards! Of course, according to the government-controlled media, everything is fine and dandy. Go Hungary!
4) Zero health care in Hungary, fleas, bedbugs and who knows what other extra residents there are.... Patients left on their own, waiting 6 hours in the waiting room, doctors abroad... but Hungary is doing better, huh???

ORIGINAL ARTICLE

DOI: 10.51480/1899-5101.18.3(41).777

Who Shouts the Loudest? Predictors of Conflict-Oriented Behavior on Social Network Sites

Martina Novotná

0000-0001-5536-4790

Masaryk University, Brno, Czech Republic

Alena Pospíšil Macková

0000-0001-5967-5166

Masaryk University, Brno, Czech Republic

Lucie Čejková

0000-0002-8598-6846

Masaryk University, Brno, Czech Republic

Abstract: Conflict-oriented behavior in the online environment is attracting increasing attention, mainly due to its widespread presence and the potential implications for shaping public opinion and attitudes. This study examines the factors that drive such behavior on social network sites and their potential implications for the online public sphere. The hierarchical binary logistic regression draws on data from a representative survey of Czech adult users of social network sites (N = 2,187; 55.3% female; mean age = 44.57, SD = 16.45), collected at the end of the first year of the COVID-19 pandemic (November–December 2020). The results indicate that conflict-oriented commenters, more likely to be less educated, younger men with low political interest, are highly active in online discussions driven by disagreement. Additionally, they are more likely than non-conflict-oriented commenters to unfriend others over differing opinions and express strong negativity toward opposing views. The study discusses implications for societal polarization and deliberative democracy.

Keywords: online discussions, cross-cutting discussions, conflict-oriented behavior, unfriending, political interest

INTRODUCTION

Engaging in (online) political discussions is a vital component of a democratic society because it allows individuals to express their opinions, encounter contrasting perspectives, and, ultimately, gain a better understanding of the world around them (Rossini & Stromer-Galley, 2019). However, the way people currently engage in online communication often does not meet the desired standards of civil discourse. Instead of productive and respectful discussions, individuals frequently engage in toxic and uncivil online communication that is characterized by verbal fighting, vulgar words, and hostility. As a result, researchers are shifting their focus from achieving an ideal form of deliberative opinion exchange to exploring actual behavior and perceptions while engaging in uncivil online discussions that are perceived as violations of communication norms (see Coe et al., 2014; Bormann, 2022; Frischlich et al., 2021; Kenski et al., 2020; Rossini, 2022). Various types of norm violations in online discussions are often blamed for negatively impacting the dynamics of online environments, influencing public opinion, shaping attitudes, and discouraging citizens from participation (Mutz, 2002). These are the main reasons for the growing attention that is now being paid to online conflict-oriented behavior (see Chen, 2017; Frischlich et al., 2021; Kim & Kim, 2019; Kim et al., 2021; Vochocová & Rosenfeldová, 2019).

Previous research has primarily focused on the effects on individuals, their perceptions of such communication (Kenski et al., 2020; Mutz, 2016; Sydnor, 2019), and their reactions (Gervais, 2015; Goyanes et al., 2021) and their relation to polarization (Hwang et al., 2014; Kim & Kim, 2019). Still, there is a lack of studies that attempt to understand the online behavior and attitudes of those who use conflict-oriented behavior (i.e., vulgar words, verbal fighting) in online discussions (but see Coe et al., 2014; Frischlich et al., 2021; Kim et al., 2021). Beyond that, we assume that the issue of conflict-oriented behavior even gains importance during crises, when society and public opinion are fragile and when emotional reactions are intense and online discussions are uncivil and polarized (Novotná et al., 2023). Thus, we expect that a certain period of data collection might shape conflict-oriented behavior that was driven by emotions, and our data from the COVID-19 pandemic might provide essential insights into the role of harsh communication strategies in online discussions.

This study contributes to existing literature by delving into the potential predictors of conflict-oriented behavior, which are often overlooked (but see Rains et al., 2017; Vargo & Hopp, 2017). We extend the previous knowledge by examining the personal characteristics of those who use conflict-oriented online behavior (Coe et al., 2014; Koban et al., 2018), commenters' attitudes towards counter-attitudinal opinions and behavior on social network sites (SNS) – the frequency of engagement in online discussions (i.e., political and

cross-cutting discussions), and selective behavior (i.e., political unfriending). Focusing on predictors of conflict-oriented behavior, which is rather unexplored, can provide a deeper understanding of the motivations behind such behaviors and their implications for individuals who tend to avoid conflict. Furthermore, this investigation sheds light on dynamics within SNS by examining how certain online practices, types of online discussions, and personal characteristics are related to conflict-oriented behavior. Opinions expressed in a harsh communication style can spread rapidly and become prevalent, especially online. Problematic content is often circulated by a relatively small group of individuals, who can nevertheless exert a substantial influence on public opinion formation (Gao et al., 2024). Therefore, it is essential to consider who may be discouraged from participating in the conversation.

Additionally, we still face a significant lack of country-contextual variety within the studies that investigate the characteristics of online discussions, even though exploring various contexts shows different behavioral patterns, different dynamics for discussions, and a variety of country- and culture-related factors that determine the character of the conflict-oriented behavior and the perception of what is and what is not acceptable (see Hmielowski et al., 2014; Vochocová & Rosenfeldová, 2019). We test conflict-oriented behavior, its predictors, in the under-researched Central and Eastern European (CEE) region. Specifically, we focused on Czechia, a society with high levels of affective polarization (Orhan, 2022) and political skepticism. Negative feelings towards political elites, institutions, and immigrants are prevalent (Pospěch, 2021), often manifested in hostility, xenophobia, and racism in Czech online discussions (Vochocová & Rosenfeldová, 2019).

Based on survey data collected during the first year of the pandemic, we show that SNS users who engage in conflict-oriented behavior in Czechia report more frequent engagement in online cross-cutting discussions and a strong tendency for politically motivated unfriending as a reaction to disagreement. Contrary to some of our more optimistic assumptions about online talk, those who use conflict-oriented behavior also have a rather negative attitude towards disagreement and the dissenting opinions of other commenters, despite (or because of) talking with them most often. Those who employ conflict-oriented behavior often contribute to political discussions, but surprisingly, they tend to have low political interest.

CONFLICT-ORIENTED BEHAVIOR IN ONLINE DISCUSSIONS

Scholars have long been interested in online discussions as arenas of opinion exchange. This interest is grounded in the deliberative model of democracy (Elster, 1998; Gutmann & Thompson, 1996), which views open, reasoned, and plural discussion as essential to upholding democratic principles (see Chen, 2017; Hwang et al., 2016; Lee et al., 2021). Exchanges of opposing opinions are often considered particularly valuable, as such interactions allow individuals to broaden their perspectives and foster critical thinking (Lu & Lee, 2021; Mutz, 2006). Nevertheless, empirical studies repeatedly show that interactions based on disagreement often diverge from deliberative ideals and exhibit various forms of conflict-oriented behavior leading to both positive and negative consequences. We understand this concept as encompassing communicative acts that express or escalate confrontation rather than mutual understanding.

Within this broader category, incivility—such as name-calling, personal attacks, or vulgar and insulting language—represents one prominent manifestation (Coe et al., 2014; Kim et al., 2021; Rossini, 2022; Stryker, 2016; Vochocová & Rosenfeldová, 2019). Another form is verbal fighting, defined as discursive exchanges characterized by hostility, reciprocal attacks, and affective escalation (Stryker, 2016). Such behaviors undermine inclusive and continuous political conversations and are often triggered by controversial or morally charged topics. In this article, we therefore use conflict-oriented behavior as an umbrella term that captures these communicative dynamics, particularly incivility and verbal fighting, as expressions of antagonistic interaction styles that stand in tension with deliberative ideals. Such conflict-oriented exchanges are particularly likely to emerge in discussions characterized by disagreement—often referred to as cross-cutting discussions—where opposing views can easily intensify and escalate into emotional confrontation. This tendency is especially common in debates over controversial or morally charged topics, such as immigration or vaccination (Vochocová & Rosenfeldová, 2019). Working with this concept allows us to capture not only the overt expressions of incivility and hostility but also their communicative role in undermining the deliberative quality of online discussions.

Among these manifestations, incivility stands out as one of the most visible and consequential forms of conflict-oriented behavior, widely discussed for its potential harmful effects on democratic discourse (Gervais, 2019; Papacharissi, 2004). Firstly, previous research shows that some people could view uncivil forms of expressions as an unacceptable discussion pattern that may lead to the avoidance of further discussions (Goyanes et al., 2021), whereas others might tend to be more resilient. According to research, personal traits, such as sex, age, and education, may impact how individuals respond to incivility in online

interactions. Research further indicates that men are more likely than women to use uncivil language (Frischlich et al., 2021; Proust & Saldaña, 2022) and that younger individuals may engage in such behavior more often, sometimes as a form of provocation (Vraga, 2015). Additionally, those with lower levels of education tend to exhibit more incivility in their online interactions (Vargo & Hopp, 2017). Harsh communication styles may discourage participation (Novotná et al., 2023; Vraga, 2015), but they can also encourage critical thinking (Chen, 2017) and be perceived as entertaining (Koban et al., 2018; Sydnor, 2019).

Secondly, there are worries about the character of the opinion exchange. Uncivil expressions may threaten discussion quality because expressive language can overshadow the argument and hinder the understanding of different points of view (Anderson et al., 2014). Overall, incivility is problematic for delegitimizing arguments, and it lacks respectful opinion exchange (Coe et al., 2014), which triggers angry reactions and hostility directed toward others (Chen, 2017; Gervais, 2019; Sydnor, 2019). Therefore, uncivil discussions may broaden polarization through hostile interactions with others and the increased polarization of the issues (Anderson et al., 2014). During times of crisis, heated exchanges may further reduce willingness to listen to opposing views, reinforcing perceived differences and mutual mistrust (Novotná et al., 2023).

Furthermore, comments that lack civility are often given high ratings by social media algorithms, which can also increase their visibility (Gervais, 2015; Kim et al., 2021) and reach a large audience, including those who only read the comments. Although algorithmic operations remain largely a black box, moderation tools may also play a role, and Mark Zuckerberg's announcement of Meta's policy changes in January 2025 could further influence the circulation of such problematic content. Ultimately, uncivil comments often provoke additional uncivil responses, reinforcing cycles of hostility in online discussions (Gao et al., 2024). These dominant uncivil opinions can significantly impact beliefs and encourage extreme opinions (Rösner et al., 2016). Certain opinions might become more prevalent, although they can be held by a minority, but are widely seen by many. This can be more problematic, especially during crisis times when the need for accurate information increases, along with the rise in social media usage for news (Aelst et al., 2021), which could potentially amplify exposure to online opinions.

CONFLICT-ORIENTED BEHAVIOR AND UNFRIENDING AS COPING STRATEGIES FOR DISAGREEMENT

Conflict-oriented behavior as a communication style is closely related to disagreement. People who disagree with someone else in conversation are more likely to be disrespectful or express negative sentiments when they have to justify their arguments, contrary to like-minded discussions (Chen, 2017; Marchal, 2021; Rossini, 2022; Rossini & Maia, 2021; Sun et al., 2021). Conflict-oriented behavior might be used as a coping strategy to respond to disagreement, and it may also be evaluated as a common aspect of communication. Therefore, we focus on cross-cutting discussions that involve disagreement to understand the reasons behind conflict-oriented behavior and its impact on online discussions. Given that disagreement in cross-cutting discussions is often associated with negative emotions such as stress or angry feelings (Anderson & Auxier, 2020; Duggan & Smith, 2016) and given that it may lead to conflict-oriented behavior as a coping mechanism (Hmielowski et al., 2014), we hypothesize:

- H1: SNS users who engage in conflict-oriented behavior in online discussions report a higher frequency of engagement in cross-cutting discussions than those who do not.

Another coping strategy for dealing with disagreement is by unfriending or blocking users. The unfriending feature in the online space allows people to avoid unpleasant content, potentially create a safe space to express their opinions (Zhu & Skoric, 2021). In addition, uncivil and conflicting disagreement seems to motivate people to unfriend others more than the actual disagreement itself (Goyanes et al., 2021; Neubaum et al., 2021; Peña & Brody, 2014). Unfriending options may be handy tools, especially for those with a higher level of conflict avoidance (see Goyanes et al., 2021; Vraga, 2015). According to this logic, we may see political unfriending that directly cuts the ties between people who have a different point of view as a reaction to a hostile environment that does not fulfill SNS users' expectations about the discussion.

Turning around the argument from previous studies that unfriending works as a strategy for avoiding disagreement and a harsher communication style (Goyanes et al., 2021; Neubaum et al., 2021; Peña & Brody, 2014), we might expect that people who often use conflict-oriented behavior would more frequently be exposed to disagreement and their environment on SNS would include more heterogeneous opinions. As they use harsh communication styles themselves, they may be more resistant to conflict and less likely to unfriend or block someone. Given that unfriending, on the contrary, limits heterogeneity, we assume that:

- H2: SNS users who engage in conflict-oriented behavior in online discussions utilize political unfriending less than those who do not.

THE DYNAMICS OF CONFLICT-ORIENTED BEHAVIOR AND OPEN--MINDEDNESS TO TALK AND LISTEN TO THOSE FROM "THE OTHER SIDE"

Most Czech SNS users (57%) do not respond to disagreement with a comment similar to data from the United States (Duggan & Smith, 2016). That is a non-negligible number of people (Macková et al., 2021). While convincing others about the relevance of different opinions, attacks and argumentative fouls take place. Following the assumption that conflict-oriented behavior is likely connected to cross-cutting discussions, as we expect in Hypothesis 1, we need to consider that offensive comments and disagreements in an online environment might trigger negative emotions. Feelings of "freaking out" may lead to an unwillingness to join talks with someone who has a different point of view (Duggan & Smith, 2016), which applies to both offline and online environments (Chen, 2017). Moreover, exposure to a harsher communication style in combination with disagreement fuels closed-mindedness toward counter-attitudinal opinions (Hwang et al., 2016).

Based on the assumption that more assertive commenters who use conflict-oriented behavior perceive the harsher communication style differently than those who instead tend not to use such a communication style, one would expect that users who use conflict-oriented behavior are more open-minded to disagreement as part of a more heated discussion and this does not discourage their engagement. Thus, we formulate the following hypothesis:

- H3: SNS users who engage in conflict-oriented behavior within online discussions are more open-minded to counter-attitudinal opinions than those who do not.

POLITICAL INTEREST AND ENGAGEMENT IN ONLINE POLITICAL DISCUSSIONS FUEL CONFLICT-ORIENTED BEHAVIOR

Political interest is one of the key motivations for engagement in online discussions that many consider to be a form of civic (online) participation (Švelch & Vochocová, 2015). Politically interested people seek to understand issues and get exposure to different points of view (Rossini & Stromer-Galley, 2019). The assumption is that politically interested (Kim et al., 2021), attentive, knowledgeable, and more vigorously partisan people are more active in online

discussions (Duggan & Smith, 2016). Equally important, research has found that individuals who frequently engage in online commenting are more prone to making uncivil remarks (Hmielowski et al., 2014; Kim et al., 2021). This could be explained by the higher likelihood of being exposed to incivility and verbal fighting in the online environment (Goyanes et al., 2021; Koban et al., 2018), which consequently contributes to incivility (Frischlich et al., 2021; Kim et al., 2021) because its presence can establish unwritten norms and set expectations about other users' behaviors and the dynamic of the SNS environment. To sum up, since politically interested people are most involved in online discussions, they are expected to be exposed to harsher communication styles, which can provoke the same conflicting reactions. Thus, we assume that:

- H4: SNS users who engage in conflict-oriented behavior within online discussions are more interested in politics than those who do not.

In addition, conflict-oriented behavior seems to be more common in political discussions, especially in discussions focused on public affairs and political issues (Hmielowski et al., 2014; Kim et al., 2021; Papacharissi, 2004; Sun et al., 2021). Though conversations about politics undoubtedly have many positive outcomes (Rossini & Stromer-Galley, 2019), there exist concerns that engagement in online political talk might lead to the higher acceptance of incivility within the discussion, simply because it is commonplace (Hmielowski et al., 2014). In line with previous findings, we propose the following:

- H5: SNS users who engage in conflict-oriented behavior in online discussions have a higher frequency of engagement in online political discussions than those who do not.

METHODOLOGY

DATA AND PARTICIPANTS

The study employs data from a Czech survey collected by the Focus (Marketing & Social Research) agency in November–December 2020, combining online interviews with respondents from Czech national panel data (n = 2,538) and personal interviews (n = 1,225). The representative dataset, which is based on quota sampling, includes responses from 3,763 Czech adults. The research agency follows the ICC/ESOMAR international research code to guarantee high research standards. The research was carried out under Czech laws, regulations, and the guidance of the Research Ethics Committee at Masaryk University.

The questionnaire consisted of measures for news consumption, media use and trust, polarization, political attitudes, and voting behavior. In this study, which is built on secondary data usage, we specifically employ a data subsample of SNS users who participated in or read discussions on SNS (n = 2,187). For this purpose, we filtered out those participants who answered "never" for the following two questions: (1) "How often are you involved in discussions on SNS?" and (2) "How often do you only read discussions on SNS without active participation?" Respondents rated their answers on a 5-point scale (1 = never, 5 = very often). In this process, a total of 1,576 cases were removed.

Regarding our research topic, it is essential to highlight that the data were collected at the end of the first year of the COVID-19 pandemic, which was its second wave in Czechia. There was an apparent 2% increase in the use of SNS compared to the previous year, which is potentially related to the pandemic protective measures, lockdowns, and quarantines that pushed people to move most of their activities online (Macková et al., 2021). Different countries (Van Aelst et al., 2021) supported the growing trend of people using social media more often during the early stages of the pandemic.

MEASURES

To identify participants with *conflict-oriented behavior*, we created a binary variable combining two essential characteristics of problematic communication style – incivility as a representation of utterances and verbal fighting, capturing the discursive character of incivility (Stryker, 2016). Both of these manifestations of conflict behavior refer to a communication style that represents a more assertive way of expressing oneself during an exchange of opinions. The first item was inspired by measures used by Rainie and Smith to assess perceived negativity (2012). To identify those who tend to use incivility in online discussions, we asked the following question: "I add negative comments in discussions, including hostile or vulgar words, in answer to other comments or posts shared by someone else" (M = 1.62, SD = 0.89). Verbal fighting was assessed by: "I fight in online discussions with someone who has an opposite opinion to me" (M = 1.73, SD = 0.90). Original items were measured with a 5-point scale (1 = never, 5 = very often). To obtain the binary dependent variable, we created a group of those who (0; 41.4%) have never participated in any of the activities (i.e., selected 1 = *never* for both questions). Those who (1; 58.6%) have at least once participated in one of the activities (i.e., selected 2–5 for at least one of the questions) formed the second group. The original variables were severely skewed toward the *never* option, and recoding them into a binary variable allowed us to obtain two relatively equal groups. Consequently, the recoded variable still informed our research aims and performed better in statistical procedures than the original variables.

Engagement in Cross-Cutting Discussions on SNS was measured by the item: "Based on your experience when you discuss (topics) on social network sites with people with whom you disagree or with whom you have different opinions, or you just read such discussions, how often is this valid for you: I react to those posts or comments with my own comment or by publishing my own content." Respondents rated their answers on a 5-point scale (1 = never, 5 = very often), which we reduced to a 4-point scale by merging the values of 4 and 5 due to low occurrence of the highest category (*M* = 2.39, *SD* = 0.88).

Political Unfriending on SNS as a reaction to disagreement in the online environment was captured with an index, including three items. All items were measured as binary (0 = no, 1 = yes), and the score was computed as the sum of values, ranging from 0–3 (*M* = 0.52, *SD* = 0.92; Cronbach's α = .74). We asked: "Did you delete, block, or hide someone from your social network site because: (1) They shared something you disagreed with about politics or public affairs (17.7% = yes); (2) They argued about politics or public affairs on your social network site with you or someone you know (15.3% = yes); or (3) They disagreed with something you shared about politics or public affairs (17.5% = yes)."

Open-mindedness to Counter-attitudinal Opinions consisted of an index (*M* = 3.09, *SD* = 0.90; Cronbach's α = .78) with two items that addressed attitudes toward cross-cutting discussions. Firstly, we asked people: "How much do you agree or disagree with the following statement, which is connected to sharing opinions or political content on Facebook and other social network sites: It is unpleasant for me when my friends and people I know share political content and opinions that I do not agree with." This was measured with a 5-point scale that ranged from "strongly disagree" (1) to "strongly agree" (5). Secondly, we measured the level of agreement on the same 5-point scale with the following statement: "It does not make any sense to be friends with or to talk to someone who has a different point of view than me." We recoded the values for a final index to capture that a higher value means higher open-mindedness to counter-attitudinal opinions, and the index was computed as a mean of the two items.

Political interest was measured on an 11-point scale (0 = not at all interested, 10 = very interested) with the question: "How much would you say you are interested in politics?" (*M* = 5.19, *SD* = 2.73).

The statement that measured the *Frequency of Engagement in Political Discussions on SNS* was: "I discuss politics and public affairs on online social network sites." The variable was measured with a 5-point scale ranging from "never" to "very often" (*M* = 2.18, *SD* = 1.16).

Lastly, the socio-demographic characteristics we examined were sex (female = 55.3%), age (*M* = 44.57, *SD* = 16.45), and education (primary education = 10.54%, vocational certificate = 34.57%, high school = 34.73%, university degree = 20.16%).

ANALYTICAL PROCEDURE

After handling the variables, we first obtained descriptive statistics and correlations of all of the variables (Table 1). In the assumptions check, all correlations were lower than 0.5, the variance inflation factors (VIF) of all variables were 2.00, and tolerance was no lower than 0.5, suggesting that the assumptions were met. Then, we ran a binary regression with conflict-oriented behavior as the dependent variable. Independent variables were added in three blocks to control for the effect of each block and see how much the blocks contribute to the prediction of the dependent variable. Block 1 comprised sociodemographic variables (sex, age, education), Block 2 included variables applicable to a broader, not only SNS context (political interest, open-mindedness to counter-attitudinal opinions), and Block 3 included SNS behavior (engagement in cross-cutting discussions, frequency of engagement in political discussions, political unfriending on SNS). IBM SPSS Statistics 30 was used in all steps of the analysis, and the ggplot2 package in R was used to visualize the results (R Core Team, 2014; Wickham, 2009).

Table 1. Correlations of All Study Variables

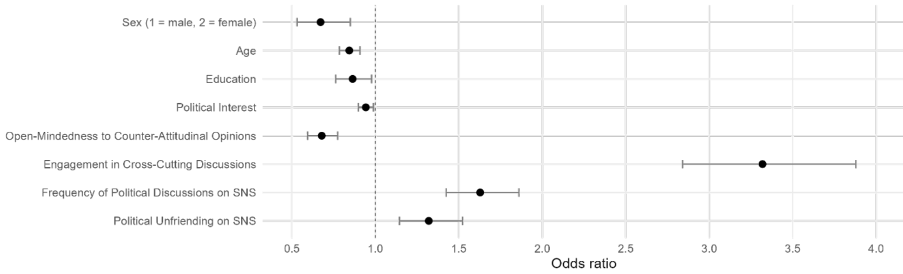
	1	2	3	4	5	6	7	8
1. Conflict-Oriented Behavior	—							
2. Sex (1 = male, 2 = female)	−.121***	—						
3. Age	−.058**	−.184***	—					
4. Education	−.115***	.027	−.009	—				
5. Political Interest	.083***	−.285***	.238***	.124***	—			
6. Open-Mindedness to Counter-Attitudinal Opinions	−.168***	.004	−.090***	.062**	.041	—		
7. Engagement in Cross--Cutting Discussions	.498***	−.112***	.033	−.088***	.153***	−.090***	—	
8. Frequency of Engagement in Political Discussions on SNS	.384***	−.187***	.068**	−.049*	.440***	−.110***	.476***	—
9. Political Unfriending on SNS	.229***	−.070**	.027	−.043	.166***	−.198***	.235***	.307***

Note. Two-tailed correlations are reported. *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

*Re*sults

After excluding cases with missing values, 1,927 cases were entered in the analysis. All three blocks significantly improved the model, with the third block improving it the most (Δ Nagelkerke $R2 = 0.326$), leading to Nagelkerke $R2 = 0.415$ in the full model. Figure 1 summarizes the results of the full model, and the model summaries and coefficients for each block entered are reported in full in Table 2.

Figure 1. Results of the model



The analysis showed that those engaging in cross-cutting discussions have the highest odds of participating in conflict-oriented behavior (*OR* = 3.319, $p < .001$, 95% CI [2.840, 3.879]), followed by a higher frequency of political discussions on SNS (*OR* = 1.628, $p < .001$, 95% CI [1.424, 1.860]) and performing political unfriending (*OR* = 1.320, $p < .001$, 95% CI [1.145, 1.522]). Therefore, we found evidence in favor of both H1 and H5 and contrary to H2. All other study variables were negatively associated with conflict-oriented behavior. In other words, it was more common among those who did not participate in conflict-oriented behavior to be more open-minded to counter-attitudinal opinions (*OR* = 0.679, $p < .001$, 95% CI [0.595, 0.775]). Thus, no support was found for H3.

Additionally, being male also proved to be associated with higher odds of behaving in a conflict-oriented way in SNS discussions (*OR* = 0.672, $p < .001$, 95% CI [0.531, 0.850]), while women are less likely to engage in such discussions, possibly due to their tendency to avoid hostility. Being more educated (*OR* = 0.864, $p = .021$, 95% CI [0.763, 0.978]), older (*OR* = 0.844, $p < .001$, 95% CI [0.784, 0.908]), and interested in politics (*OR* = 0.943, $p = .016$, 95% CI [0.899, 0.989]) were less articulated in terms of odds ratios, but still significant for those who do not participate in conflict-oriented behavior. These results suggest that men and younger individuals with lower levels of education and low interest in politics

are more likely to engage in heated or confrontational online exchanges. Lastly, H4 was not supported.

Table 2. Results of the Hierarchical Binary Logistic Regression

				95% CI for *OR*					95% CI for *OR*					95% CI for *OR*	
	B	*p*	*OR*	Lower	Upper	*B*	*p*	*OR*	Lower	Upper	*B*	*p*	*OR*	Lower	Upper
Sex (1 = male, 2 = female)	−0.539	<.001	0.583	0.483	0.705	−0.469	<.001	0.626	0.513	0.764	−0.398	<.001	0.672	0.531	0.850
Age	−0.091	.002	0.913	0.860	0.968	−0.146	<.001	0.864	0.812	0.920	−0.170	<.001	0.844	0.784	0.908
Education	−0.240	<.001	0.787	0.711	0.871	−0.257	<.001	0.773	0.696	0.859	−0.146	.021	0.864	0.763	0.978
Political Interest						0.078	<.001	1.081	1.041	1.122	−0.059	.016	0.943	0.899	0.989
Open-Mindedness to Counter--Attitudinal Opinions						−0.422	<.001	0.656	0.588	0.731	−0.387	<.001	0.679	0.595	0.775
Engagement in Cross-Cutting Discussions											1.200	<.001	3.319	2.840	3.879
Frequency of Engagement in Political Discussions on SNS											0.487	<.001	1.628	1.424	1.860
Political Unfriending on SNS											0.278	<.001	1.320	1.145	1.522
Constant	2.082	<.001	8.020			3.133	<.001	22.947			−0.469	.413	0.625		
Hosmer and Lemeshow Test	χ2 = 5.250; *df* = 8; *p* = .731					χ2 = 9.446; *df* = 8; *p* = .306					χ2 = 10.516; *df* = 8; *p* = .231				
Nagelkerke R^2	.040					.089					.415				

DISCUSSION

Building on prior research and theoretical assumptions that address the consequences of a harsher communication style in the online public sphere, we investigated predictors of the usage of conflict-oriented behavior in discussions on SNS. Some scholars (e.g., Chen, 2017; Rossini & Maia, 2021; Rossini, 2022) view such behavior, including incivility and verbal confrontation, as a natural element of online discourse that can foster deliberation. Others, however, emphasize its potential democratic risks, such as radicalization and polarization (Hwang et al., 2014; Kim & Kim, 2019). These perspectives are not mutually exclusive but represent two sides of the same coin. While for some people harsh communication is essential for engagement in online discussions, others see it as a barrier to deliberation (Novotná et al., 2023).

Firstly, in line with previous research (Chen, 2017; Rossini & Maia, 2021), we found that conflict-oriented users tend to be more active in cross-cutting discussions. Such exchanges, grounded in disagreement, often involve incivility and verbal fighting that can help overcome communicative barriers between differing perspectives (Chen, 2017; Marchal, 2021; Rossini, 2022; Rossini & Maia, 2021; Sun et al., 2021). Similarly, conflict-oriented commenters might be fascinated by disagreement, and cross-cutting discussions can provide them with a discussion arena to enjoy heated discussions (Vraga, 2015). However, a harsher communication style may discourage participation among less assertive individuals who prefer to avoid conflict (Mutz, 2002; Kim et al., 2021). Consequently, those who refrain from conflict-oriented behavior tend to participate less in online political discussions, possibly due to their sensitivity to hostility and doubts about the effectiveness of online dialogue for meaningful exchange or public awareness.

Secondly, we examined political unfriending as a less pro-deliberative response to disagreement, as it can reduce exposure to diverse perspectives (Goyanes et al., 2021). Contrary to our expectations, SNS users who engage in conflict-oriented behavior are more likely to unfriend others. Previous research indicates that one of the key drivers of political unfriending is the avoidance of confrontation and uncivil discussions (Goyanes et al., 2021). Our findings extend this understanding by showing that even conflict-oriented users may use unfriending as a strategy to manage their online environment. By engaging in uncivil exchanges more often, they may be more likely to experience escalations that prompt boundary-setting when content is perceived as norm-violating, leading to selective behavior. Thus, even though they are active in cross-cutting discussions, they simultaneously constrain heterogeneous exposure.

Thirdly, contrary to our assumptions that conflict-oriented users could be more open-minded to cross-cutting discussions, we found that those who use a harsher communication style tend to be less tolerant of disagreement than those who

are not conflict-oriented. This suggests that conflict-oriented users in Czechia tend to hold undemocratic and negative attitudes toward deliberative dialogue. The optimistic view that incivility can serve as an expressive form of argument, fostering mutual understanding (Rossini, 2022), appears to be only partially supported, as these users often perceive cross-cutting discussions as meaningless. Although they engage more frequently in such exchanges, they lack openness and hope for mutual understanding – key benefits typically associated with dialogue based on disagreement (Chen, 2017).

People's negative views towards discussions that involve different perspectives may stem from their skepticism, which could be a result of negative encounters that they had in online conversations with people who hold opposing views. Such a strong and negative experience can lead to concerns about the value and purpose of engaging in dialogue with individuals who hold different beliefs. This assumption stems from negative affective responses, including frustration and stress, which often accompany engagement in online cross-cutting discussions (Anderson & Auxier, 2020; Rosenberg, 2024).

Fourthly, while the frequency of engagement in political discussions is positively related to conflict-oriented behavior, political interest is evidently not. We built our hypotheses on the expectation that users who engage in conflict-oriented behavior are also more interested in politics. However, we found evidence for the opposite direction, as conflict-oriented behavior relates to lower political interest. Moreover, this finding weakens the deliberative potential of conflict behavior usage, and it must be carefully reconsidered that, although conflict-oriented users encounter dialogue and political discussions, their interest in politics is lower.

Looking from the other side, participants with non-conflict-oriented behavior in discussions who are politically interested avoid participation in political discussions, which could be caused by their senseless or uncivil character (see Goyanes et al., 2021). While political interest involves actively learning about public affairs, participating in discussions requires no such knowledge. Research shows that online disagreement tends to attract people who are less politically interested but enjoy conflict (Vraga, 2015). As a result, the absence of informed participants can lower the quality of debate and foster uncivil exchanges marked by limited understanding. This is especially concerning during crises, such as the COVID-19 pandemic, the Russia–Ukraine war, or other global issues, like the climate crisis or migration waves, when discussions become highly polarized (Schmid et al., 2022). Ultimately, discouraging participation by politically aware individuals risks deepening polarization and amplifying one-sided views.

Lastly, consistent with prior research, conflict-oriented behavior is more typical among men (Proust & Saldaña, 2022), whereas women, who often avoid hostility in political discussions, engage less in such exchanges (Abendschön & García-Albacete, 2021). Participants displaying conflict-oriented behavior also

tend to be younger and have lower education levels, groups that have previously been found to enjoy heated debates more and show less concern about incivility (Min & Shen, 2023). Also, young adults usually enjoy heated discussions more and find them joyful (Vraga, 2015).

Our study is not without limits. The measurement of conflict-oriented behavior is broadly defined and assessed by only two items. These items have their own limitations, such as lower reliability due to the difficulty of recalling online discussions and the different perceptions of what is considered uncivil. More precisely, we admit that various types of conflict-oriented behavior should be examined (e.g., intolerance, hate speech, toxicity) regarding participation in online discussions. Additionally, recoding the two items of conflict-oriented behavior into one variable that captures only the performance and non-performance of such a behavior limits nuances in interpretation, especially if the aim was to look at the differences in intensity of this behavior. We believe that our study makes a valuable contribution to the field because it highlights specific trends related to conflict-oriented behavior in the under-researched CEE region, particularly in a country shaped by a post-communist legacy. This historical background may shape how citizens engage in conflict-oriented discussions, perceive disagreement in the public sphere, and participate in civic and online communication during times of crisis. Our findings may be transferable to countries with similar post-communist legacies and low institutional trust (Pospěch, 2021). As the data were collected during a crisis, conflict behavior may have been context-dependent; future research should therefore examine how these patterns emerge in other crises or non-crisis settings. However, we cannot rule out the causal effects and their directions. Thus, we cannot say if particular activities fuel conflict-oriented behavior within SNS or if conflict-oriented participants seek certain activities – further research could be done. Additionally, exploring the barriers that prevent individuals who shy away from conflict from taking part in online discussions would be beneficial. This could explain why those with a strong political interest may not engage in these conversations, despite being open to different perspectives.

Our research highlights the double-edged sword of conflict-oriented behavior in online discussions. We argue that this finding adds an important dimension to scholarly debates that value high participation in cross-cutting discussions but often overlook how participants' perceptions of these exchanges can fundamentally shape their overall benefits and, consequently, influence the quality of deliberative democracy. This phenomenon underscores the need for strategies that balance open dialogue with respectful discourse to foster a healthier public sphere and call for problematic content moderation. Artificial intelligence has recently become a widely discussed tool in this process, although its role in online deliberation remains disputed. It can promote civility and filter harmful content,

but also risks reinforcing biases, overlooking context, and silencing underrepresented voices (Carstens & Friess, 2024), underscoring the importance of research on the predictors of incivility and other contributing factors.

Funding: The work on this article was supported by the NPO "Systemic Risk Institute" No. LX22NPO5101, funded by European Union – Next Generation EU (Ministry of Education, Youth and Sports, NPO: EXCELES).

Acknowledgments: Grammar corrections and style refinements were made with ChatGPT 4.0 and Grammarly.

REFERENCES

Abendschön, S., & GarcíaAlbacete, G. (2021). It's a man's (online) world. Personality traits and the gender gap in online political discussion. *Information, Communication & Society, 24*(14), 2054–2074. https://doi.org/10.1080/1369118X.2021.1962944

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of ComputerMediated Communication*, *19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Anderson, M., & Auxier, B. (2020). *55% of U.S. social media users say they are 'worn out' by political posts and discussions*. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/08/19/55-of-u-s-social-media-users-say-they-are-worn-out-by-political-posts-and-discussions/

Bormann, M. (2022). Perceptions and evaluations of incivility in public online discussions—Insights from focus groups with different online actors. *Frontiers in Political Science*, *4*. https://doi.org/10.3389/fpos.2022.812145

Chen, G. M. (2017). *Online Incivility and Public Debate: Nasty talk*. Palgrave Macmillan.

Carstens, J. A., & Friess, D. (2024). AI within online discussions: Rational, civil, privileged? Minds and Machines, *34*(2). https://doi.org/10.1007/s11023-024-09658-0

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Duggan, M., & Smith, A. (2016). *The Political Environment on Social Media*. Pew Research Center. https://www.pewresearch.org/internet/2016/10/25/the-political-environment-on-social-media/

Elster, J. (1998). *Deliberative democracy*. Cambridge University Press.

Frischlich, L., SchattoEckrodt, T., Boberg, S., & Wintterlin, F. (2021). Roots of Incivility: How Personality, Media Use, and Online Experiences Shape Uncivil Participation. *Media and Communication*, *9*(1), 195–208. https://doi.org/10.17645/mac.v9i1.3360

Gao, Y., Qin, W., Murali, A., Eckart, C., Zhou, X., Beel, J. D., Wang, Y., & Yang, D. (2024). A Crisis of Civility? Modeling incivility and its effects in political discourse online. *Proceedings of the International AAAI Conference on Web and Social Media*, *18*(1), 408–421. https://doi.org/10.1609/icwsm.v18i1.31323

Gervais, B. T. (2015). Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-based Experiment. *Journal of Information Technology & Politics*, *12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Gervais, B. T. (2019). Rousing the Partisan Combatant: Elite Incivility, Anger, and Antideliberative Attitudes. *Political Psychology*, *40*(3), 637–655. https://doi.org/10.1111/pops.12532

Goyanes, M., Borah, P., & Gil de Zúñiga, H. (2021). Social media filtering and democracy: Effects of social media news use and uncivil political discussions on social media unfriending. *Computers in Human Behavior*, *120*. https://doi.org/10.1016/j.chb.2021.106759

Gutmann, A., & Thompson, D. (1996). *Democracy and Disagreement*. Harvard University Press.

Hmielowski, J. D., Hutchens, M. J., & Cicchirillo, V. J. (2014). Living in an age of online incivility: examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society*, *17*(10), 1196–1211. https://doi.org/10.1080/1369118X.2014.899609

Hwang, H., Kim, Y., & Huh, C. U. (2014). Seeing is Believing: Effects of Uncivil Online Debate on Political Polarization and Expectations of Deliberation. *Journal of Broadcasting & Electronic Media*, *58*(4), 621–633. https://doi.org/10.1080/08838151.2014.966365

Hwang, H., Kim, Y., & Kim, Y. (2016). Influence of Discussion Incivility on Deliberation: An Examination of the Mediating Role of Moral Indignation. *Communication Research*, 1–28. https://doi.org/10.1177/0093650215616861

Kenski, K., Coe, K., & Rains, S. A. (2020). Perceptions of Uncivil Discourse Online: An Examination of Types and Predictors. *Communication Research*, *47*(6), 795–814.
https://doi.org/10.1177/0093650217699933

Kim, Y., & Kim, Y. (2019). Incivility on Facebook and political polarization: The mediating role of seeking further comments and negative emotion. *Computers in Human Behavior*, *99*, 219–227. https://doi.org/10.1016/j.chb.2019.05.022

Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The Distorting Prism of Social Media: How SelfSelection and Exposure to Incivility Fuel Online Comment Toxicity. *Journal of Communication*, *71*(6), 922–946. https://doi.org/10.1093/joc/jqab034

Koban, K., Stein, J.-P., Eckhardt, V., & Ohler, P. (2018). Quid pro quo in Web 2.0. Connecting personality traits and Facebook usage intensity to uncivil commenting intentions in public online discussions. *Computers in Human Behavior*, *79*, 9–18. https://doi.org/10.1016/j.chb.2017.10.015

Lee, J., Choi, J., & Kim, J. (2021). Effects of online incivility and emotions toward in-groups on cross-cutting attention and political participation. *Behavior & Information Technology*, 1–15. https://doi.org/10.1080/0144929X.2021.1969429

Lu, Y., & Lee, J. K. (2021). Determinants of cross-cutting discussion on Facebook: Political interest, news consumption, and strong-tie heterogeneity. New Media & Society, *23*(1), 175–192. https://doi.org/10.1177/1461444819899879

Marchal, N. (2021). "Be Nice or Leave Me Alone": An Intergroup Perspective on Affective Polarization in Online Political Discussions. *Communication Research*, *49*(3), 376–398. https://doi.org/10.1177/00936502211042516

Mutz, D. C. (2002). The Consequences of CrossCutting Networks for Political Participation. *American Journal of Political Science*, *46*(4), 838–855. https://doi.org/10.2307/3088437

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Mutz, D. C. (2016). *In-your-face Politics: The Consequences of Uncivil Media*. Princeton University Press.

Neubaum, G., Cargnino, M., & Maleszka, J. (2021). How Facebook Users Experience Political Disagreements and Make Decisions About the Political Homogenization of Their Online Network. *International Journal of Communication*, *15*, 187–206.

Macková, A., Novotná, M., Procházková, K., Macek, J., & Hrbková, L. (2021). Češi na sítích, důvěra a polarizace v době pandemie. [Czechs on social networking sites, trust and polarization in pandemic]. https://doi.org/10.13140/RG.2.2.20950.78401/1

Min, C., & Shen, F. (2023). Online incivility, argument quality and public expression in China: Exploring the moderating role of education level and opinion congruency. *Telematics and Informatics*, *82*, 102010. https://doi.org/10.1016/j.tele.2023.102010

Novotná, M., Macková, A., Bieliková, K., & Rossini, P. (2023). Barriers to participation in polarized online discussions about COVID-19 and the RussoUkrainian war. *Media and Communication, 11(*3), 274–284. https://doi.org/10.17645/mac.v11i3.6657

Orhan, Y. E. (2022). The relationship between affective polarization and democratic backsliding: Comparative evidence. *Democratization*, *29(*4), 714–735. https://doi.org/10.1080/13510347.2021.2008912

Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Peña, J., & Brody, N. (2014). Intentions to hide and unfriend Facebook connections based on perceptions of sender attractiveness and status updates. *Computers in Human Behavior*, *31*, 143–150. https://doi.org/10.1016/j.chb.2013.10.004

Pospěch, P. (2021). *Neznámá společnost: Pohledy na současné* Česko [Unknown Society: Contemporary views of the Czech Republic]. Host.

Proust, V. & Saldaña, M. (2022). "What a Nasty Girl!" Incivility and gendered symbolic violence in news discussions. *Feminist Media Studies*, 1–23. https://doi.org/10.1080/14680777.2022.2142640

Rainie, L., & Smith, A. (2012). *Politics on Social Networking Sites.* Pew Research Center. https://www.pewresearch.org/internet/2012/09/04/politics-on-social-networking-sites/

Rains, S. A., Kenski, K., Coe, K., & Harwood, J. (2017). Incivility and political identity on the internet: Intergroup factors as predictors of incivility in discussions of news online. *Journal of ComputerMediated Communication*, *22*(4), 163–178. https://doi.org/10.1111/jcc4.12191

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing [Computer software]. http://www.R-project.org/

Rosenberg, A. (2024). "Incivility makes me angrier than uncivil disagreement": a survey experiment using news comments. *Journal of Information Technology & Politics*, 1–15. https://doi.org/10.1080/19331681.2024.2407636

Rossini, P., & Stromer-Galley, J. (2019). Citizen deliberation online. In E. Suhay, B. Grofman, & A. H. Trechsel (Eds.), *The Oxford handbook of electoral persuasion* (pp. 689–712). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190860806.013.14

Rossini, P. (2022). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*, *49*(3), 399–425. https://doi.org/10.1177/0093650220921314

Rossini, P., & Maia, R. C. M. (2021). Characterizing Disagreement in Online Political Talk: Examining Incivility and Opinion Expression on News Websites and Facebook in Brazil. *Journal of Deliberative Democracy*, *17*(1), 90–104. https://doi.org/10.16997/10.16997/jdd.967

Rösner, L., Winter, S., & Krämer, N. C. (2016). Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior*, *58*, 461–470. https://doi.org/10.1016/j.chb.2016.01.022

Schmid, F., Treib, O., & Eckardt, F. (2022). The virus of polarization: online debates about Covid-19 in Germany. *Political Research Exchange*, *5*(1). https://doi.org/10.1080/2474736x.2022.2150087

Stryker, R., Conway, B. A. & Danielson, J. T. (2016). What is political incivility? *Communication Monographs*, *83*(4), 535–556. https://doi.org/10.1080/03637751.2016.1201207

Sun, Q., Wojcieszak, M., & Davidson, S. (2021). OverTime Trends in Incivility on Social Media: Evidence From Political, NonPolitical, and Mixed Sub Reddits Over Eleven Years. *Frontiers in Political Science*, *3*. https://doi.org/10.3389/fpos.2021.741605

Sydnor, E. (2019). *Disrespectful democracy: The psychology of political incivility.* Columbia University Press.

Švelch, J., & Vochocová, L. (2015). Sociální média jako nová výzva pro výzkum politické participace [Social Media as a New Challenge for Political Participation Research]. *Sociologický Časopis / Czech Sociological Review, 51*(1), 65–87. http://doi.org/10.13060/00380288.2015.51.1.154

Van Aelst, P., Toth, F., Castro, L., Štětka, V., De Vreese, C. H., Aalberg, T., Cardenal, A. S., Corbu, N., Esser, F., Hopmann, D. N., KocMichalska, K., Matthes, J., Schemer, C., Sheafer, T., Splendore, S., Stanyer, J., Stępińska, A., Strömbäck, J., & Theocharis, Y. (2021). Does a Crisis Change News Habits? A Comparative Study of the Effects of COVID-19 on News Media Use in 17 European Countries. *Digital Journalism*, *9*(9), 1208–1238. https://doi.org/10.1080/21670811.2021.1943481

Vargo, C. J., & Hopp, T. (2017). Socioeconomic Status, Social Capital, and Partisan Polarity as Predictors of Political Incivility on Twitter. *Social Science Computer Review*, *35*(1), 10–32.
https://doi.org/10.1177/0894439315602858

Vochocová, L., & Rosenfeldová, J. (2019). Ridiculed, but safe: What e-mothers' discussion on migration tells us about the potential of "third spaces' for the political communication of women. *European Journal of Communication*, *34*(2), 142–158. https://doi.org/10.1177/0267323118810865

Vraga, E. K., Thorson, K., KliglerVilenchik, N., & Gee, E. (2015). How individual sensitivities to disagreement shape youth political expression on Facebook. *Computers in Human Behavior*, *45*, 281–289. https://doi.org/10.1016/j.chb.2014.12.025

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer.

Zhu, Q., & Skoric, M. M. (2021). From Context Collapse to "Safe Spaces": Selective Avoidance through Tie Dissolution on Social Media. *Mass Communication and Society*, *24*(6), 892–917. https://doi.org /10.1080/15205436.2021.1883671

ORIGINAL ARTICLE

DOI: 10.51480/1899-5101.18.3(41).783

Navigating Narratives: The Dynamics of Disinformation Exposure and Believability to Information in Kosovo

Dren Gërguri

University of Prishtina, Kosovo

Darren Lilleker

Bournemouth University, United Kingdom

Abstract: This paper examines individuals' exposure to disinformation in Kosovo and how they judge its credibility. We investigated exposure to and belief in six of the most popular disinformation narratives through a survey of 600 respondents. The results show that trust in news media does not significantly influence disinformation belief, suggesting that cognitive and situational factors are more impactful. Although social media remains the main source of exposure, the source typically does not affect belief unless there are some high-salience political stories. Additionally, repeated exposure to a narrative increases the chance that it will be believed. Fact-checking and media literacy efforts have limited reach, as most respondents rarely verify information that aligns with their views. These findings highlight the need to improve media literacy and critical thinking education to help individuals become more discerning in how they evaluate information.

Keywords: disinformation, believability, exposure, narratives, trust

INTRODUCTION

The circulation of disinformation within any society poses a threat to the notions of truth and fact, thereby undermining the principles of informed deliberation at the heart of democratic life (McKay & Tenove, 2021). We define disinformation as intentionally false or manipulated content designed to deceive; disinformation differs from misinformation, which is misleading but circulates without the intention to deceive others (van der Linden, 2022). Disinformation is a threat

to transparency, institutional trust, and polarization because it blends facts and falsehoods (Iyengar & Massey, 2019). The circulation of misinformation and disinformation, particularly across digital platforms, has led to the current era being defined as the post-truth age, where truth becomes contingent on believability, not its evidential basis (Lilleker, 2018). Although the term post-truth has a short lineage, disinformation has long been a feature of the communication environment and is not purely a feature of politics in the digital age. Disinformation now reaches wider audiences and shapes public opinion by undermining their ability to distinguish truth from falsehood (Bastick, 2021). Our paper examines the ability of citizens in the Kosovan highly polarized environment where disinformation regularly circulates to determine what forms of disinformation are seen as believable, if there are socio-demographic or source-based explanations for belief in disinformation, and how repeated exposure impacts the likelihood of believability. We explore this specific context to develop more widely applicable findings to help understand the dynamics of post-truth environments. The disinformation in Kosovo is reflective of the trends worldwide, and the information landscape is polluted by the spread of manipulated information in almost every field, such as science (Scheufele & Krause, 2019), health (Suarez-Lledo & Alvarez-Galvez, 2021), and politics (Jerit & Zhao, 2020).

One essential component of disinformation is the narrative of a specific story and the fit within a broader meta-narrative, for example, low trust between ethnic communities, which enhances the believability of the single piece of disinformation (Bánkuty-Balogh, 2021). Such narratives have the capacity to persuade people and influence their beliefs and behaviors, making them effective in eroding people's trust in institutions and undermining the truth. Studies highlight how disinformation can prove sticky, through repetition, and through fitting to highly prominent myths or tropes, and thus can have destructive consequences on both individual and societal levels, such as undermining trust in democratic institutions (Lewandowsky & van der Linden, 2021) or increasing vaccine hesitancy during the COVID-19 pandemic (Pierri et al., 2022). Hence, examining the believability of these narratives offers insights into the quality and stability of pro-democratic attitudes.

KOSOVO: CASE STUDY CONTEXT

Kosovo is a country where ethnic tensions and political polarization are constant, and often, the political elites may use these divisions to their advantage (Qerimi & Gërguri, 2022). In Kosovo, the trend of low trust towards institutions remains stable, with 33% trusting the Kosovo Government and even less (25%) trusting the Assembly (Avdiu 2022). This paper analyzes how citizens navigate the complex

post-truth environment by examining key disinformation narratives that have been independently determined as false by fact-checkers, who aim to reduce the spread and impact of false information to the public (Kuś & Barczyszyn--Madziarz, 2020).

With few exceptions, there is limited research on how citizens read disinformation narratives (Suau & Puertas-Graell, 2023), as most research explores the spread of disinformation during elections (Chen et al., 2021; Pierri et al., 2020) or other major political or societal events (Perez-Escolar et al., 2023; Reddi, et al., 2023). Previous studies show that exposure to factual information related to major political or societal events can lead people to reject false claims (David, 2023; de Saint Laurent, et al., 2022), for instance, exposure to the debunking of conspiracy theories related to COVID-19 vaccination leads to a higher willingness to accept a vaccination (Buturoiu et al., 2021). However, in this study, we conducted the research at a point when there were no major socio-political events or controversies taking place and focused on narratives that regularly circulate within the Kosovar information environment. The study is interested in the way its citizens interact and analyze information during their daily media intake by studying disinformation under relatively normal conditions. It is concerned with the common patterns of information processing – how individuals evaluate the credibility and factuality of the news, whether they believe what they read, and the relationship with judgments related to the reliability of the media sources. These are the daily cognitive and behavioral reactions, exploring them indicates how disinformation can subtly influence public attitudes during their everyday consumption of news. Hence, this paper contributes to knowledge by showing how disinformation can impact citizens' evaluations of information accuracy, credibility, and trust in sources during a period of relative normality, within the complex Kosovar information environment. Rather than focusing on general political views or orientations towards distinct ideologies, the research explores the mechanisms that shape the believability of disinformation.

NAVIGATING A 'DISORDERED' INFORMATION ENVIRONMENT

A large amount of research has been devoted to what is termed information pollution and information disorder. These terms describe a situation where inaccurate information circulates widely within a society and impacts the attitudes and behaviors of citizens. Citizens of democratic societies must be able to recognize the difference between facts, ideological interpretations of data, opinions, and outright lies. The problem is that in many societies, particularly polarized environments such as Kosovo, a variety of actors spread claims of fact that have some degree of inaccuracy, leading to a mistrustful and cynical citizenry.

There are debates regarding the scale of the problem of disinformation, although largely based on evidence from studies in Western Europe or America. There is evidence that disinformation is produced and shared by a limited number of users of social media, and the problem is exaggerated through the attention given to isolated but dramatic examples (Allen et al., 2020). However, the increased focus on disinformation during the COVID-19 pandemic revealed that the spread of disinformation could have serious implications (Roozenbeek et al, 2020). The most serious cases are when preventative measures such as lockdowns or vaccinations become politicized (Lilleker et al, 2021). In nations with highly polarized political environments, low trust in political and media institutions, and where competing arguments were given equal weight within the information environment, there were higher degrees of uncertainty about core facts relating to the pandemic. The evidence from national studies of the spread of misinformation and disinformation during the pandemic reinforces longstanding notions of the problem when the information environment becomes polluted with false claims (Pérez Escolar et al., 2023). Citizens must have some form of mental schema for testing what facts and sources of facts are credible and reliable, perhaps based on their perceptions of what news sources are credible and trustworthy. However, where the media is highly mistrusted, the blurring of fact and fiction becomes prevalent and highly problematic, more so when that blurring is strategically manufactured for political gain. When there is low trust, citizens rely on confirmation and selection bias and strengthen already held opinions (Thorbjørnsrud & Figenschou, 2022). Within the polarized and low-trust Kosovar information environment, we hypothesize that:

- *H1: People who have a high level of trust in news media are more likely to believe in narratives that are disseminated by their trusted source.*

It is long observed that the trend is for more people to access news via social media (Nielsen & Schrøder, 2014; Swart, 2023). However, this reliance introduces significant challenges for knowledge outcomes, and social media has also been found to be a significant source of disinformation (Olan et al, 2024). The studies show that although digital interventions can have a moderate positive effect on the capacity of a user to differentiate credible and fake news (Guess et al., 2020), the overall impact of social media use is mixed, and some of its consequences are harmful to digital literacy (Chen et al., 2021). The detection of disinformation remains underdeveloped and usually reliant on individual user reporting, and hence it can spread rapidly within information environments where the conditions are ripe (Diaz Ruiz & Nilsson, 2023). In low-trust environments, ordinary users are often more trusted than official sources and news shared by close peers is viewed as an endorsement (Karlsen & Aalberg,

2023). Arguably, when disinformation is shared by other users, it becomes more credible in an environment such as Kosovo, where societal trust is crucial. The digital environment provides a space where any individual, mainstream media organization, digital 'alt-news' outlet, or political actor can post or share content without gatekeepers. The individual alone decides which sources they use and what content they believe (Melchior & Oliveira, 2024). Although people tend to use digital/social and traditional media, we are particularly concerned with the inverse relationship between a major dependence on unverified social media platforms and the susceptibility to being misled by disinformation. Therefore, we hypothesize:

- *H2: Kosovars who are more reliant on gathering news from social media platforms are more likely to find disinformation more believable than those who rely on digital or traditional media (newspaper, radio, and TV).*

People are exposed to disinformation on different channels within the information environment: the broader media ecosystem encompassing traditional, digital, and social media platforms. However, extant research suggests that mere exposure does not always result in persuasion (Su et al, 2022). Thus, our study concentrates on the question of the socio-political environment's mediating role: to what extent does a highly polarized environment predetermine the probability of repeated exposure to disinformation being converted into influence on citizens' attitudes and beliefs. Recent work has found that emotions are shaped by exposure to certain frames (David, 2023) and that exposure can complement impulses for confirmation bias (Xu et al, 2023). The pre-existing beliefs and attitudes among individuals contribute to their selective exposure and selective attention, which direct human beings to content, including disinformation, which supports their prior held beliefs and attitudes (Stroud, 2008).

Repeated exposure to narratives over time leads them to be accepted as truths, regardless of how false the content is (Guess et al., 2020). Hence, repeatedly being exposed to false narratives can lead people to develop distorted worldviews, which become confirmed through their selective use of sources and desire for confirmation bias. These psychological factors, often shaped by lived experiences within a particular socio-political environment, can make certain forms of disinformation especially resonant if they circulate frequently. Therefore, we hypothesize that:

- *H3: Kosovars who have previously encountered a disinformation narrative are more inclined to believe it compared to those who are unfamiliar with it.*

Extant research demonstrates that those who mistrust institutions, rely on social media for news, and are frequently exposed to disinformation will find it more believable. However, some agencies actively attempt to combat the spread of misinformation and disinformation and attempt to debunk specific claims while inoculating citizens to reduce their future susceptibility (Van der Linden et al., 2017). Fact-checkers are a key defence against the spread of false content and attempt to ensure their results gain publicity. There are limits to how many claims can be checked, hence, fact-checkers focus on high-profile individuals and claims, attempting to prevent some false claims from gaining traction (Birks, 2019). However, often the repetition of a claim is sufficient for it to be more memorable than the counterargument; also, counterclaims can be dismissed by citizens when the original claim confirms their existing beliefs. Hence, any form of debunking, including the work of fact-checkers, is not wholly successful (Hameleers, 2024). Brautovic and John (2023) also found that fact-checkers were ineffective, as their website and social media pages were rarely visited. Media literacy initiatives attempt pre-bunking strategies, making people aware of common tropes or sources of disinformation, and are designed to engage the critical faculties of citizens (Bertolotti & Catellani, 2023). The challenge for initiatives countering the spread of disinformation is reaching sufficient numbers of the population (Birks, 2019) and engaging with those who have low trust in societal institutions (Su et al., 2022). Disengagement with systems designed to protect the information environment is much more likely within societies where trust is low, and polarization is high, spaces where selection bias and confirmation bias are equally likely to be high, such as Kosovo.

Individuals tend to fact-check information that opposes their preexisting opinions but not that which supports them, which is also in line with motivated reasoning theory (Lewandowsky et al., 2020). Meanwhile, prior beliefs also influence perceptions of the accuracy: people are more likely to believe information that supports their worldview even though it may not be factual (Gërguri, 2022). To differentiate these two processes – behavioral engagement (fact-checking) and cognitive evaluation (believability), we hypothesize:

- *H4: Kosovars are less likely to verify information that supports their existing views.*

METHODOLOGY

To find the most relevant disinformation narratives, an analysis was conducted during a period of relative normality in Kosovo; there were no election campaigns, controversies, or tensions in the situation on the border with Serbia. For 18 days, from 10th to 28th April 2023, 86 fact-checked articles were collected from Kosovo's two IFCN-affiliated fact-checkers, *Krypometer*[1] and *Hibrid*[2].

The initial stage involved analyzing the articles and identifying the six most repeated disinformation narratives during this period (see Table 1). To measure the prominence of disinformation narratives, we conducted a content analysis, exploring how frequently those narratives appeared within fact-checked claims during the analyzed period. Of 86 fact-checked articles, 49 articles were related to Kosovo-Serbia relations, including the topic of ethnic and political conflicts, and foreign actors' involvement.

A survey was then conducted, which tested the familiarity and believability of the six most prominent disinformation narratives. Following an initial set of sociodemographic questions (including party identification, ideology), respondents were asked about media consumption sources and the frequency of news consumption in order to measure media engagement. Media trust was measured by asking "do you trust the information published by news media/digital media/social media?" using a 7-point scale (1 = never trust, 7 = always trust). In the second part, for each narrative, there were four questions asking respondents about exposure, source and believability following our hypotheses. Each narrative was followed by the questions: (1) Had the respondent read the narrative before? (2) Does the respondent agree with the statement; on which media or platform did they read it, choosing only one option (if this was the case)? and (3) Had they shared related content?

The identified narratives are representative of common disinformation themes. Unsurprisingly, four of them (1, 2, 4, & 5) were concerned with Kosovo-Serbia relations, ranging from political claims to religious issues. Stories relating to relations with Serbia are a constant theme of news, and disinformation relating to these issues is a permanent feature of the Kosovar information environment, even when there are no obvious tensions between the two countries. The other repeated narratives are about migration (N6), relating to job opportunities for young people thinking of migrating into the EU, and celebrity gossip (N3) related to the arrest of a celebrity and reinforcing the narrative that celebrities often enjoy privileges when dealing with state institutions and so this narrative indirectly undermines trust in institutions.

[1] Krypometer was launched in December 2016 and is part of the online media, Kallxo.com.
[2] Hibrid was founded in September 2020 by the NGO "Action for Democratic Society"

Table 1. The six most repeated disinformation narratives

Number	Narrative	Code	
1	The West was an accomplice in the ethnic cleaning of Serbs in Kosovo	Kosovo-Serbia relations	Political issue
2	Russian group, „Wagner", fights against Kosovo police in the North	Kosovo-Serbia relations	Political issue
3	The famous designer Valdrin Sahiti was arrested and is in custody in Turkey	Celebrities	Celebrity Gossip
4	Orthodox churches in Kosovo are under political attack	Kosovo-Serbia relations	Political issue/ Religion issue
5	Serbs in Kosovo face intimidation, physical attacks and are targets of crime in Kosovo meanwhile institutions don't deliver justice	Kosovo-Serbia relations	Political issue
6	A company is looking for workers in Germany, the company will organize work visas	Immigration	Employment issue/ Immigration issue

The narrative headlines were inserted into a survey, which was conducted online during June 2023. The survey was completed by 600 respondents among Kosovar internet users between 18 and 65 years old from seven main regions. In Kosovo, daily use of the internet is 95.6%, the highest in the region and higher than the 84% average for European Union member states (Eurostat 2023). There are variances in internet usage between different age groups and different education levels, with older and less educated people being less active online (STIKK 2019).

Table 2. Sociodemographic characteristics of the sample

PLACE OF RESIDENCE	%	AGE	%	GENDER	%	EDUCATION	%
Ferizaj	9.5	**18-29**	44.7	**Man**	50.3	**Elementary school**	1.3
Gjakova	6.7	**30-44**	44. 2	**Woman**	49.3	**High school**	13.5
Gjilan	8.8	**45-64**	11.2	**Other**	0.3	**University degree**	67.0
Mitrovica	10.7						
Peja	9.3						
Prishtina	45.0						
Prizren	10.0						

The digital divides by age and education are reflected in the sample of respondents (see Table 2). There is an imbalance when it comes to age and education, with a lower number of elderly people and also individuals with lower education included in the sample. However, the sample reflects Kosovars who are heavy users of social media platforms and so are most likely to be exposed to disinformation online as part of their daily information diet.

FINDINGS

TRUST IN MEDIA AND BELIEF IN DISINFORMATION NARRATIVES (H1)

The data demonstrates varying levels of trust and distrust in different narratives, reflecting the diversity of public attitudes on the issues and differing levels of critical thinking and fact-checking, which might be used to assess the credibility of information. Some respondents fall into the group of undecided, indicating a degree of skepticism or uncertainty about the accuracy of the disinformation, which sometimes is the main goal of those who disseminate these narratives (Gërguri, 2022). The narratives regarding international or political events (e.g., involvement of the West, Russian group "Wagner") tend to have higher levels of distrust, while the narrative about the company offering jobs and work visas in Germany has a more balanced distribution of trust levels, with a notable percentage of respondents believing the story.

In Kosovo, there are considerable differences in how much individuals trust the news media. The majority do not trust the news media, with 47.5% having very low or low trust, while only around 20% have high or very high levels of trust in the news media. To examine whether this trust influences susceptibility to disinformation, a series of chi-square tests of independence was conducted to test Hypothesis 1, which posited that individuals with higher trust in news media are less likely to believe in disinformation narratives. Trust in news media was dichotomized as tend to trust (scores 5–7) versus tend to distrust (scores 1–4) on a 7-point scale, and belief in each narrative was coded as believes (scores 5–8).

The results are presented in Table 3 and Figure 1. No significant associations were found between media trust and belief in any of the disinformation narratives (all $p > .29$), with uniformly small effect sizes (Cramér's $V < .05$). The largest—but still non-significant—difference was observed for the "Orthodox churches" narrative, where 12.5% of high-tend to trust respondents believed the story compared to 9.3% among low-tend to distrust respondents. Figure 1 illustrates these findings, showing near-identical belief rates across trust levels, with overlapping 95% confidence intervals.

Complementary correlation analyses, using continuous trust scores, reveal modest but statistically significant positive correlations for certain narratives. Specifically, trust in media correlated positively with belief in Narrative 1 ("The West was complicit in the ethnic cleansing of Kosovars") ($r = .128^{**}$, $p < .01$) and Narrative 5 ("Discrimination against Serbs") ($r = .089^{*}$, $p < .05$). These findings suggest that higher media trust may, in some instances, coincide with greater belief in particular disinformation narratives, possibly reflecting trust in domestic rather than international media sources or differing perceptions

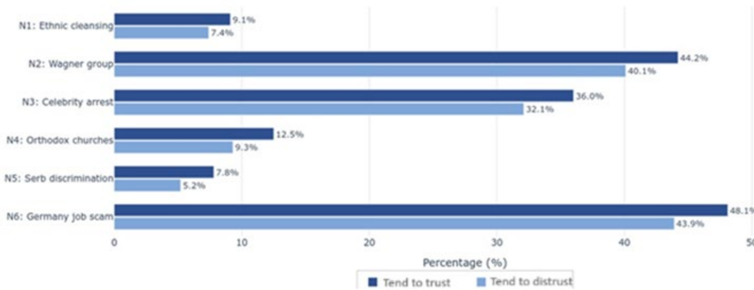
of what constitutes credible information. However, these exploratory linear correlations do not translate into significant categorical differences when the hypothesis is tested using binary trust and belief variables. Accordingly, the chi-square analyses provide no evidence of a robust relationship between media trust and belief in disinformation narratives.

Table 3. Association Between Media Trust and Belief in Disinformation Narratives

	Tend to trust (5–7)	Tend to distrust (1–4)	χ^2 (df=1)	*p*
N1	9.1%	7.4%	0.42	.517
N2	44.2%	40.1%	0.81	.368
N3	36.0%	32.1%	0.71	.399
N4	12.5%	9.3%	1.12	.290
N5	7.8%	5.2%	1.01	.315
N6	48.1%	43.9%	0.79	.374

Overall, H1 was not supported. Trust in news media does not reliably predict belief in disinformation narratives among respondents in Kosovo. This finding suggests that belief formation may be shaped more by other psychological or contextual factors, such as exposure to misleading content (H3) or ideological congruence with the narrative (H4), than by general levels of trust in traditional media institutions.

Figure 1. Belief in Disinformation by Media Trust Level



SOCIAL MEDIA'S ROLE IN SPREADING DISINFORMATION (H2)

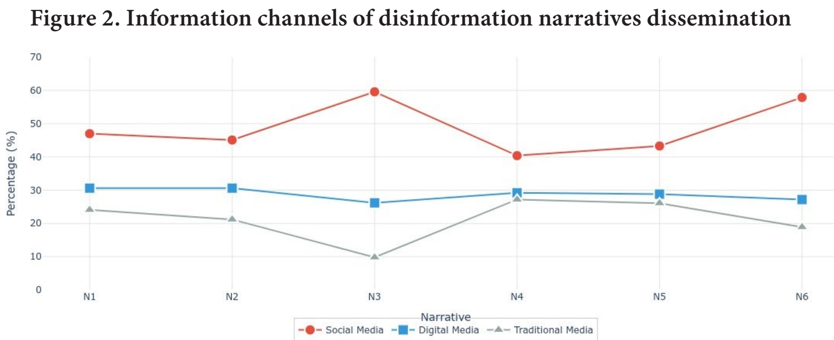
Despite the data representing a range of different topics, including disinformation, and that we know disinformation can be manifested in various forms and be disseminated through a range of media channels, it is striking that those who use social media to be informed are most likely to find these narratives familiar. Across all the narratives, social media consistently emerges as a primary channel for exposure to these disinformation narratives. This finding indicates

a significant portion of the audience encountered these narratives on platforms like Facebook, Instagram, and underscores the potential influence of social media in shaping public perceptions and discussions on various topics.

To test Hypothesis 2, a chi-square test was conducted among respondents who reported prior exposure to each narrative (n = 1,308 exposed responses). Belief was dichotomized (5–8 = believes on 8-point scale), and primary source was categorized as Social Media (Facebook, TikTok, Instagram, etc.), Digital Media (online news portals), or Traditional Media (TV, radio, newspapers).

Figure 2 presents the percentages of each of the six narratives viewed on different channels. On the horizontal axis are the six narratives (1 to 6), while the vertical axis is the percentage of respondents who learned about the story from each source.

Figure 2. Information channels of disinformation narratives dissemination



As shown in Figure 2, dissemination channels are extremely varied but social media proves most dominant exposure, with a 59.6 percent encounter with the celebrity arrest (N3) and 57.9 percent with the Germany job scam (N6), both non-political and high-salience topics. Geopolitical narratives (N1, N2, N4, N5), in their turn, feature a more balanced pattern of exposure. Television remains relevant as a source of exposure, especially when it comes to political issues, respondents heard the narratives more on TV (13% to 19.9%) than the celebrity arrest (N3 is 5.4%) and work opportunity (N6 is 2.3%). Respondents were exposed to the four political issue narratives somewhat less on social media and found this information via traditional or digital channels.

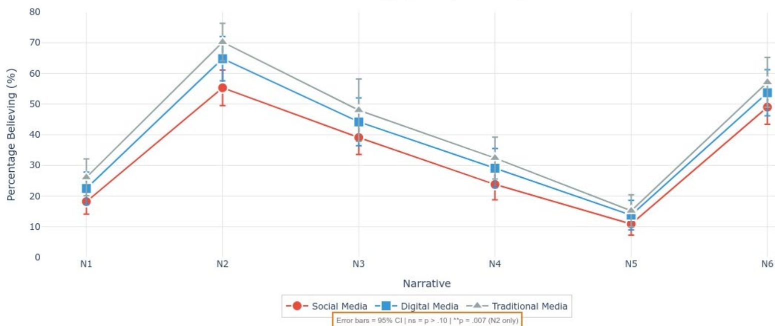
Although social media dominated exposure, the chi-square tests gave no empirical evidence to Hypothesis 2. Table 4 shows belief rates based on source-weighted using the exposure proportions. In five of the six stories (N1, N3-N6), the type of source did not have any relationship with belief (all $p > .34$, Cramer $V < .09$). In the story about the Wagner group (N2), there was a significant correlation but in the opposite direction to the hypothesis, which indicated that social media users were less likely to believe the statement (55.3) than those who used digital (64.8) and traditional media (70.2).

Table 4. Results of chi2 tests, association between primary source of exposure and belief in disinformation narratives (among exposed respondents only)

N	Topic	Social media	Digital media	Traditional media	Association measures	
		Believe *(Exposure %)*	Believe *(Exposure %)*	Believe *(Exposure %)*	Cramer's V	p-value
N1	Ethnic cleansing	18.2% *(47.0%)*	22.5% *(30.6%)*	26.1% *(24.1%)*	.091	.343
N2	Wagner	**55.3% *(45.1%)***	64.8% *(30.6%)*	**70.2% *(21.2%)***	.198	.007
N3	Celebrities	39.1% *(59.6%)*	44.2% *(26.2%)*	48.0% *(9.8%)*	.088	.366
N4	Orthodox churches	23.8% *(40.4%)*	29.1% *(29.2%)*	32.4% *(27.2%)*	.087	.372
N5	Serbs discrimination	10.9% *(43.3%)*	13.8% *(28.8%)*	15.2% *(26.1%)*	.074	.484
N6	Work in Germany	49.0% *(57.9%)*	53.7% *(27.2%)*	57.1% *(18.9%)*	.082	.413

Figure 3 shows that the Wagner narrative (N2) was the only narrative to display a significant difference across media sources ($\chi^2(1) = 8.21$, $p = .004$, Cramér's V = .179). No other narrative exhibited a significant effect of source.

Figure 3. Belief by Primary News Source (Among Exposed)



Taken together, Figure 3 and Table 4 present consistent evidence that exposure to disinformation and belief are subject to different media dynamics. Although the primary exposure channel, especially in the case of non-political, high-engagement stories (N3 and N6), is social media, it does not determine belief. Indeed, the most geopolitically provocative and broadly shared statement (N2) is believed less by those exposed to the narrative on social media (55.3%) than among users who were exposed via traditional media (70.2%). The lower level of belief might be due to corrective diffusion occurring quickly within social media ecosystems due to peer commentary, algorithmic counter-narratives, or exposure to other perspectives, especially among younger and digitally literate individuals. In Kosovo's case, social media spreads disinformation

widely, but does not necessarily persuade—and may even foster greater skepticism toward implausible or externally driven claims. Therefore, Hypothesis 2 was rejected. Although the majority of narratives failed to demonstrate any substantial connection between news source and belief, the narrative about the Wagner group (N2) reported the reverse influence to the predicted finding: it was respondents who mostly used social media that were less likely to believe the statement than traditional or digital media. The finding indicates that in some high-stakes geopolitical stories, social-media consumers can become more skeptical and less vulnerable.

FREQUENCY OF EXPOSURE AND BELIEVABILITY (H3)

Our data shows that the reach of narratives and the frequency of exposure vary widely across the six pieces of disinformation. Some of the narratives were encountered virtually every day or are similar to highly familiar tropes that circulate. However, we could also suggest that familiarity can also relate to the fact that some topics are of higher interest or significance to respondents, as recall can indicate the salience of a story to the individual. Narrative 6, concerning career prospects in Germany, is the narrative most frequently seen or recalled, perhaps indicating a high level of salience among respondents. Other narratives, such as those about the Russian group "Wagner", were found to have moderate recall, indicating the level of interest in this individual story could be lower. Familiarity, therefore, reflects not only information circulation but also issue salience and personal relevance within the Kosovo context.

To test H3, chi-square tests were conducted for each narrative (see Table 5), comparing belief rates (5–8 = believes) between respondents who had previously encountered the claim (exposed) and those who had not (not exposed).

Table 5. Results of chi2 tests, association between prior exposure and belief in disinformation narratives

N	Topic	% Believes *(Exposed)*	% Believes *(Not exposed)*	Association measures	
				Cramer's V	p-value
N1	Ethnic cleansing	21.5%	8.3%	0.281	0.000
N2	Wagner	61.8%	42.4%	0.570	0.000
N3	Celebrities	42.5%	34.4%	0.352	0.000
N4	Orthodox churches	27.7%	11.0%	0.361	0.000
N5	Serbs discrimination	12.8%	6.6%	0.233	0.000
N6	Work in Germany	52.7%	46.2%	0.332	0.000

The findings indicate that there is a statistically significant relationship between previous exposure and belief for all six disinformation narratives. The chi-square tests were very significant (all p <.001), and the effect sizes (Cramer's V) were found to be between .233 and .570. This trend shows that recalled exposure is a good predictor of belief, supporting the mere exposure effect, where repeated exposures positively influence perceived credibility despite the implausibility of claims.

The findings are somewhat mixed, with a higher level of belief when recalling exposure (21.5%-12.8%), despite generally a lower baseline of belief levels for some storylines (Ethnic cleansing 8.3%; Serbs discriminating 6.6%). The impact of exposure is significant in other situations, such as the Wagner group in Kosovo (61.8% vs. 42.4%, V = 0.570), highlighting how security-related narratives may quickly gain credibility once they start to spread. These results show substantial empirical evidence to support Hypothesis 3. The effects of prior exposure are significant in believing disinformation, as cognitive accounts of familiarity and truth bias expect. This is strongest on the geopolitically charged stories (N2, N4), where repetition can make unrealistic assertions but support polarized identities. Even in the case of less political stories (N3, N6), the exposure-belief relationship is high, that is, repetition, no matter the subject matter, increases credibility.

MODELING BELIEF IN DISINFORMATION NARRATIVES (H4)

Table 6 below shows the only significant correlations between the sociodemographic variables and believing each of the disinformation narratives. For the narrative related to the Russian group, "Wagner", males are more likely to believe this narrative (.155**), and also the more educated (.112*). In the case of the celebrity narrative, females find this more believable (.150**), and the more educated (.081*), but the latter is statistically weaker.

Table 6. Correlations of sociodemographic data with believability of each narrative

Narratives	Ethnic cleansing	Wagner	Celebrities	Orthodox churches	Serbs discrimination	Work in Germany
Age	-.152**	.085	-.013	.035	.001	-.069
Gender	.051	.155**	-.150**	.031	.050	.053
Education	-.008	.112*	.081*	-.035	-.052	.010

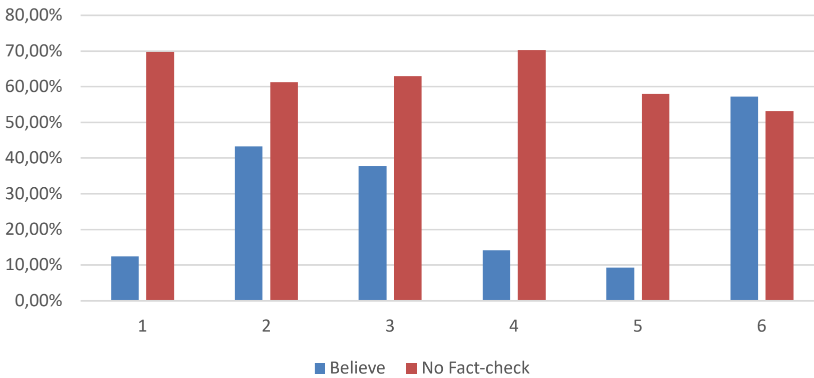
* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

This shows that different forms of disinformation, on different topics, are more or less believable to differing internet users across the Kosovar community. Perhaps they tend to be less likely to believe disinformation about the Serb/

Kosovo tensions unless it is something new and links to a wider context (i.e., the Wagner group), but they can be taken in by celebrity-related disinformation due to its novelty and the globally pervasive interest in celebrity gossip. The premise of inoculation theory is that people will become more resistant to disinformation if they learn narratives are misleading. In the Kosovo context, only a minority of respondents believe the disinformation associated with Kosovo-Serbia relations (N1, N2 & N5); therefore, due to the sheer level of disinformation and its exposure as false by fact-checkers, it seems likely that some Kosovars have built up an innate skepticism of stories that inflame relations with Serbians. If true, then it is positive for this highly contentious political issue if Kosovars have become more resistant to accepting false claims when they see similar stories. For instance, the narrative that the West is supporting Kosovo institutions for the ethnic cleansing of Serbs is believed only by 12.4%, however, exposure can still prove to be significant for the believability of these narratives, suggesting the circulation of disinformation can have an impact on beliefs, and consequently, behavior.

A related problem, which may exacerbate the impact of disinformation, is that Kosovars say they are largely very unlikely to refer to fact-checkers, confirming the findings of previous studies in the region about fact-checkers being ineffective at reaching wider audiences, as often these sources are not visited frequently, if at all. Our data (Figure 4) shows that for most narratives, Kosovars believe the disinformation without referring to any fact-checking systems. This can be explained by confirmation bias as well as selective attention bias. However, while believable and confirming negative attitudes of Kosovars towards Serbia, these narratives may not be seen as consequential or novel and so are unlikely to impact attitudes or behavior. This is relevant when considering how Kosovars view stories about the relations with Serbia. For example, the narrative about the involvement of the Russian paramilitary group, Wagner, was believed by 43.3% of respondents, even though the majority of respondents did not fact-check the story. We surmise that people who have strong pre-existing beliefs that a new war with Serbia is possible, and that Russia will support Serbia over Kosovo, fueled by the Russian invasion of Ukraine, are more inclined to accept narratives that align with their existing view and do not fact-check them. However, greater attention and critical thinking are devoted to narratives that are of immediate consequence. This is quite evident with N6, the narrative related to work opportunities in Germany. A majority of respondents believed it, but some, due perhaps to its relevance, visited a fact-checker to search if it is debunked. Hence, where information may be of benefit to them directly, like N6, they will fact-check it, suggesting that the decision to fact-check might be influenced by the perceived importance of the information and whether it is true.

Figure 4. Percentages of believing and no fact-checking the information



CONCLUSIONS

This study provided insights into the multifaceted nature of information consumption in Kosovo. Through an exploration of six prominent narratives, we show the evidence of how living under a polarized and low-trust society, people learn to live in a complicated information landscape where false information mixes with news.

A key assumption drawn from the data is the substantial potential influence of cognitive biases, including confirmation bias, on people's inclination to accept disinformation narratives, especially if they are familiar to them. Exposure frequency emerged as the most consistent predictor of belief, confirming that repetition normalizes even implausible claims. People are more inclined to accept some material without giving it critical examination or fact-checking when it supports pre-existing views. This is likely a strong factor in Kosovo, where negative narratives relating to Kosovo–Serbia relations have been circulating for decades. However, our analyses also show that belief is not simply a product of bias or trust—but of repeated exposure. This suggests that familiarity itself can make information appear credible, especially when tied to national identity or geopolitical themes.

The tests of Hypothesis 1 did not find a significant relationship between trust and belief when measured on a categorical scale and showed that media trust in itself does not predict disinformation acceptance. This upholds the fact that in societies where there is low trust, the formation of belief may be significantly influenced by existing cognitive/ideological beliefs rather than institutional confidence. Hypothesis 2, which predicted that users of social media were more likely to believe disinformation, is not supported. Although social media was the most used source of exposure, it did not have a positive effect on perception.

The narrative of social media users having even lower belief than traditional media, as seen in one reported case, the Wagner Group. These results indicate that social media increases exposure but is not likely to persuade. The data show strong support for Hypothesis 3, which assumes that repetition of exposure enhances disinformation belief. The most predictable and strong predictor of belief was exposure frequency. The more someone was exposed to a narrative, whether or not it is factual, the more it seemed to be truthful. Acceptance was often boosted by familiarity as opposed to trust or accuracy. This mechanism supports the existing worldviews and normalizes disinformation in the situation in Kosovo, where ethnicity and politics are historically enshrined. Lastly, Hypothesis 4, showing Kosovars are much more likely to confirm information when it fits with their political worldview, is proved by the findings. The level of fact-checking was also very low, and respondents would only confirm information that would be of interest to them, like the migration or employment opportunities. Verification declined drastically when content conformed to their beliefs. This is an expression of motivated reasoning and confirmation bias, in which the process of fact-checking takes on the quality of an identity-based process, instead of becoming an accuracy-based process.

Our data confirms the theoretical viewpoint of much post-truth literature (Lilleker, 2018; McKay and Tenove, 2021) because it shows that disinformation can influence the thinking of the citizens by involving the use of emotions and the mediation of trust instead of rational thinking. In Kosovo, the dynamics are especially noticeable in the polarized, low-trust society, where the repetition of disinformation narratives, particularly in the context of Kosovo-Serbia relations, intensifies the divisions that were already present, and shapes the way citizens perceive political and social realities. The acknowledgment of the relationship between exposure, belief, and trust in such an environment provides insight into how citizens process information in post-conflict democracies. Finally, the research paper is part of a larger debate on the topic of disinformation resilience, which proposes that enhancing media literacy and critical thinking is crucial not only to Kosovo but to the rest of transitional democracies struggling with the same issues.

Comprehending the complexities of disinformation in Kosovo, including its dissemination, sources, and the impact of cognitive biases, facilitates consideration of more efficacious ways to counteract disinformation and foster an educated, media-literate society. This work, by relating the empirical findings to the larger theoretical paradigm, contributes to the research on the origins of belief in disinformation, at the point of exposure, trust, and motivated cognition, and has implications not only for Kosovo but also for other low-trust, transitional democracies. To increase media and digital literacy in Kosovo, it is important to implement a multi-level solution as the country is characterized by a polarized

media space and institutional mistrust. The involvement of media literacy as a mandatory course from elementary through high schools would equip the young generations with the necessary tools to evaluate sources critically and being able to identify manipulation. For the moment, media literacy is taught in some schools across Kosovo as an elective subject (Jahiri et al., 2024). It can also be important to have civil society organizations and media organizing the public awareness campaigns, community workshops, and journalist training programs that are centered on disinformation identification and countering. Besides, the partnership between educational institutions, fact-checking organizations, and media might serve as a good solution to the existing disparity between the practice of verification and the daily intake of information. With such activities backed by uniform policy-backed public communication strategies, this might slowly foster a more critical and resistant citizenry who may handle the intricate Kosovo information space.

While the present research highlights the specific challenges within the Kosovar information environment, a post-conflict society with low institutional trust and high political polarization, the results, especially the influence of media trust, frequency of exposure, and motivated reasoning in the development of disinformation belief, are not exclusive to Kosovo. They can provide important clues into how citizens of other societies, where there are similar low trust levels and polarized media systems, may react to false information. Therefore, the empirical range of the study is context-dependent, whereas theoretical implications apply to the larger discourse of disinformation resilience in weak democracies and transitional information environments.

Whereas our study provides valuable information, certain limitations should be considered. First, the research is focused on a single country context defined by certain historical and political processes, and thus, the findings are to be applied with caution to other societies that possess different media and institutional trust rates. Second, because the study relied on an online survey, younger and more educated respondents were more accessible and likely overrepresented in the sample. This may have influenced the overall belief and verification patterns observed, particularly if these groups display higher digital literacy or political engagement than the general population. Third, the research has examined six narratives, which, though varying in subject matter and ideological inclination, may not capture the full spectrum of disinformation circulating in the Kosovar media environment. The future studies ought to include larger and more diverse collections of narratives focusing on differing national contexts to determine the broader applicability and robustness of the results. These limitations can be addressed in future studies by using longitudinal designs, experimental manipulations, or cross-country comparisons to test more fully the results found here. Regardless of these limitations, the current results provide useful insights

regarding how people determine credibility when exposed to disinformation in a low-trust and post-conflict information environment.

Acknowledgements: This publication is part of the project "DISINFTRUST," funded by the European Media and Information Fund (EMIF). The authors would like to express their gratitude to all cooperation partners who contributed to the development and implementation of the survey that provided the data used in this work.

REFERENCES

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science advances*, *6*(14), eaay3539.

Avdiu, P. (2022). *Public perception on trust, corruption, and integrity of public institutions in Kosovo.* Western Balkans Security Barometer.

Bánkuty-Balogh, L. S. (2021). Novel Technologies and Geopolitical Strategies:

Disinformation Narratives in the Countries of the Visegrád Group. *Politics in Central Europe*, *17*(2), 165–195. https://doi.org/10.2478/pce-2021-0008

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*, 106633. https://doi.org/10.1016/j.chb.2020.106633

Bertolotti, M., & Catellani, P. (2023). Counterfactual thinking as a prebunking strategy to contrast misinformation on COVID-19. *Journal of Experimental Social Psychology*, *104*, 104404. https://doi.org/10.1016/j.jesp.2022.104404

Birks, J. (2019). *Fact-checking journalism and political argumentation: A British perspective*. Springer Nature.

Brautovic, M., & John, R. (2023). Limitations of Fact-Checking on Debunking COVID-19 Misinformation on Facebook: the Case of Faktograf.hr. *Central European Journal of Communication, 16*(1), 40–58. https://doi.org/10.51480/1899-5101.16.1(33).3

Buturoiu, R., Udrea, G., Dumitrache, A. C., & Corbu, N. (2021). Media Exposure to Conspiracy vs. Anti-conspiracy Information. Effects on the Willingness to Accept a COVID-19 Vaccine. *Central European Journal of Communication*, *14*(2), 237–258. https://doi.org/10.51480/1899-5101.14.2(29).3

Chen, E., Chang, H., Rao, A., Lerman, K., Cowan, G., & Ferrara, E. (2021). COVID-19 misinformation and the 2020 U.S. presidential election. *Harvard Kennedy School (HKS) Misinformation Review*.

David, Y. (2023). The effects of exposure to gendered stereotypes on emotions toward immigrants and attitudes toward refugees. *Journal of Ethnic and Migration Studies*, *49*(19), 4828–4849. https://doi.org/10.1080/1369183X.2022.2064840

de Saint Laurent, C., Murphy, G., Hegarty, K., & Greene, C. (2022). Measuring the effects of misinformation exposure and beliefs on behavioural intentions: a COVID-19 vaccination study. *Cognitive Research* 7, 87. https://doi.org/10.1186/s41235-022-00437-y

Diaz Ruiz, C., & Nilsson, T. (2023). Disinformation and echo chambers: how disinformation circulates on social media through identity-driven controversies. *Journal of Public Policy & Marketing*, *42*(1), 18–35. https://doi.org/10.1177/07439156221103852

Gërguri, D. (2022). *Fake News, Information Disorder in the Digital Age.* Dukagjini.

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour, 4*(5), 472–480. https://doi.org/10.1038/s41562-020-0833-x

Hameleers, M. (2024). The state-of-the-art in combating mis-and disinformation: Lessons from pre-and debunking approaches. In Frau-Meigs, D., & Corbu, N. (Eds.) *Disinformation Debunked*, (pp. 19–36). Routledge

Iyengar, S., & Massey, D. S. (2019). Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences*, *116*(16), 7656–7661. https://doi.org/10.1073/pnas.1805868115

Jahiri, M., Qerimi, G., & Gërguri, D. (2024). Exploring critical thinking of smartphone use from elementary school students in Kosovo. *Journal of Media Literacy Education, 16*(3), 107–119. https://doi.org/10.23860/JMLE-2024-16-3-8

Jerit, J., & Zhao, Y. (2020). Political misinformation. *Annual Review of Political Science, 23*(1), 77–94. https://doi.org/10.1146/annurev-polisci-050718-032814

Karlsen, R., & Aalberg, T. (2023). Social media and trust in news: An experimental study of the effect of Facebook on news story credibility. *Digital Journalism*, *11*(1), 144–160. https://doi.org/10.1080/21670811.2021.1945938

Kuś, M. & Barczyszyn-Madziarz, P. (2020). Fact-checking initiatives as promoters of media and information literacy: The case of Poland. *Central European Journal of Communication*, *13*(2), 249–265. https://doi.org/10.19195/1899-5101.13.2(26).6

Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology, 32*(2), 348–384. https://doi.org/10.1080/10463283.2021.1876983

Lilleker, D. (2018). Politics in a post-truth era. *International Journal of Media & Cultural Politics*, *14*(3), 277–282. https://doi.org/10.1386/macp.14.3.277_2

Lilleker, D., Coman, I. A., Gregor, M., & Novelli, E. (2021). Political communication and COVID-19: Governance and rhetoric in global comparative perspective. In D. Lilleker, I. Coman, M. Gregor & E. Novelli, (Eds.). *Political Communication and COVID-19* (pp. 333–350). Routledge.

McKay, S., & Tenove, C. (2021). Disinformation as a threat to deliberative democracy. *Political Research Quarterly*, *74*(3), 703–717. https://doi.org/10.1177/1065912920938

Melchior, C., & Oliveira, M. (2024). A systematic literature review of the motivations to share fake news on social media platforms and how to fight them. *New Media & Society*, *26*(2), 1127–1150. https://doi.org/10.1177/14614448231174224

Nielsen, R. K., & Schrøder, K. C. (2014). The relative importance of social media for accessing, finding, and engaging with news: An eight-country cross-media comparison. *Digital Journalism*, *2*(4), 472–489. https://doi.org/10.1080/21670811.2013.872420

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*(2), 443–458. https://doi.org/10.1007/s10796-022-10242-z

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), 76–87. https://doi.org/10.17645/mac.v11i2.6453

Pierri F, Artoni A, Ceri S (2020). Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections. *PLoS ONE 15*(1), e0227821.

Pierri, F., Perry, B. L., DeVerna, M. R., Yang, K.-C., Flammini, A., Menczer, F., & Bryden, J. (2022). Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. *Scientific Reports, 12*(1), 5966. https://doi.org/10.1038/s41598-022-10070-w

Qerimi, G., & Gërguri, D. (2022). Infodemic and the crisis of distinguishing disinformation from accurate information: Case study on the use of Facebook in Kosovo during COVID-19. Information & Media, *94*, 87–109. https://doi.org/10.15388/Im.2021.94.56

Reddi, M., Kuo, R., & Kreiss, D. (2023). Identity propaganda: Racial narratives and disinformation. *New Media & Society*, *25*(8), 2201–2218. https://doi.org/10.1177/14614448211029293

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., van der Bles, A. M., & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10), 201199. https://doi.org/10.1098/rsos.201199

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences, 116*(16), 7662–7669. https://doi.org/10.1073/pnas.1805871115

STIKK. (2019). Kosovo IT Barometer 2019. December 2019. https://stikk.org/wp-content/uploads/2020/01/IT-Barometer-2019.pdf

Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior, 30*(3), 341–366. https://doi.org/10.1007/s11109-007-9050-9

Su, Y., Borah, P., & Xiao, X. (2022). Understanding the "infodemic": Social media news use, homogeneous online discussion, self-perceived media literacy and misperceptions about COVID-19. *Online Information Review*, *46*(7), 1353–1372. https://doi.org/10.1108/OIR-06-2021-0305

Suarez-Lledo, V., & Alvarez-Galvez, J. (2021). Prevalence of health misinformation on social media: Systematic review. *Journal of Medical Internet Research, 23*(1). https://www.jmir.org/2021/1/e17187/

Suau, J. & Puertas-Graell, D. (2023). "Disinformation narratives in Spain: Reach, impact and spreading patterns". *Profesional de la información*, *32*(5). https://doi.org/10.3145/epi.2023.sep.25

Swart, J. (2023). Tactics of news literacy: How young people access, evaluate, and engage with news on social media. *New Media & Society*, *25*(3), 505–521. https://doi.org/10.1177/14614448211011447

Thorbjørnsrud, K., & Figenschou, T. U. (2022). The alarmed citizen: Fear, mistrust, and alternative media. *Journalism Practice*, *16*(5), 1018–1035. https://doi.org/10.1080/17512786.2020.1825113

Van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine, 28*(3), 460–467. https://doi.org/10.1038/s41591-022-01713-6

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the Public against Misinformation about Climate Change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008

Xu, S., Coman, I. A., Yamamoto, M., & Najera, C. J. (2023). Exposure effects or confirmation bias? Examining reciprocal dynamics of misinformation, misperceptions, and attitudes toward COVID-19 vaccines. *Health Communication*, *38*(10), 2210–2220. https://doi.org/10.1080/10410236.2022.2059802

ORIGINAL ARTICLE

DOI: 10.51480/1899-5101.18.3(41).821

Mobilizing Distrust: Persuasive Strategies in Czech Disinformation Email Subject Headings

Todd Nesbitt

0000-0002-7757-9271
University of New York in Prague, Czech Republic

Tess Slavíčková

0000-0003-2083-4187
University of New York in Prague, Czech Republic

Veronika Zavřelová

0009-0001-2848-1593
University of New York in Prague, Czech Republic

Abstract: This article examines the persuasive function of subject headings in Czech-language disinformation chain emails, an understudied component of cross-platform disinformation in Central Europe. The study focuses on the first six months of the full-scale Russian invasion of Ukraine in 2022 and analyzes a corpus of email subject headings collected by the civic initiative (Czech Elves) Čeští Elfové. Drawing on critical discourse analysis, relevance theory, and presupposition theory, the article explores how subject headings establish relevance, mobilize shared assumptions, and encourage recipients to access and disseminate disinformation content. The analysis identifies recurring thematic frames centered on national victimhood, fear of the "other", distrust of political and media institutions, and the construction of an imagined community of resistance. Linguistic and typographic strategies, including informal address, hyperbolic punctuation, and directive speech acts, presuppose ideological alignment and familiarity with alternative information networks. The findings demonstrate how highly condensed subject headings function as effective ideological triggers within email-based disinformation practices.

Keywords: disinformation; chain emails; presupposition; critical discourse analysis; Czech Republic

INTRODUCTION

Disinformation and its impact on the public sphere has been the subject of numerous studies since the late 2010s predominantly due to new types of influence campaigns brought about by computational propaganda. Following significant news media attention as a result of allegations of foreign interference in the lead-up to the American presidential election in 2016, many nations began to examine the potential for harm in their own domestic environments (Palmer & Wilner, 2024; Ohiln, 2021; Fidler, 2021; Boulos, 2021). Much of the concern has been with the "weaponization" of social media (Kuś & Barczyszyn--Madziarz, 2020).

In the Czech Republic, in addition to the webpages and social media accounts used to attract readers, a chain (disinformation) email environment also became very active in this period, culminating in approximately 1000 different disinformation emails being sent monthly in the winter of 2022 (Pravidelný přehled české dezinformační scény, 2022). The Czech Republic is rather prominent in terms of the vigorous nature and reach of these campaigns. For example, one study showed that 90% of senior citizens had been reached by disinformation emails (Elpida, 2018).

This study focuses on the disinformation chain email phenomenon in the Czech Republic. Specifically, it discusses the persuasive factors involved in accessing disinformation email content. To understand the incentives at work in this engagement, email subject headings were analyzed for persuasive thematic indicators, and functional linguistic patterns. The project examines the first 6 months of the full-scale Russian invasion of Ukraine in 2022.

The study aims to answer the following questions:

- RQ1: What sort of knowledge/prior understandings are the authors of the subject headings expecting readers to have that would motivate them to access/open the email upon reading the subject heading?
- RQ2: What do repeating thematic frames used in subject headings indicate about presupposed views of the audience?
- RQ3: What do evident functional linguistic patterns in the subject headings indicate in terms of presupposed views of potential audiences for the emails?

The article aims to contribute to the research in two main ways: firstly, to emphasize the significance of chain mails in the cross-platform disinformation mix in the Czech Republic; secondly, to break down persuasive factors at work for recipients.

CROSS PLATFORM PROPAGANDA AND DISINFORMATION

Although propaganda has been discussed as part of scholarly work for a century, disinformation became prominent only in the latter part of the second decade of the twenty-first century (Freelon & Wells, 2020). Disinformation is often conceived of as coordinated attempts to spread false information by political players, whether domestic or foreign (Guess & Lyons, 2020). Tucker et al. (2018:22) define disinformation broadly, describing the diverse content it can contain, specifically "trolls, bots, fake-news websites, conspiracy theorists, politicians, highly partisan media outlets, the mainstream media, and foreign governments". The text in this study can be best defined as disinformation, but it seems important to recognize the role of state propaganda as a significant stand-alone element.

The Czech Republic is an interesting area for the study of propaganda, due to its strategic position on the historical border of the former Soviet Bloc, and therefore a volatile region in light of the often-discussed suggestions that the Russian Federation seeks to return its sphere of influence to Soviet-era times (Studzinska, 2015). The country is also interesting to examine in terms of the medium and platforms used for propaganda, since the demographic make-up of the country means that the less technically savvy have an overwhelming influence in elections, those aged 65+ making up 20% of citizens ("Age Structure", n.d.). Additionally, the growth of malaise with democracy and increased belief in disinformation as described by Bennett and Livingston (2018) clearly apply in the Czech Republic, with only 34% placing trust in news media (Mihelj & Štětka, 2022), and numbers trusting disinformation reaching dramatic levels at 40% (STEM, 2021).

The country for many constitutes fertile ground for disinformation and propaganda not only because of its geopolitical positioning, but also due to a combination of historical, cultural, and institutional factors. Long-standing skepticism toward official authority, shaped by experiences of state socialism and reinforced by post-1989 political and economic disillusionment, sharpened by instances of capture of the media by oligarchs may contribute to low levels of trust in political institutions and mainstream media. Research shows that this environment has facilitated the growth of alternative interpretive frameworks that present politics through narratives of elite betrayal, national victimhood, and hidden power structures (Štětka et al., 2021; Mihelj et al., 2022).

In light of this context, we consider the medium used for accessing news. In addition to traditional channels and social media platforms, news by email is also a common method of distributing information by a broad range of news organizations. In some ways resembling a curation process related to theme/topic, the success of the model seems to be related to the connection made between the reader and the news organization, or individual journalists themselves (Seely

& Spillman, 2021). Finding it to be a financial model that works has provided encouragement (Jack, 2016). Significant numbers of people claim to receive news by email, ranging from one in ten to one quarter of the population in a recent study (Newman, 2022).

Email communication is also common in political campaigning, with manipulative and deceptive tactics being the norm (Mathur et al., 2023). Further, chain emails used for disinformation purposes, a natural cousin to news by email, as is fake news to "real" news, are a surprisingly understudied part of the cross-platform propaganda approach, particularly since news email "audiences" can be found that are not necessarily social media users, and for some of the population, for example those of retirement age, it is a gateway to internet communication (Madden & Zickuhr, 2012). Chain disinformation emails often use a deflective source model (Jowett and O'Donnell, 2019), attracting the attention of the receiver with a personalized message in the subject heading, but sharing information from a third source – not the email author. Commonly referred to as astroturfing (Howard, 2002; Chan, 2022), the overall aim is to allow for messages to be spread farther.

Chain emails written in the Czech language contain socio-political content intended to provoke a reaction in receivers, including encouragement to share (forward) the email to others. The source of the emails being analyzed in this study is very often the object of speculation. Some observers feel that a major source of the emails is a hostile government (often the Russian Federation) paying local citizens to spread content in order to bring about unrest. Others feel it is local (Czech) citizens with sympathies toward Russia, while others have alluded to mainstream domestic political parties being involved. Thus, elements of state propaganda and non-state localized / non-state organized disinformation form the backdrop of the manipulative use of email for persuasion.

THEORETICAL FRAMEWORK

Critical Discourse Analysis (CDA) is typically characterized as a unit of language which is greater in length than one sentence, but whose meaning extends beyond the unit itself. This conceptualization enables analysis of text, co-text and wider context, including the sociocultural realities underpinning discourses.

Van Dijk (2005) has argued for the articulation of CDA to cognitive analysis and pragmatics and the negotiation of meaning. Any discourse-based analysis of news reporting (including news by email) entails exploration of the relationship between writer/speaker and reader/listener, engagement with the pragmatic dynamics underlying this relationship, and uncovering the ideological messages concealed beneath the immediate text. We need to understand the

assumptions that message authors make about their readers in terms of their existing knowledge and likely ideological attitudes. Bekalu (2006), advocating a greater engagement by analysts with the cognitive dimension of discourse processing, provides a useful framework following Van Dijk's (2005) lead, linking presupposition with cognition.

One key linking approach is via relevance theory (see Sperber & Wilson, 1995 [1986]), which accounts for the ways in which writers (including subject heading authors) secure audience attention. The value of an utterance is determined if: "(1) it is sufficiently relevant to be worth the audience's processing effort, and (2) it is the most relevant one compatible with the communicator's abilities and preferences" (Bekalu, 2006, p. 155). Using presupposition to establish relevance is thus one important component in uncovering the ideological content of a media message. Althusser's (1971) concept of interpellation proposes that everyday cultural knowledge is ideological, and that communicators construct individuals as subjects with shared assumptions regarding power and authority. For example, if a news source describes a public figure as "hardline", it assumes that its readers/viewers have been positioned to share, and judge, this "knowledge" of what constitutes "hardline" behavior. Van Dijk (2005) proposes deployment of a "context model" that posits knowledge as "shared beliefs satisfying the specific (epistemic) criteria of an (epistemic) community" (Van Dijk, 2005, p. 73). This approach recognizes the impossibility of absolute truth in epistemology, adding a flexible dynamic to the cognitive schemata we draw upon to function in the everyday social world. Context accepts the importance of pragmatic negotiation of knowledge, "because many aspects of discourse depend on what the speaker assumes the hearer to know or not to know" (Van Dijk, 2005, p. 76). Van Dijk goes on to assert that ideological presuppositions are easily formed or reinforced by the media:

> Since people learn by the acquisition of new knowledge through public discourse, they may thus be manipulated into believing that such presumptions of authoritative sources are in fact forms of knowledge that no longer need to be certified (demonstrated, proved, etc.) (Van Dijk, 2005, p. 88).

Of course, the process of reading an email heading must be cognitively rather different to reading an entire email or a mainstream news report. With such minimal verbal content and context, the amount of processing needed is highly dependent on the familiarity of readers with chain emails and an expected ideological alignment with their content. Although we might accept that at least the previous link in the chain is a person known to the receiver, the subject heading plays an important functional role in inducing a reader to open it. Until the email is opened, they are largely unaware of its content. As is the case with

conventional reports, news headlines (and by extension email subject headings) establish relevance via the promise of new information pertaining to a (probably) familiar subject. The role of the headline creator is to pique readers' curiosity, that is to make a topic "relevant" while reassuring them that their cognitive and ideological presuppositions will be satisfied by the content.

METHODOLOGY

Our goal was to understand the persuasive factors at play in the subject headings in chain disinformation emails in the Czech language distributed by Czech servers via personal email accounts. The aim was to try and discover linguistic and even typographic patterns that motivate people to open seemingly unknown emails, although they may have been forwarded by persons known to the recipient. In this study, the focus of attention was limited to email subject headings as discrete items, generally shorter in length than a sentence. Some potential data regarding context and wider meaning that could be found by reading the emails in full, or searching for their provenance, is absent. The research team felt it was important to approach each chain email as a typical reader would find it among their everyday personal emails.

The researchers gathered data for the period starting from the Russian invasion of Ukraine, thus from February 24, 2022, to August 24, 2022. The data was provided via a database compiled by the Czech Elves, a volunteer civic movement that, inspired by similar 'Elf' movements from the Baltic countries, maps, analyzes and actively fights against foreign disinformation campaigns on the Czech internet. The Elves keep a protected database of chain emails and provide access for researchers. The database, called Eldariel, catalogues disinformation chain emails circulating within Czech cyberspace and is available in both public and restricted formats. The public version offers anonymous access and presents messages in a machine-shortened form to prevent their further dissemination. A complete, unedited version is reserved for journalists, academic researchers, and comparable professionals. Access to this full dataset must be formally requested and justified, and applicants' identities are verified prior to approval. In line with state security requirements, the Czech Elves do not release copies of the database to third parties, prohibit automated remote processing, and do not allow automated downloading or replication of its contents (Čeští Elfové, 2026).

The database allows for filtering, and thus researchers filtered the emails by 'Ukraine' and 'refugees' in the database in the time period described above, leading to thematic and functional analysis of presuppositions about recipients' views.

The gathered data enabled identification and description of persuasive thematic and functional indicators, revealing common features which may be influential

in persuading readers to open the email. The researchers then refined the data further to identify and discuss key presuppositions about email readers' views and expectations inherent in the subject headings.

DATA ANALYSIS

THEMES

Thematic analysis confers familiarization with patterns of discursive strategies. In this case, we can assume that email authors and disseminators make choices, based on what they know or surmise about the typical profile of the target reader of disinformation emails, in order to direct their attention in the most favorable manner. We may also choose to go beyond the thematic profile of the subject headings to consider their speech act functions.

All subject headings filtered for *Ukrajina/Ukraine* and *úprchlíci/refugees* from February-August 2022 were counted and categorized thematically. Given that the focus was on headings only, the filtered corpus was small (3,882 words), thus, thematic and functional patterns could be discerned with relative ease by eyeball scrutiny. The filter used also flags up articles about Ukraine and refugees when those keywords do not occur in the subject headings. However, in most cases, we can assume that the regular reader has already inferred this, either from linguistic and other signs in the headings, or based on knowledge about the email's source (e.g., a friend or acquaintance). The categorizations were eventually narrowed down to four overarching themes (plus a fifth, "other" category which is predominantly functional rather than thematic).

OVERVIEW OF COMMON THEMES

Anecdotal evidence and hearsay are frequently deployed, drawing on testimony of eyewitnesses who are often unnamed (a friend, a witness, etc.) *Svědectví řidiče z Brna při převozu tzv. uprchlíků z Ukrajiny/Testimony of a driver from Brno relating to transport of so-called refugees from Ukraine.* At the same time, the theme of negative change appeared via a search for words and phrases such as invaze/invasion, *je zlá doba/it's a bad time/Je zlá doba, bude hůř?/Times are bad, will it be worse?*

Individuals and institutions commonly named in subject headings were subdivided into "heroes" and "villains". Regular disinformation email readers are likely to be triggered by known names in subject headings. "Heroes" are typically bloggers, far right influencers, lawyers and journalists whose names may be familiar in news reports or have become familiar via social media to

regular disinformation readers. Czech government ministers and officials, as well as mainstream media, are typically portrayed in non-mainstream news media as “villains”: *FIALO, zodpovídej se národu, lidu této země, všech 499 mých kontaktů souhlasí s OKAMUROU, a ti to rozeslali dalším celkem 10,117 lidí a ani jeden není při vás. / [PM] FIALA, take responsibility for the nation, for the people of this country, all 499 of my contacts agree with [far right politician] OKAMURA, and they have forwarded to a further 10,117 people and not one of them is with you.*

The theme of “undeserving refugees” comprises two main sub-categories: a) the common populist trope of “[national majority] first”, b) Czechs as victims, or the “we’re the fools” subcategory. Authors/senders attempt to reinforce the validity of their case often by making parallels with other countries, especially neighboring Germany, or by poking fun at the perceived gullibility of Czech citizens. *Ukrajina z vděčnosti k České republice a její pomoci zdražila poplatky za tranzit ropy… In gratitude to the Czech Republic and for her help Ukraine has increased tariffs for transport of oil………Nejsme vaši sluhové!!! We’re not your servants!!!*

The “other” category denotes subject headings of zero or minimal semantic content (e.g., punctuation markers such as exclamation marks, emojis, single words such as *video*, *info*, or *zpráva/news*, etc). Some headings with minimal semantic content might be categorized as “exclusives”; a common clickbait trope here indicated by phrases such as *Málokdo ví…/Not many people know…, Praha/Prague, or Sdílejte/Share*. These were categorized as providing little or nothing in the heading that would indicate the email’s topic. Punctuation such as multiple exclamation points, or the occurrence, for example, of single trigger words such as *záhada/mystery* might also act as clickbait without providing any indication of content.

FUNCTIONS

Functional analysis focuses on interplay between the following: presuppositions, implicature, speaker/reader-meaning, content (semantic content), grammar (grammatical mood). We distinguish between a meaningful (a referential) utterance, or locutionary act, in Austin’s terms, and its illocutionary force (its intended or perceived meaning). Searle (1969) suggests that some locutionary acts also behave as illocutionary acts. Certain locutionary meaning determines their illocutionary force (i.e., what is intended by the utterance). This necessitates a contractual alignment of meaning between writer and reader. Readers may interpret this subject heading (“po přečtení budete plakat”/*after reading this you will weep*) as a prediction, a threat, or even a hidden command.

Similarly, a heading referencing anecdotal testimony as in “kamarádka ze Vsetína viděla”/*a friend from Vsetín [a town] saw* comprises a grammatically

coherent (if incomplete, without an object) utterance, comprised of meaningful words, however, its pragmatic meaning, function and significance may be understood by different readers in different ways.

As would be expected from disinformation, the primary function of many subject headings is to warn, or remind, readers that mainstream media is failing to keep the population informed about "the truth", and that the situation regarding Ukraine policy is worse than it appears (as in most examples below). The use of colloquialisms, non-standard grammar and punctuation may be deliberately deployed as a contrast to the perceived elitism of mainstream journalism. This is prominent in the *we're the fools* narrative in which Czechs self-represent as victims. Some headings take the form of directives (forward, resend, or watch). Underlying the warning/alerting function is the promise that readers will have access to the "real" version of events, and that they will thereby continue to seek out "alternative" messages, including disinformation.

PRESUPPOSITION

In this section subject headings are considered in relation to presuppositions about readers. Typographical anomalies in the table are retained.

Table 1 (below), the largest category, deals with presuppositions that underlie nomination strategies used in the headings. Some individuals are named and assumed to be known to the reader as public figures, for example, Vaclav Klaus (former prime minister), Jaroslav Foldyna (politician of the far-right Freedom and Direct Democracy party), Robert Vašíček (local Prague politician), and Petr Štěpánek (actor). These politicians are presupposed to be aligned with the populist opinions of the sender, and target readers are those who have expressed opposition to or skepticism about the Czech government and EU policy regarding Ukrainian refugees. Others in the list are deemed to be unknown; these include "normal-thinking citizens", "elderly Czech lady", "Slovak Serbian war veteran", "Ukrainian Liudmila", "Petr", "Jan", "Pavel" and "lawyer Mgr. Jana Zwyrtek Hamplová" (with emphasis on her academic title and profession as a reliable witness). To the reader of these emails, even less well-known opinion leaders may be recognized from previous email content elsewhere.

The current Czech PM (Fiala) is named twice, but as the object rather than subject of an action (by an "elderly Czech lady" and (former president) Vaclav Klaus).

Table 1. "Reader will approve of or respect these opinion leaders or witnesses"

Poslechněte si to – advokátka Mgr. Hamplová	Listen to this – from the lawyer Mrs. Mgr. Hamplová, M.A.
Česká stařenka…dotaz na Fialu	Old Czech lady – a question to Fiala
Slovenský veterán ze srbské války – to je bohužel realita !!!	Slovak veteran from the Serbian war – this is unfortunately the reality
Ukrajinka Ludmila k současné situaci … to si přečtěte všichni ,TO JE DÍLO	Ukrainian Ludmila to the current situation – everyone read this – IT IS SOMETHING
Petr posílá video z Vaclaváku	Petr sends a video from Vaclavák (the colloquial term for Wenceslas Square)
Názor z druhé strany – od normálně myslících občanů..	An opinion from the other side – from normal thinking people
Krátce hovoří Bc. Robert Vašíček – zástupce Prahy 11 – video – a je moc pravdivé …	A brief talk from Robert Vašíček, B.A., a government representative of Prague 11 – video – it is very truthful
Foldyna k Ukrajině, kde bere tu odvahu? Velmi zajímavé..	Foldyna to Ukraine – where does he get the courage? Very interesting…
MACEK – STAREJME SE O NAŠE LIDI ! – k vyjádření k rozpravě p. Moravce skutečnost	MACEK – LET´S TAKE CARE OF OUR OWN PEOPLE! – in response to the Mr. Moravec conversation – the reality
Klaus napsal Fialovi a jeho vládě – přepošli dál, ať to ví co nejvíce ˇČechů	Klaus wrote to Fiala and his government – send it on, so as many Czechs as possible know about this
Jan Vymazal na TikTok ,toto většina národa neví ,jaká	Jan Vymazal on TikTok, most of the nation don't know ,how
Petr Štěpánek. Jeden z mála, kterému zůstal rozum v „hlavě"!	Petr Štěpánek. One of the few that still has common sense!
Dopis od Pavla – pracoval dlouho v Rusku a ….. přehledně a výstižně!	A letter from Pavel – he worked in Russia for a long time – clearly and precisely!
TOHLE SE NA MAINSTREAMU NEDOČTETE !!!…výborný článek od Pavla ŠTÁSTKA..,,,,..	YOU WON'T FIND THIS IN THE MAINSTREAM!!! – excellent article by Pavel Šťástek
Vladimír Vladimírovič Putin děkuje Zelenskému, že dokázal skoro nemožné!!!	Vladimir Vladimirovic Putin thanks Zelensky that he achieved almost the impossible!!

Source: work of the authors

Table 2. Disrespect for various authorities (predominantly the current government) is formally expressed by the second person singular (informal you), and the absence of title in connection with names (*voc.* Fialo), which is rather pejorative. The vocabulary contains derogative terms and slurs (gypsies), insults connected to the cognitive and general mental capacities of the current government (stupid, stupidity). The form also creates a sense of urgency through the usage of capital letters and the absence of proper punctuation (one may only speculate if it is due to the lack of accurate knowledge). The headlines suggest that no matter what the government says and does, it is and always will be incorrect, inappropriate, or fraudulent (a trip by government representatives to Kiev was "fraud"). There is an acute division between 'us' (the senders of the emails, normal, right people) versus 'them' (wrong, inaccurate, unfair government).

Table 2. "Reader will disapprove of or disrespect these opinion leaders or witnesses"

DOPIS Rakušanovi od zdr. sestry – Pan Rakušan nás nazval šváby!	A LETTER to Rakušan from a nurse – Mr. Rakušan called us cockroaches!
Rakušan se omlouvá za příkoří cikánům.	Rakušan apologizes for cruelty to gypsies
Nouzový stav – vláda může nařídit ubytovat přidělený počet tzv. uprchlíků u Vás doma.a nebo zabavit nemovitost.	State of emergency – the government can order the accommodation of an allocated number of so-called refugees at your home or confiscate real estate.
Hloupost našich vládních představitelů – Agresivita místních „hrdinů" se zvyšuje…	The stupidity of our government representatives – The aggressiveness of local „heroes" is increasing…
Cesta premiéra Fialy do Kyjeva je podvod	Prime Minister Fiala's trip to Kyiv is a scam
FIALO TO SI PŘEČTI A SPOLUTVOŘITELÉ KONFLIKTU TÉŽ Otevřený názor M.Knížáka.*	FIALO, READ IT AND CO-CREATORS OF THE CONFLICT TOO Open opinion of M. Knížák
Opakovaná výzva starostky obce premiérovi Fialovi	Repeated challenge of a village mayor to Prime Minister Fiala
FIALO, zodpovídej se národu, lidu této země, všech 499 mých kontaktů souhlasí s OKAMUROU, a ti to rozeslali dalším celkem 10,117 ldií a ani jeden není při vás. Další volby už neobsadíte	FIALA, answer the nation, the people of this country, all 499 of my contacts agree with OKAMURA, and they sent it to another total of 10,117 people and not one is with you. You will never win another election

Source: work of the authors

Table 3. A common trope of populism is the narrative of victimhood, as proposed, for example, by al-Ghazzi (2021), who describes the ways in which perceptions of historical injustices are woven into contemporary national political discourses (e.g., Taking Back Control in Brexit-supporting Britain). Victimhood is commonly expressed in small-nation populist discourse as the desperate struggle to maintain a discrete, independent cultural identity, and wrest political control away from elites and "back" to "the people". In the Czech Republic, this issue is heightened by the migration of large numbers of Ukrainian refugees, perceived government ineptitude, and pressure from the European Union.

Victimhood is deployed in some way in most of the presuppositions here, although it can be seen as a reactive response rather than one tied to any proactive, future-oriented movement, and which is tinged with self-criticism. Self-criticism manifests as skepticism regarding the Czechs' willingness to fight for their endangered future, as disinformation readers see it. As examples below presuppose, the government of the Czech Republic is a laughing stock in the eyes of other countries, too weak to stand up to the wave of refugees set to target Czech citizens, cast them into the streets, deny them access to their own national healthcare system and even prevent their children from using their local playgrounds.

Table 3. ”Reader will presuppose Czechs are suffering”

Česká země je … JEDEN BY BREČEL,TU NEJDE O VTIPY…	The Czech lands are … BRINGS ONE TO TEARS, THIS IS NOT FUNNY…
Pět minut kruté pravdy – Češi jsou pro Evropu pouhá pakáž !!!	Five minutes of cruel truth – Czechs are the trash of Europe!!!
Ukrajinské děti vyhánějí české dítě ze hřiště, musel zasáhnout otec…Sláva ukrajině…a to je teprve začátek	Ukrainian children kicked a Czech child out of the playground, the father had to intervene…Glory to Ukraine…and this is only the beginning
VIDEO: Češi končí na ulici, sdílej to!	VIDEO: Czechs will end up on the street, share it!
Ukrajinští „utečenci" v Česku…to si poslechni , opravdu by stím už měli něco dělat , aby nás přestali využívat!!	Ukrainian „refugees" in the Czech Republic… listen to this, they really should finally do something so they stop using us!!
Zneužívání zdravotní péče	Health care abuse
Nejsme vaši sluhové!!!	We are not your servants!!!
Ekonomické a zdravotní důsledky války na Ukrajině pro Česko a Slovensko! Analýza, co nás asi čeká.	Economic and health consequences of the war in Ukraine for the Czech Republic and Slovakia! An analysis of what awaits us.
ČR jako terč !	ČR as a target!
Pro blbou vládu!!! Celý svět se vám směje!!! Ve světě si neškrtnete a když tak jen se smíchem všem, kdož vás využijí a vy to při své temnotě neuvidíte For the stupid government!!!	The whole world is laughing at you!!! You can't cut yourself off from the world, if so only by laughing at everyone who uses you and you won't see it in your darkness

Source: work of the authors

Table 4 shows the premise that refugees pretend that they are in a difficult situation to take advantage of the Czech social system and misuse it into their advantage (buying expensive cars, going for expensive vacations). They pretend to be injured, poor, etc. in order to demand systematic help. Czechs who offer any help are considered stupid. Institutional help is mocked. Meanwhile, the Czech government does not help the Czech people. There is the implication of the Czech self-stereotype using envy, and not wishing well to those in need, foregrounding the "Czechia first" trope.

Table 4. "Reader thinks refugees take advantage of Czechs' generosity"

Superdovolené Ukrajinců…	Ukrainians' superholiday
Rozbor dávek pro ukrajinské běžence, nebo válečné uprchlíky nebo sociální turisty ???	Benefit trends for Ukrainian fugitives, or war refugees or welfare tourists ???
Cizinec není našinec	A foreigner isn't one of us
Šok: Romové z Ukrajiny se přeplněnými vlaky valí do Prahy„Pořád se ptají, kdy dostanou nějaká eura… když nastupuji s jízdným zdarma do pendolína z Košic do Prahy na nikoho s místenkami neber	Shock: Roma [ethnic minority] from Ukraine are filling trains and heading to Prague. They keep asking when they will get some Euros…when they get on the Pendolino [trains] from Kosice to Prague for free
Už se lidé začínají ptát proč jsou zde…???	People are already starting to ask why they are here…???

Ukrajinští „utečenci" v Česku…to si poslechni , opravdu by stím už měli něco dělat , aby nás přestali využívat!!	Ukrainian "refugees" in the Czech Republic… listen. They really should do something so they stop taking advantage of us!!
Ukrajinci se vrací domů v nově nakoupených autech	Ukrainians are returning home in newly-bought cars
Utíkají i ti, co to nepotřebují – jak to s námi dopadne?! :((	Even those that don't need to are fleeing – how will that work out for us?! :((
Šok: Romové z Ukrajiny se přeplněnými vlaky valí do Prahy	Shock: Roma [ethnic minority] from Ukraine are filling trains and heading to Prague

Source: work of the authors

Table 5. Arguably one of the most popular, oldest strategies in persuasion is trying to convince the audience that the true reality is not the one that they have been accustomed to accepting. Although the approach may not be novel, the end of scarcity in information and communication technologies and the explosion of content creation tools mean it has become increasingly powerful. Presupposing that the audience is skeptical and willing to accept alternative narratives is thus an obvious avenue for exploration. The sample provides a number of trends, ranging from metaphors which possibly provide a sense of layperson authenticity ('open [your] eyes', 'the Czech media won't show you this!'), to 'truth' synonyms and antonyms ('reality!!!', 'truth', or 'lies') the use of punctuation to emphasize the seriousness of the 'disinformation' (multiple exclamation marks, question marks), reference to 'reputable witnesses' ('Slovak veteran of the Serbian war', 'Slovak lady from Kiev) and key nation states which would have the motivation, in the context of the invasion, to not show the 'truth' (Ukraine, USA, and Germany).

Table 5. "Reader believes things are not what they seem"

Posílám video co otevře oči	I'm sending a video which will open your eyes
Slovenský veterán ze srbské války – to je bohužel realita !!!	Slovak veteran of the Serbian war – this is unfortunately reality !!!
pokud to je pravda…😏	If this is true…
Zničená nemocnice v Mariupolu: Kdo tady vytváří dezinformace?	Destroyed hospital in Mariupol: Who here is creating disinformation?
V Polsku o migranty policie nestojí a lid ano????	In Poland the police don't take care of migrants and the people do????
Cesta premiéra Fialy do Kyjeva je podvod	[Czech] Prime Minister's trip to Kiev is a fraud
Pravda o majdanu na Ukrajině_sdílejte!	Truth about the Ukraine maidan _ share!
Schůzka Zelenského summitu v Kyjevě byla ffalešná – schůzka se konala v polském Przemyslu	Zelensky summit meeting in Kiev was fake – the meeting took place in Przemysl in Poland
Jak vznikla Ukrajina – je dobré vědět něco z historie…	How Ukraine was formed – it's good to know some history
Slovenka z Kyjeva poslala video – toto vám česká media neukáží !!!!	Slovak from Kiev sends video – Czech media won't show you this!!!

Source: work of the authors

DISCUSSION

In this section we return to the research questions posed at the outset to assess findings from the data analysis.

Van Dijk (2005: 95-96) describes what he calls the K-device, "the knowledge of language users about the knowledge of the recipient", which determines the nature and scope of discursive strategies used between interlocutors, and affecting how new knowledge is mapped. This includes assumptions regarding types of knowledge ("personal, interpersonal, group, institutional, national or cultural"). He continues:

> Although the various strategies allow speakers [authors] to make more specific assumptions about what recipients know, the overall (meta) strategies are surprisingly simple, such as: When the recipients are members of my community, assume that they know all that I know, except the information about personal experiences or sources not (yet) used by the recipients. This will account for everyday storytelling and news in the press. (Van Dijk, 2005 p. 96).

We might therefore surmise that disinformation email readers, impacted by unfolding news about war in Ukraine and the arrival of refugees in the Czech Republic from early 2022, will probably need to process the new situation and undertake cognitive adjustments to their existing or common ground knowledge framework.

News reporting comprises what is taken for granted (presupposition) and what is (expected to be) new information. However, within what is a narrowcast medium, email headings are even more tightly anchored to dominant readings (author-preferred) than would be the case with news headlines. The element of presupposition in email subject headings is likely to be even less nuanced, that is there is an absence of discursive hedges or modality to express degrees of truth and certainty. This may be partly for stylistic economy (normatively, most email headings are no more than a few words), but it can also be argued that the lack of nuance reflects confident presuppositions by email senders about readers' ideology.

The emergence of repeated themes raised the question about presuppositions by heading authors regarding the audience in their knowledge of societal affairs, and what would then motivate them to open the emails. Three main areas seem to emerge: 1) traditional fear of the 'other', 2) the position that traditional institutions of authority cannot be trusted, 3) the view that a small community of resistance exists that shares their views.

Fear of the 'other' has existed as long as humanity itself and is a key sociological concept. More narrowly, it is central in nationalism and propaganda studies,

in a mass media context possibly most famously addressed by Anderson in the formation of his concept of imagined communities (1983). Modern, mediated propaganda campaigns and extremist movements typically use this approach to gather support against an imagined external threat, presented as the source of a problem (Berger, 2018). Thus, we find this fear of the 'other' or 'external threats' as a repeating frame in disinformation email subject headings. Specifically, in presupposition tables 3 ("Czechs are victims") and 4 ("Refugees are undeserving") we see the fear of the other repeating – from Ukrainian children kicking a Czech child out of the playground, to the suggestive, "*Czechs will end up on the street!*" to "*Ukrainian "refugees" in the Czech Republic… listen to this, they really should finally do something so they stop using us!!*", the existential threat from the refugees is made clear. "*Benefits*", "*superholidays*" going home in "*newly bought cars*", the message is clear: "*we*" are being "*taken advantage of*" by them. From children to "*Ukrainian fugitives*" to "*welfare tourists*", the persuasive nature of such selected language entices the reader to open the email.

Following the emergence of computational propaganda, scholars in recent years have made efforts to connect the decline of trust in institutions, government and the media to the growth and intensification of disinformation campaigns (Bennet, 2018; Mounk, 2018; Lazer et al., 2018; Walter et al., 2020). In the headline frames studied, this skepticism towards traditional institutions of authority emerges as one of the most common frames. In table 1 ("Will approve of or respect these opinion leaders or witnesses") we see repeated allusions to figures that fit: from the average Joe sending a "real" video from a demonstration ("*Petr from Vaclavák*"), to the "*old lady*" grilling the Prime Minister, to the "*admirably courageous*" (far-right) politician, the frame is clear. In table 2 "Will disapprove of or disrespect these opinion leaders or witnesses"), we see, for example, derogatory inferences toward the government ("*The stupidity of our government representatives*"), members of parliament ("*Prime Minister Fiala's trip to Kyiv is a scam*", "*A LETTER to Rakušan from a nurse – Mr. Rakušan called us cockroaches!*"). In table 5 ("Things are not what they seem") we find the emphasis on the "*lying*" media, with plenty of suggestions reflecting this ("*summit meeting in Kiev was fake*", "*Czech media won't show you this!*", "*truth about Maidan*", "*…a video which will open your eyes*". All these examples clearly show evidence of the presupposition that potential readers feel they are being lied to by democratic institutions, pawns in greater, orchestrated conspiracies, and that political parties and the mainstream media are colluding in concealing the truth.

Finally, presuppositions imply that potential email readers feel that a small community of resistance exists that is fighting for them, and that they ostensibly may be a part of it. Both Blee (2016) and Ebner (2023) discuss the significance of identification and fusion of personal and group identities in both social movements and extremist activities, including online radicalization. In our categories,

this was also a relevant context, logically in table 1 ("Will approve of or respect these opinion leaders or witnesses"), and table 3 ("Czechs are the victims").

A final question we sought to answer was on the functional side, seeking to find approaches or trends in language in the subject headings which may be persuasive in encouraging the audience to open the emails, that is what role functional linguistic patterns play in presupposition. What is assumed about the audience in terms of the use of language?

Other common textual features that may affect function include the use of hyperbolic punctuation (e.g., overuse of question marks and exclamation points); non-standard morphology (e.g., "mladej"/young, not standard mladý); frequent interrogatives; non-standard spelling and typography (including emojis and overuse of capital letters, general overuse of conversational idiolect.

CONCLUSIONS

This study highlights the central role of presupposition, audience positioning, and linguistic form in the construction of disinformation email subject headings in a topic-specific and time-limited analytical frame. As such, it provides a limited account of disinformation activity in the Ukraine war. Undoubtedly, disinformation changes over time, and the passage of time will also have impacted its flow and content, requiring further investigation. The findings demonstrate that these headings are not neutral previews but strategic entry points that activate recipients' existing knowledge, fears, and ideological alignments in order to prompt engagement. Authors presuppose shared attitudes among their audiences, most notably fear of the "other," deep distrust of political and media institutions, and identification with a small, oppositional community, and deploy thematic frames that exploit these assumptions to elicit affective responses and encourage email opening.

At the same time, recurring functional linguistic patterns, including hyperbolic punctuation, informal and conversational language, capitalization, and typographic disorder, serve to create urgency, simulate interpersonal familiarity, and lower critical distance between the sender and recipient. These stylistic choices reinforce the persuasive force of the message and support its ultimate purpose: the rapid and wide dissemination of disinformation through forwarding and sharing practices. Taken together, the analysis shows that disinformation email subject headings reflect a sophisticated understanding of audience cognition and pragmatic manipulation, positioning them as a key yet often overlooked mechanism in contemporary disinformation ecosystems.

The findings may have broader implications. The effectiveness of highly condensed, presupposition-heavy subject headings in mobilizing distrust

suggests a structural vulnerability for democratic public spheres: when political communication increasingly operates through minimal cues that activate shared grievances and in-group identities, deliberative engagement may be displaced by affective alignment and rapid dissemination. This has clear implications for media education and literacy interventions, which may need to move beyond fact-checking or source evaluation toward a deeper focus on recognizing presuppositional cues, emotional triggers, and pragmatic manipulation in everyday digital communication formats as simple and unassuming as subject field lines in emails.

Going forward, these dynamics are likely to be further exacerbated by AI-assisted content creation and dissemination. Generative AI tools can already automate the production of highly tailored, emotionally resonant micro-texts at scale, potentially accelerating the circulation of disinformation that is finely tuned to specific audiences' fears, identities, and cognitive shortcuts. While a systematic analysis of AI-mediated disinformation is beyond the scope of this study, the patterns identified here suggest that future research and policy interventions will need to account for how automation may intensify existing persuasive strategies.

REFERENCES

Althusser, L. (1971). *Ideology and ideological state apparatuses (Notes towards an investigation).* In *Lenin and philosophy and other essays* (B. Brewster, Trans., pp. 1–60). Monthly Review Press.

Anderson, B. (2016). *Imagined communities.* Verso.

Austin, J. L. (1962/2018). *How to do things with words* (J. O. Urmson, Ed.). Martino Fine Books.

Bekalu, M. A. (2006). Presupposition in news discourse. *Discourse & Society, 17*(2), 147–172. https://doi.org/10.1177/0957926506060248

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication, 33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Berger, J. M. (2018). *Extremism.* MIT Press.

Boler, M., & Nemorin, S. (2013). Dissent, truthiness, and skepticism in the global media landscape: Twenty-first century propaganda in times of war. In J. Auerbach & R. Castronovo (Eds.), *The Oxford handbook of propaganda studies* (pp. 395–417). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199764419.013.022

Boulos, S. (2021). Does foreign cyber intervention in electoral processes violate international law? In J. M. Ramírez & B. Bauzá-Abril (Eds.), *Security in the global commons and beyond* (pp. 173–193). Springer. https://doi.org/10.1007/978-3-030-67973-6_10

Chan, J. (2022). Online astroturfing: A problem beyond disinformation. *Philosophy & Social Criticism, 48*(6), 1–20. https://doi.org/10.1177/01914537221108467

Chovanec, J. (2014). The functional differentiation of the standard language. In J. Chovanec (Ed.), *Chapters from the history of Czech functional linguistics* (pp. 27–40). Masarykova univerzita.

Čeští elfové. (2022, October 19). *Čeští elfové.* https://cesti-elfove.cz/uvodni-strana/

Czech Statistical Office. (n.d.). *Age structure (Census 2021).* Retrieved February 12, 2023, from https://www.czso.cz/csu/scitani2021/vekova-struktura

Doob, L. (1950). Goebbels' principles of Nazi propaganda. *Public Opinion Quarterly, 14*(3), 419–442.

Ebbelind, A. (2015). Systemic functional linguistics as a methodological tool when researching patterns of participation. In *Proceedings of CERME 9: Ninth Congress of the European Society for Research in Mathematics Education* (pp. 3185–3191). Charles University.

Fidler, D. P. (2021). Foreign election interference and open-source anarchy. In J. D. Ohlin & D. B. Hollis (Eds.), *Defending democracies: Combating foreign election interference in a digital age* (online ed.). Oxford University Press. https://doi.org/10.1093/oso/9780197556979.003.0014

Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication, 37*(2), 145–156. https://doi.org/10.1080/10584609.2020.1723755

Giles, H., & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293–310). Lawrence Erlbaum Associates.

Golovchenko, Y., Buntain, C., Eady, G., Brown, M. A., & Tucker, J. A. (2020). Cross-platform state propaganda: Russian trolls on Twitter and YouTube during the 2016 U.S. presidential election. *The International Journal of Press/Politics, 25*(3), 357–389. https://doi.org/10.1177/1940161220912682

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). Academic Press.

Howard, P. N. (2003). Digitizing the social contract: Producing American political culture in the age of new media. *The Communication Review, 6*(3), 213–245. https://doi.org/10.1080/10714420390226270

Jack, A. (2016, November 10). There are at least eight promising business models for email newsletters. *Nieman Journalism Lab.* http://www.niemanlab.org/2016/11/there-are-at-least-eight-promising-business-models-for-email-newsletters/

Jowett, G., & O'Donnell, V. (2014). *Propaganda & persuasion* (6th ed.). SAGE.

Kuś, M., & Barczyszyn-Madziarz, P. (2020). Fact-checking initiatives as promoters of media and information literacy: The case of Poland. *Central European Journal of Communication, 13*(2), 249–265. https://doi.org/10.19195/1899-5101.13.2(26).6

Madden, M., & Zickuhr, K. (2012). *Older adults and internet use.* Pew Research Center. https://www.pewresearch.org/internet/2012/06/06/older-adults-and-internet-use/

Mathur, A., Wang, A., Schwemmer, C., Hamin, M., Stewart, B. M., & Narayanan, A. (2023). Manipulative tactics are the norm in political emails: Evidence from 300K emails from the 2020 US election cycle. *Big Data & Society, 10*(1). https://doi.org/10.1177/20539517221145371

Mihelj, S., Kondor, K., & Štětka, V. (2022). Establishing trust in experts during a crisis: Expert trustworthiness and media use during the COVID-19 pandemic. *Science Communication, 44*(3), 292–319. https://doi.org/10.1177/10755470221100558

Mounk, Y. (2018). *The people vs. democracy: Why our freedom is in danger and how to save it.* Harvard University Press.

Mýtům a konspiracím o COVID-19 věří více než třetina české internetové populace. (2021, March 2). *STEM.* https://www.stem.cz/mytum-a-konspiracim-o-covid-19-veri-vice-nez-tretina-ceske-internetove-populace/

Newman, N. (2022, June 15). Email news: Its contribution to engagement and monetisation. *Digital News Report 2022.* https://www.reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/email-news-its-contribution-to-engagement-and-monetisation

Ohlin, J. D. (2021). A roadmap for fighting election interference. *AJIL Unbound, 115*, 69–73. https://doi.org/10.1017/aju.2020.87

O'Shaughnessy, N. (2020). From disinformation to fake news: Forwards into the past. In *The SAGE handbook of propaganda* (pp. 55–70). SAGE. https://doi.org/10.4135/9781526477170

Palmer, J. M., & Wilner, A. (2024). Deterrence and foreign election intervention: Securing democracy through punishment, denial, and delegitimization. *Journal of Global Security Studies, 9*(2), Article ogae011. https://doi.org/10.1093/jogss/ogae011

Persily, N., & Tucker, J. A. (Eds.). (2020). *Social media and democracy: The state of the field, prospects for reform.* Cambridge University Press. https://doi.org/10.1017/9781108890960

Pravidelný přehled české dezinformační scény. (2022, March). *Čeští elfové.* https://cesti-elfove.cz/wp-content/uploads/MM_03_2022.pdf

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Řetězové e-maily dostává devět z deseti seniorů. (2018, September 24). *Elpida.* Retrieved February 12, 2023, from https://www.elpida.cz/fake-news-po-cesku

Searle, J. R. (1970). *Speech acts: An essay in the philosophy of language.* Cambridge University Press.

Seely, N., & Spillman, M. (2021). Email newsletters: An analysis of content from nine top news organizations. *Electronic News, 15*(3–4), 123–138. https://doi.org/10.1177/19312431211037681

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.

Štětka, V., Mazák, J., & Vochocová, L. (2021). "Nobody tells us what to write about": The disinformation media ecosystem and its consumers in the Czech Republic. *Javnost, 28*(1), 90–109.

Studzińska, Z. (2015). How Russia, step by step, wants to regain an imperial role in the global and European security system. *Connections, 14*(4), 21–42.

Taddeo, M. (2012). Information warfare: A philosophical perspective. *Philosophy & Technology, 25*, 105–120. https://doi.org/10.1007/s13347-011-0040-9

Taylor, P. M. (2003). *Munitions of the mind: A history of propaganda from the ancient world to the present era* (3rd ed.). Manchester University Press.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3144139

Van Dijk, T. A. (2005). Contextual knowledge management in discourse production. In R. Wodak & P. Chilton (Eds.), *A new agenda in (critical) discourse analysis* (pp. 71–100). John Benjamins.

Winseck, D. (2008). Information operations "blowback": Communication, propaganda and surveillance in the global war on terrorism. *International Communication Gazette, 70*(6), 419–441. https://doi.org/10.1177/1748048508096141

Woolley, S. C., & Howard, P. N. (2019). *Computational propaganda: Political parties, politicians and political manipulation on social media.* Oxford University Press.

ORIGINAL ARTICLE

DOI: 10.51480/1899-5101.18.3(41).805

Deep Symmetrism? A Populist Strategy for Political Communication and Beyond

Jacek Kołodziej

0000-0002-6655-4597

Jagiellonian University, Cracow, Poland

Abstract: The concept of *symmetrism*, even more so *deep symmetrism*, does not appear in academic literature as a distinct field or topic. The purpose of this article is to elucidate the essence of this phenomenon, believing it constitutes an important concept that bridges politics and persuasion in contemporary politics. Deep symmetrism is assumed to be the generative mechanism of political struggle in the context of the growing strength of populist argumentation, supported by the actions taken by autocratic politicians when they are allowed to utilize wider resources. Symmetrism, originally a metanarrative, has evolved into a comprehensive system of duplicating social institutions (doppelgangering) for the purpose of orchestrating political influence. As such, it closely connects to the history of propaganda and disinformation. The foundation for this article is an analysis of political communication and its contexts during the 2023 election campaign in Poland, as reflected in mediated political discussions.

Keywords: deep symmetrism, disinformation, propaganda, post-truth, doppelgangering

INTRODUCTION

The notion of *symmetrism* denotes a significant manifestation in contemporary political communication. Nevertheless, it has yet to attain recognition as an established academic term within the disciplines of political science or media and communication studies. The term has gained traction in Polish public discourse, particularly among journalists and political commentators, who employ it to critique other participants in the ongoing debate, seeking to voice concerns for the well-being of democracy based on the argument that it poses

a threat to democratic norms (Janicki & Władyka, 2023). Or conversely, in line with the argument that accusations of false symmetry are misplaced, because the objective is to overcome polarization (Czepiel, 2022). Owing to the heated nature of these disputes, *symmetrism* has become a highly charged term in public discourse, one that carries strong emotional undertones – most frequently of a pejorative kind.

Who, then, is labelled a *symmetrist* in the Polish disputes that have unfolded since 2016? Typically, it is a journalist or politician who

> …incomprehensibly maintains an equally critical distance from both Prawo i Sprawiedliwość [Law and Justice] and Platforma Obywatelska [Civic Platform][1] […] in the name of so-called objectivity. […] This becomes the perfect alibi for autocrats, allowing them to become part of the 'rules of the game', to impose their terminology, and to enforce acceptance of a new political order (Janicki & Władyka, 2023).

Thus, polarization provides the framework underlying the definition. But beyond this, the notion reveals two underlying layers for anchoring the phenomenon within theoretical frameworks. First, *symmetrism* is realized as a communicative practice that can be deployed instrumentally to obscure unethical behavior through the manipulation of public perception and awareness. Second, it is framed by the principles of objectivity and truth, so it invokes journalistic standards. Consequently, the concept of *symmetrism* embodies two inherent dimensions: one communicative, the other normative.

At its core, *symmetrist* conduct rests upon the persuasive use of the argument to symmetry – a form of cognitive proportionality perceived as natural or intuitive. It is so, because symmetry is deeply embedded as a constant motif in nature and evolution, as well as a fundamental principle in art, mathematics, or architecture. Neither the natural world nor the world we have created can be imagined without symmetry (Hahn, 1998). In social relations mirroring has become the natural defensive mechanism – as reversing the situation, rejection, and repression. It is correlated with the substance of projection, which concerns the assignment of social blame.

One of its defining features is that it operates as a second-mover tactic, deployed by an actor responding to accusations of having done or said something wrong. The tactical objective lies in turning the accusation back on the attacker, mirroring

[1] At the time of writing, the largest and deeply polarized Polish political parties are Prawo i Sprawiedliwość (PiS) – hereafter referred to as Law and Justice – and Platforma Obywatelska (PO) – hereafter referred to as Civic Platform.

the charge, thereby shifting blame and neutralizing the original allegation (Van Eemeren, Garssen, & Meuffels, 2009; Roselle, Miskimmon, & O'Loughlin, 2014).

This objective is served by a range of well-honed rhetorical techniques that vary in tone and intensity. The identification and systematization of these techniques constitute the analytical core for this analysis.

Redirecting attention to the opposing side may well stem from objectively sound reasoning or even from honorable intentions to balance the picture. Yet, equally, it may serve as a deliberate distortion of reality. In such cases, it ceases to be a neutral act of balance and becomes a weapon of propaganda. Having said that, it must be acknowledged that *symmetrism*, like propaganda, presents a similar dilemma to scholars including myself: how to reconcile a neutral term – capable of denoting phenomena that are ostensibly grave and detrimental in their consequences? How can one reconcile the legitimate, and in many respects commendable, call to safeguard democracy with the task of describing and assessing a phenomenon that may, in fact, pose a threat to it?

Naturally, context remains crucial – as ever – not only for grasping the underlying nature of the process under study but also for delineating the theoretical assumptions that guide its analysis. It is from this convergence that the theoretical framework for analyzing *symmetrism* is derived.

At the broadest level, there is the premise that the rapid spread of *symmetrism* has been enabled by the mounting crisis of liberal democracy across the world, the resurgence of autocratic mechanisms of governance, the success of populism as a pathway to power, and the intensification of political and affective polarization (Reiljan, 2020) as cumulative social effects. It is assumed here that phenomena such as *authoritarianism*, *polarization*, *populism*, and *post-truth* are mutually reinforcing and frequently co-occur, to form a new and distinctive ecosystem of the *new authoritarianism* (Applebaum, 2025).

What is important, this ecosystem contributes to the erosion of conventional norms of truth and lie – ushering in the condition of *post-truth*, which can be understood as the cumulative result of populist persuasive communication and the increasing isolation of communities within information and opinion silos (McIntyre, 2018). Another prominent author making a similar point is Moisés Naím, who argues that this system exhibits the integral, three-part nature of the "3Ps": *populism*, *polarization*, and *post-truth* (2022)[2].

Symmetrism fits significantly within this constellation for at least three key reasons: it offers a simple yet effective model of argumentation – highly compatible with populism; it relies on binary oppositions – perfectly aligned with the

[2] Interestingly, populism is understood as providing climate for seizing power, polarization has the aim of maintaining power by exacerbating conflicts within the framework of resentment and fear, and post-truth constitutes the cumulative effect of strategically deploying techniques aimed at generating confusion about factuality and truth within public life (Naím, 2022).

logic of polarization; and it adopts a relaxed, if not dismissive, attitude toward truth-based argumentation, easily linked to mechanisms of discreditation and the rhetoric of contempt.

The body of collected material serves as the foundation for the systematization of descriptive categories, as well as for the formulation of models of *symmetrist* communication. The database also includes purposefully chosen case-specific examples drawn from different national contexts, most notably: Russia (Vladimir Putin, since 2012), Hungary (Viktor Orbán and Fidesz, since 2010), Italy (Matteo Salvini's *Lega Nord* after 2013, and *Fratelli d'Italia* under Giorgia Meloni since 2022), the United States (Donald Trump during 2017–2021 and currently), the United Kingdom (Boris Johnson between 2019–2022), Brazil (Jair Bolsonaro, 2019–2022), and Venezuela (Hugo Chávez and Nicolás Maduro from 2000 to the present) – not to mention numerous individual cases from across the globe.

METHODOLOGY

The analysis is based on two collections of mediated interactions that were initially identified as related to *symmetrism*: one comprising cases from Poland, and the other consisting of international examples.

The case-specific representations drawn from international contexts were made based on a targeted search of websites belonging to media outlets and non-governmental organizations which took part in fact-checking and clarifying issues that gave rise to heated public controversies: climate change, Covid-19, Brexit, governmentally sponsored organizations (GONGO), RyanAir flight 4978, alleged secret operations of the Russian spies, and war in Ukraine[3]. The collection of international examples was intended to explore generic mechanisms of controversy in which *symmetrist* reasoning is employed.

The purpose of the Polish collection was different. It provides the primary dataset at a pragmatic level. Over the period from 2015 to 2023, the author systematically compiled a collection of examples illustrating the rhetorical strategies and persuasive techniques employed by leading political actors. Specifically, it was assumed that the most appropriate unit of analysis would be *mediated political interaction* (Thompson, 1995), within the genre of opinion journalism, specifically in mediated political disputes. This format accurately reflects the nature of binary conflict and polarization, quite often with a journalist being

[3] The following websites were included: time.com, theguardian.com, the-independent.com, lemonde.fr/en, bbc.com, itv.com, reuters.com, euronews.com, voteleavetakecontrol.org, gijn.org [Global Investigative Journalism Network], epde.org [European Platform for Democratic Elections], osce.org [Organization for Security and Co-operation in Europe], amnesty.org [Amnesty International], discuss-data.net.

an active participant. These premises served to explore the recurring patterns of persuasion. Therefore, five programs were selected from the reputable media known for presenting both sides of the political cleavage[4]. For this article, the analysis was limited to a set of 100 examples. This number proved sufficient, as the core mechanisms began to recur, while no new patterns were emerging.

The applied method of final analysis would be that of a qualitative content analysis, making use of the reconstruction of the pragmatic mechanisms of constructing *symmetrist* argumentation. One might find here the idea described by Kathy Charmaz as focused coding (2014) – with the focus placed on formal structures of argument, as well as the categorial reconstruction understood as explained by Virginia Braun and Victoria Clark in their approach to thematic analysis (2019, 2021, 2022).

The analysis enabled the description of the generic forms of *symmetrist* argumentation along with the identification of the characteristic modes of its application. Approaching *symmetrism* as a persuasive mechanism, its full potential becomes visible once an autocratic politician attains power. This leads to a fundamental question: has *symmetrism* exhausted its function at the rhetorical level, or does it transcend beyond that? Why wouldn't autocratic politicians in power take full advantage of their resources to showcase their successes?

GENERIC MODEL OF SYMMETRIC POLITICAL COMMUNICATION AND ITS PRAGMATIC EXECUTIONS

Symmetrism, as was suggested, takes on a defensive-offensive nature as a specific response to an accusation. The reversal of the argument draws upon the rhetorical figure of *tu quoque*. We find it in the biblical principle "He that is without sin among you, let him first cast a stone at her" (John 8:7, New International Version). Initially, Jesus' metaphor was an appeal to empathy, honesty, and restraint in the exercise of authority or judgment (Brown, 1966), thus constituting the very essence of humanism – human imperfection. Yet, in terms of ethical standards, Jesus' words have become a metaphor employed to denounce double standards, that is, hypocrisy.

The *tu quoque* principle is widely applied in sophistic traditions as a generator of *ad hominem* argumentation. Its simplicity generates its power, so it is hardly surprising that it started to be systematically facilitated in politics.

[4] Kawa na ławę [Straight Talk Over Coffee] and Fakty po Faktach [Facts After the Facts] on the liberal channel TVN 24, Minęła 20 [After 8 PM] on the Law and Justice-biased TVP Info, Siódmy dzień tygodnia [The Seventh Day of the Week] on the liberal Radio ZET, and Wybory w TOKu [Elections on TOK FM] on the left-leaning progressive radio station TokFM.

In modern times, the technique has been refined to perfection in Russia, where its origins date back to the Tsarist era. One of the classic examples is the famous Russian counter-argument: "And you are lynching Negroes." Its source lies in the actual words of the Russian Minister of the Interior, Vyacheslav von Plehve, in 1903, in response to Western accusations against Tsarist Russia following the Kishinev pogrom. Von Plehve is reported to have said: "The Russian peasants were driven to frenzy. Excited by race and religious hatred, and under the influence of alcohol, they were worse than the people of the Southern States of America when they lynch negroes" (Lindemann, 2000, p. 378).

Equally meaningful are the efforts undertaken by Russia in the early 1930s to mask this strategy by joking. We find an example in the campaign run by the satirical magazine of the Russian Communist Party Moscow Committee *Bezbozhnik u Stanka* (Pospielovsky 1988) – read in Great Britain and the U.S. – which in the 1930s popularized the slogan as a stock response of the "Russian computer" to questions posed by Western journalists about the low standards of living in Russia. The accompanying campaign spread across Europe, and the joke gained resonance in numerous languages worldwide.

Tu quoque, consciously employed, is said to be closely integrated with the practices of Russian disinformation since its institutional set up in 1923 (Pomerantsev, 2014; Rid, 2020). In XXI century it still finds its crucial place as a one of the principal mechanisms in hybrid warfare, where creating *post-truth* has become one of the strategic goals (Galeotti, 2019).

Tu quoque has worldwide application. It was used to refute one of the three charges against Admiral Karl Dönitz at the Nuremberg Trial, accused of ordering German submarines to sink neutral vessels without warning. Dönitz's counsel demonstrated that the Allies had engaged in the same practice – as in 1940, when the British Admiralty ordered to sink all vessels in the Skagerrak on sight. Similar orders were given by Admiral Chester W. Nimitz, the U.S. commander-in-chief, when the United States initiated unrestricted submarine warfare against Japan (Zabecki, 2007). Here, the *tu quoque* strategy eased the severity of the accusation by suggesting that "your side does the same", in a fair criminal trial.

But as a key generative mechanism of *symmetrism* it unlocks a full spectrum of persuasive stratagems[5], structured around a simple model of rebuttal, which can be illustrated using the following elements: p, q – issues discussed, A and B – politicians of opposite sides, *neg* – negative value (blame assign), *pos* – positive value (credit assign), J – journalist (as: interviewer, mediator, biased interlocutor, symbolic prosecutor etc.), Public – public opinion.

[5] Cambridge Dictionary (n.d.) defines a stratagem as "a carefully planned way of achieving a particular goal or of dealing with a particular problem, often by using a trick". So it is more than a mere rhetorical device – it takes the form of a move in a game, assuming it carries a strategic dimension combined with tactical cunning.

The mediatized interaction usually consists of several rounds, which, for this study, are here reduced to three basic ones:

STAGE 1. ACCUSATION
A/J→B (A/J frame p with *neg*, assigning blame to B)

STAGE 2. DEFENSE AND COUNTER-ATTACK
B→A/J: B rejects the accusation and mirrors it back with *neg*, claiming A is equally or more guilty

STAGE 3. POLARIZING SELF-DEFINITION
B→Public: B reframes their own role in terms of *pos/neg*, by manufacturing blame to A and mutually assigning credit to themselves

This sequence is reproduced in a limited number of variants, including the more or less active role of a journalist. Politicians perform specific communicative roles within a polarized spectacle. They aim to dominate their opponents through blame attribution and stigmatization, aligning these tactics with the party's message of the day and spin. This type of interaction kept resurfacing across all the examined programs.

Model-wise, politician B, in their rebutting an accusation, essentially limits themselves to one of the following four options: *denial of source or actor's credibility*, *denial of facts*, *whataboutism*, *and bothsidesism*. These four manifestations of *symmetrism* merit a more in-depth analysis.

The collected data show that denial remains the most frequently used application. It manifests itself in two versions. The first and most common version refers to the dichotomy between truth-telling and lying, as well as its variant – the dichotomy between invoking a credible versus a deceptive source of information. Below is a script for this stratagem.

STAGE 1. ACCUSATION
A/J→B (A/J frame p with *neg*, assigning blame to B)

STAGE 2. DENIAL BY DISCREDITATION OF INTERLOCUTOR / SOURCE
B→A/J: B does not directly deny *p*, but undermines A's or J's credibility as a truth-teller, or their source

STAGE 3. POLARIZING SELF-DEFINITION
B→Public: B reframes their own role in terms of *pos/neg*, by manufacturing blame to the other actor and mutually assigning credit to themselves in terms of truth vs lies / credibility vs unreliability

A notable example is the Polish case involving the unauthorized use of Pegasus during the 2019 election campaign and subsequent period. The endless media duels between Law and Justice and the liberal opposition provide numerous pragmatic manifestations of *denial*. The model below focuses on the generic variants of Politician B's rebuttal.

STAGE 1. ACCUSATION
J→B: We have learned that the head of the Civic Platform's campaign (…) was surveilled using the Pegasus system. Doesn't this story shock and disgust you?

STAGE 2. DENIAL (PRAGMATIC VARIANTS)
B1: It did not happen. It is a fake. Russian propaganda.
B2: Which "Pegasus" are you talking about? The horse with wings?
B3: No, I haven't heard about it, I don't know, I am busy these days…
B4: Perhaps it was so, but it needs to be checked.
B5: These are just media reports, not facts. I don't comment on that.
B6: It surely comes from *Gazeta Wyborcza*, so it's not credible.
B7: How do you know that? That's impossible.

Apart from option (B1), which is a fundamental denial, the subsequent variants allow for a soft avoidance of the accusation while retaining the opportunity for a rebuttal, in line with reasoning based on the fallacy of false conditioning: "Since I don't know anything about p, therefore p is false, and if you claim that p happened, you are manipulating". These primary versions of playing ignorant (B3, B4, B7) complement mockery and ridicule (B2), a perfect trick to avoid substantive discussion. A less elegant variation is the discrediting of the source of information, either in the most general form (B5) or by using the stigmatized symbol of the liberal voice (B6) associated with Law and Justice.

It confirms that a characteristic feature of contemporary public debates is the overuse of references to truth-telling and lying. Of course, this is a double-edged sword. The overuse of such a weapon may lead to a severe devaluation of truth, which is definitely worth noting, due to broader social consequences.

A similar stratagem, *denial of facts*, has also firmly established itself. The method of denying a real event may seem nonsensical; yet, its irrational nature only reinforces its persuasive appeal. Striking examples include the denials issued by states such as China or Russia when accused of violating international law and human rights.

Starting from China, according to activists, there is a network of mass detention camps holding more than a million ethnic minorities as prisoners in the western region of Xinjiang. Former detainees have described being tortured during interrogation there, living in crowded cells, and being subjected to a brutal daily

regimen and party indoctrination that drove some people to suicide (Martina, 2019). The provincial governor replied: “These kinds of statements are completely fabricated lies, and are extraordinarily absurd.” “They are the same as boarding schools,” and the personal freedoms of the “students” were guaranteed. These “boarding schools” are not concentration camps – that are “home to millions of Uighurs and other Muslim ethnic minorities” (Martina, 2019).

When ex-Russian spy Sergei Skripal and his daughter Yulia were found poisoned in Salisbury in 2018, British commentators pointed to the Russian secret services’ use of the Novichok nerve agent. Prime Minister Theresa May declared that this was the pure form of Novichok, and it was highly likely Moscow stands behind the attack. Russian Foreign Minister Sergei Lavrov’s reply could serve as a textbook example of *denial* plus *tu quoque*. He said that the nerve agent was never produced in the Soviet Union or Russia, but also “this formulation was in the inventory of the United States, Britain, and other NATO states” (Reuters, 2018).

The stratagem of *denial of facts* operates according to the following scheme:

STAGE 1 – ACCUSATION (AS ABOVE)

STAGE 2 – DENIAL OF THE FACT
B→A: B rejects the factual basis of the accusation, claiming it never happened or the evidence is false.

STAGE 3. POLARIZING SELF-DEFINITION
B→Public: B reframes their own role in terms of *pos/neg*, by assigning credit to themselves and manufacturing blame to the interlocutor by questioning their sense of realism

The next stratagem, *whataboutism*, is arguably the most well-known one from the literature on international relations, and stories about diplomats redirecting public attention toward other issues when blamed for something. Interestingly, this stratagem sometimes functions as a synthesis of all forms of *symmetrism* – e.g., in the empirical analysis by Wilfred M. Chow and Dov H. Levin (2024). We contend, however, that this is not entirely justified, since there are apparent differences in the underlying mechanisms of argumentation among them. Whereas the basic principle of *tu quoque* is strictly mirroring, *whataboutism* consists in shifting attention to an issue different from p. It fully deserves the name of “distractive *symmetrism*”.

Whataboutism’s name derives from the English expression “what about”, a ready-made formula for launching a *symmetrist* rebuttal. Linguist Ben Zimmer (2017) notes that the term was first coined as “whataboutery” and “the whatabouts” in reference to the Irish Republican Army in the 1970s; however, it gained

widespread currency in Russian propaganda during the Cold War. It unfolds in accordance with the following model:

STAGE 1. ACCUSATION (AS ABOVE)

STAGE 2. DIVERSION OF ATTENTION
B→A/J: B does not refute the charge, but diverts attention by raising a different issue (q) linked to A, attaching *neg* to it

STAGE 3. POLARIZING SELF-DEFINITION
B→Public: B reframes their own role in terms of *pos/neg*, by redirecting the audience's attention to an alternative issue, which casts A in a negative light.

The *Washington Post* reporter Dan Zak opens his illuminating article on *whataboutism* with the following sequence taken from Donald Trump's responses to accusations:

> What about antifa? What about free speech? What about the guy who shot Steve Scalise? What about the mosque in Minnesota that got bombed? What about North Korea? What about murders in Chicago? What about Ivanka at the G-20? What about Vince Foster? If white pride is bad, then what about gay pride? What about the stock market? What about those 33,000 deleted emails? What about Hitler? What about the Crusades? What about the asteroid that may one day kill us all? What about Benghazi? (Zak, 2017)

These examples show how deeply this persuasive structure is intertwined with a natural defensive reaction. A specific, slightly more detailed example of Trump's rebuttal on Charlottesville's riots in August 2017 provides contextual depth. After the clashes in August 2017, white supremacists, neo-Nazis, and members of the Ku Klux Klan gathered near the statue of Confederate leader Robert E. Lee for the "Unite the Right" rally and violently confronted counter-protesters. President Trump, siding with the radicals, opposed the idea of dismantling Lee's monument by arguing: "Robert E. Lee? What about George Washington? George Washington was a slave owner – are we going to take down statues to George Washington? How about Thomas Jefferson?" (Zak, 2017).

The last stratagem, *bothsidesism*, adds one more dimension. It skillfully waters down the attribution of blame. In this case, the essence of the persuasive mechanism is relativization, producing the impression that both sides are equally guilty, or just that "everybody is the same". Whereas *whataboutism* seeks to divert attention from B's guilt, *bothsidesism* advances the claim that guilt belongs to everyone.

This creates a particular difficulty in identifying unequivocal examples of *bothsidesism*. It often becomes a well-disguised stratagem – since it is crucial to distinguish between the deliberate distortion of blame attribution by a propagandist and the natural, good-faith attempt to identify fault in a wider context.

The EU-mandated Tagliavini inquiry into the 2008 Russia–Georgia conflict clearly demonstrates this. The report indicates that Georgian forces were the first to open fire (but did so in response to prior Russian provocations). For this reason, much media coverage portrayed the situation as one of shared responsibility. As the report states: "In the mission's view, it was Georgia which triggered off the war" (while acknowledging earlier Russian provocations) (Radio Free Europe/Radio Liberty, 2009). One might ask, were Russia and Georgia truly equally at fault at the time?

Bothsidesism often triggers relativization through generalization. This can be observed, for example, in UN reports on the Yemeni civil war from 2019, which characterized the conflict as a "collective failure, collective responsibility," and detailed possible war crimes committed by multiple parties involved (United Nations Human Rights Council, 2019). However, one can find a paradigmatic example of *bothsidesism* in the renowned Ben Rhodes' book, "After the Fall" (2021), with a point made, that Vladimir Putin came to power not by convincing people that he was honest, but by convincing people that… everyone was corrupt. Such generalization disarms any accusation of corruption – once it becomes the norm. What matters is that "he is really our guy."

The structure of *bothsidesism* unfolds in the following way:

STAGE 1. ACCUSATION (AS ABOVE)

STAGE 2. MANUFACTURING EQUALITY OF BOTH SIDES (FALSE BALANCE)
B→A/J: Instead of directly rejecting the accusation, B reframes the issue by asserting that both sides are equally guilty.

STAGE 3. POLARIZING SELF-DEFINITION
B→Public: B reframes their own role in terms of *pos/neg*, by positioning the issue as a general systemic problem rather than their own responsibility

The four identified stratagems relate directly to *symmetrism*; however, there is no doubt that, at the level of rhetorical composition, they can intersect with other tropes, such as those documented in the context of relativization – e.g., *minimization*, *false normalization*, *trivialization*, *temporal distancing* ("it was a long time ago"), *ironization*, *mockery,* or *sarcasm* (EUvsDisinfo, 2021).

Building on what has been said thus far, let's venture an operational definition of *symmetrism*. It is a political persuasive strategy, whereby communicative

actions are aligned with the principle of mirroring to disarm the arguments of the opponent. It is basically a second-mover activity employed by actors against whom credible accusations of wrongdoing have been raised. Its essence lies in neutralizing the charge by shifting guilt and shame onto the accuser, indicting them in turn on the grounds that they are, in fact, in a similar position (Van Eemeren, et al., 2009). The intended effect of such a strategy is to undermine the opponent's credibility, to arouse doubts about their motives and integrity – in short, to portray them as a hypocrite.

In Table 1 we take into account all of these variants of *symmetrism*, to re-affirm their key distinguishing features.

Table 1. Symmetrist stratagems, their specific objectives and modes of execution

Stratagems	Core mechanism – tactical objective	Core mechanism – mode of execution
tu quoque [general idea]	shifting blame back by claiming symmetry of guilt	"You accuse us of mishandling p, but you did the same (or worse)"
denial of credibility	undermining accuser's standing as truth-teller	"Why believe A? They lie, they lack credibility"
denial of facts	rejecting factual basis of accusation	"That never happened, p never occurred / evidence is false"
whataboutism	diverting to different and negative issue tied to A	"Talking about p? What about q under A? That was worse"
bothsidesism	equalizing guilt (claim: both sides are the same)	"Problems with p exist on both sides; A is no better"

Source: Own elaboration

FROM SYMMETRISM TO DEEP SYMMETRISM

The forms of *symmetrism* described have been rhetorical in nature. However, discourse is not just language, but rather a multilayered social practice that connects language with power, ideology, and cognition. Discourse, therefore, involves connecting words and ideas to institutional processes that encompass power relations and ideological legitimization (van Dijk, 2008).

When Dmitry Dubrovsky, a professor at Columbia University claims that *whataboutism* extends beyond rhetoric, adding that "it is not only a narrative practice; it is real policy" (Dubrovsky, as cited in Kurtzleben, 2017), he supports the claim that the logic of *symmetrism* ought to be interpreted comprehensively,

with the inclusion of a fourth stage, as only then does its capacity to permeate reality and exert a totalizing force become fully apparent:

> STAGE 1. ACCUSATION
>
> STAGE 2. DEFENSE AND COUNTER-ATTACK
>
> STAGE 3. POLARIZING SELF-DEFINITION
>
> STAGE 4. THE PERFORMANCE OF DISCOURSE LEGITIMIZATION (in ways determined by the resources and the politician's confidence and sense of impunity)
> B→Public: B grounds its policy by setting up the environment of controlled normative power institutions

The essence of Stage 4 lies in extending rhetoric to institutional normative guarantees, which shows why taking control over the judiciary and the media is so crucial. These are the bodies that define the normative order: the courts, the tribunals, the watchdogs. Public narratives surrounding current developments reveal an increasing use of *symmetrism*-related terminology particularly that associated with that stage – such as *Doppelgänger* or *Doppelstaat*.

The German *Doppelgänger* means "double-goer" – a person's ghostly double or counterpart[6]. The term *Doppelstaat* points to the normative and institutional state dualism – a legal concept that legitimized the takeover of political control of the Nazi German state during WWII, when the politicized "will of the people" was an excuse for mirroring legal institutions by the totalitarian rulers (Fraenkel, 2006).

We have been witnessing a similar process of state dualism in Poland after 2015, with the mirroring of institutions like a new chamber of the Supreme Court, a new formula for the National Council of the Judiciary, or the National Media Council – all introduced by Law and Justice. It is telling how difficult it was for language to keep up with institutional changes, as its users tried to encompass an entire range of *symmetric* entities: judges and *neo-judges* or *duplicate judges*, courts and *non-courts*, decisions and *non*-decisions, journalists and *non-journalists*, all framed within the added context of blurring the distinction between reality and *neo-reality* of a *double state*.

[6] In eighteenth-century Romantic literature it was used by Jean Paul in Siebenkäs, where he described a sinister double of a living person (Oxford University Press, n.d.). Since then, the term has also appeared in relation to propaganda, for instance, in reference to the current stage of Russia's disinformation war of 2024, named with the expression *Operation Doppelgänger* (Frühwirth et al., 2024).

The idea of a *symmetric* reality was perhaps best represented in the division of the Cold War, where two systems competed with one another in every detail. However, striking examples still prevail, such as the Intervision Song Contest. Initially launched from the 1960s to 1980 on the Eastern side, it was intended to mirror the original Eurovision Song Contest. It has recently been reinstated by Vladimir Putin, who on 3 February 2025 ordered the return of the Intervision by decree. Only friends were invited to promote "the real music"[7].

Vivid illustrations of institutional mirroring can be found worldwide in many autocracies. They include key state authorities, such as the National Constituent Assembly in Venezuela (2017), the All-Belarusian People's Assembly (a second parliamentary chamber, constitutionally framed in 2024), or the Constitutional Tribunal in Poland, in which the previous bench of judges was replaced by biased politicians appointed by the parliamentary majority.

One of the more intriguing domains is the mirroring of non-governmental organizations, by setting up the so-called GONGOs (*government-organized non-governmental organizations*). In Hungary, it is the CÖF–CÖKA (Civil Cooperation Forum), a government-funded organization that organizes "peace marches." Its purpose is to demonstrate that there also exists a "pro-government variant of civil society" (Narsee, Negri, et al., 2024).

Russia offers an abundance of examples, some of which stand out for their boldness – such as the non-governmental World Without Nazism, founded by Boris Spiegel, a member of the Russian Federation Council. Its declared purpose is the "monitoring of Nazism," especially in the activities of institutions critical of Moscow, such as Human Rights Watch, Amnesty International, or Memorial (Kirchick, 2015). The operations of WWN make it possible to adapt the scope of the concept of Nazism to the objectives of the Federation's policy. Another definitely notable example is the Russian Institute for Democracy and Cooperation, established in New York (2009–2015) and Paris (2009-present), and sponsored by President Putin. Its stated mission was "to study Western democracy and offer recommendations for its improvement," as well as "to provide a symmetrical response to the allegations of Freedom House about human rights violations in Russia" (Kanevskaya, 2014; Wayback Machine, 2024).

The next example comes from Azerbaijan. In 2024, the incumbent president, Ilham Aliyev, who has held office since 2003, won a fifth consecutive term with over 92% of the vote. The Organization for Security and Co-operation in Europe stated that "the elections did not offer voters genuine political alternatives [and were] overly restrictive of fundamental freedoms." By contrast, a *symmetrist* institution, the Shanghai Cooperation Organization, hired by the authorities

[7] Sergey Lavrov stressed that "Intervision would be free of perversion and abuse of human nature, as we saw in the Paris Olympics" (Novaya Gazeta, 2005).

and composed of 16 local members together with an ad hoc group of friendly "international observers", issued a statement declaring that "the elections have been transparent, credible, and democratic" (Andrusz & Ivanishcheva, 2024).

All of these examples support the assumption that the institutional layer of *symmetrism* cannot be separated from its rhetorical layer, as the two mutually condition and reinforce each other. In view of what is at stake – put grandly, the future political order and the shape of democracy – it is worth pursuing the most meticulous possible explanation of this fascinating phenomenon.

DISCUSSION AND CONCLUSIONS

The experimental analysis of the impact of *symmetrism* in public diplomacy on American society (Chow & Levin, 2024) demonstrates its effectiveness. The American public is becoming increasingly inclined to reject U.S. practices toward other states as justified, and even shifting its perception of what constitutes the norm in international relations. These results are relevant to all comparable societies.

The analysis of *deep symmetrism* presented here leads to three questions open for discussion. First, why are the potential long-term and adverse effects of symmetrism reasonably considered relevant? Second, if these hypothetical effects indeed constitute a threat, how might they be countered? Moreover, how does deep *symmetrism* fit into the normative standards and best practices of journalism, due to its alleged impact on perceptions of truth and objectivity?

From the perspective of normative concern for the common good, it is justified to claim that *symmetrism* is a tool of propaganda that proves harmful when driven by the ill will and cynicism of politicians or journalists. Furthermore, that has probably always been – and now more than ever – timely.

At this point, one may attempt to sketch, using the broadest strokes, the fundamental issues at stake. One such area is the domain of social knowledge, traditionally grounded in facts perceived as objective events. Through *symmetrism*, what is claimed here, the boundaries of objectivity and factuality begin to blur: any event can be denied, dismissed as fake, interpreted as manipulation, or framed as a hostile media construction; relativized, or cloned with a mirrored evaluation. Furthermore, any well justified judgment can be recast as falsehood, and any fabricated news can be sold as fact.

Thus, the overdose of *symmetrism*, together with polarization, in the service of bad intentions may ultimately dismantle the very conditions for perceiving truth – the central cognitive value upon which social order and consensus depend. One piece of evidence pointing to justifiable cause for alarm is provided by the findings of the Pew Research Center (2019), showing that voters of the two largest

U.S. parties could agree on the truth of basic facts only in 23 percent of cases combined. This must be considered striking, given that the phenomenon is surely not confined to the United States – but develops in every system with rising levels of autocracy, lowering trust in public institutions, and electoral cynicism.

Can anything be done to counteract the tendency, and if so, how? As always, in the face of the prevailing political winds, the first response that comes to mind is one of skepticism. Yet it is also clear that, as always, there remains the possibility of broadening our knowledge about *symmetrist* mechanisms, of learning to recognize their potentially harmful manifestations in public and political communication. It is also possible to ensure that each *symmetrist* tool is explicitly named and promptly denounced as soon as it is deployed. Expanding social awareness at this level of political communication is always greatly needed.

One might even consider initiating a serious discussion involving opinion-forming journalistic and media environments, widely regarded as independent and responsible, as well as bodies that might reconsider professional media standards, with the aim of more actively banning practices marked by a distinctly *symmetrism* with anti-democratic intent.

The above proposal is bold, and may appear as naïve as it looks unrealistic – for in a market-based system of independent media, any attempt to impose operational rules inevitably amounts to interference with conventional freedoms. That is true, but let us look at the other side – the compelling grounds to recognize that *symmetrist* practices indeed converge with the standards of mainstream media, even those regarded as paragons of journalistic objectivity, with the BBC at the forefront.

This became apparent in public debates on such topics as climate change, vaccines, migration, or the aggression against Ukraine. In the case of climate, *symmetrically* interpreted professional standards led to justification of presenting scientific knowledge on an equal footing with the views of climate deniers and unscientific assumptions. The "death of expert knowledge" resulting from populism (Nichols, 2017) generates a void that is frequently occupied by self-proclaimed pseudo-experts. In their report, the British Parliamentary Committee on Science and Technology wrote: "Some editors appear to be particularly poor at determining the level of scientific expertise of contributors in debates, putting up lobbyists against top scientists as though their arguments on the science carry equal weight" (Bawden, 2014). It is exactly about the BBC presentation of climate change. The BBC, however, did not relent, arguing that "even if science dominates, political decisions on climate are about values, choices, and policies, so skeptics must be given space in order for the audience to see resistance to climate policy and debate about its costs and consequences" (Sweney, 2018).

In other words, the professional principles of impartiality and pluralism served the BBC as justification – when claims grounded in broad scientific

consensus were treated as only an option. BBC Radio 4 – for example – was many times giving the floor to Lord Nigel Lawson, a popular politician in Margaret Thatcher's government, the Brexit supporter, and climate skeptic (Sweney, 2018).

The use of false equivalence to ensure media appeal, under the guise of upholding journalistic standards, is no joke – as in the oft-quoted parody of a television talk show: "Tonight, we'll be talking to a scientist about the solar system. To balance her views, we've invited a member of the Flat Earth Society" (Wikimedia Commons contributors, n.d.). This makes a fundamental problem, closely linked to the populist idea to replace the voices of experts with the voices of "ordinary people," pronouncing their own truths.

Symmetrist "second-opinion science" and journalistic practices of false balancing – or even the privileging of advocates of "alternative truths" – ought to be recognized as violations of professional standards of news journalism. This issue could be pushed further, in line with the continuation of our present discussion, to the question of whether it makes sense to invite anti-democrats and populists into democratic debate at all.

REFERENCES

Andrusz, K., & Ivanishcheva, A. (2024, September 2). *Azerbaijan's elections devoid of real competition amid diminishing respect for fundamental freedoms, but efficiently prepared: International observers*. Organization for Security and Co-operation in Europe. https://www.osce.org/odihr/elections/575509

Applebaum, A. (2025). *Autocracy, Inc.: The dictators who want to run the world*. Penguin Random House.

Bawden, T. (2014, April 3). MPs accuse the BBC of creating "false balance" on climate change with unqualified sceptics. *The Independent*. https://www.independent.co.uk/climate-change/news/mps-accuse-bbc-of-creating-false-balance-on-climate-change-with-unqualified-sceptics-9231176.html

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology, 9*(1), 3–26. https://doi.org/10.1037/qup0000196

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.

Brown, R. E. (1966). *The Gospel according to John I–XII: Introduction, translation, and notes* (Vol. 29). Yale University Press.

Cambridge University Press. (n.d.). *Stratagem*. In *Cambridge Dictionary* (Online). Retrieved May 29, 2025, from https://dictionary.cambridge.org/dictionary/english/stratagem

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE Publications.

Chow, W. M., & Levin, D. H. (2024). The diplomacy of whataboutism and US foreign policy attitudes. *International Organization, 78*(1), 103–133. https://doi.org/10.1017/S002081832400002X

Czepiel, A. (2022). *Ontologia symetryzmu. Umiarkowanie polityczne jako nowy romantyzm* [*The ontology of symmetrism: Political moderation as a new romanticism*]. Instytut Wydawniczy ZNAK.

EUvsDisinfo. (2021). *Modus Trollerandi*. https://euvsdisinfo.eu/modus-trollerandi

Fraenkel, E. (2006). *The dual state: A contribution to the theory of dictatorship* (E. A. Shils, Trans.). Transaction Publishers. (Original work published 1941)

Frühwirth, L., Nazari, S., Terroile, M., Lesser, C., Stradner, I., de Sessa, C., DiStasi, L., Savranska, K., Uhlig, D., Szymkiewicz, A., Siedin, O., Pivtorak, O., Roussel, F., & Bouchaud, P. (2024). *Fool me once: Russian influence operation Doppelgänger continues on X and Facebook* [Counter Disinformation Network Report]. Center for Monitoring, Analysis & Strategy; Alliance4Europe. https://alliance4europe.eu/wp-content/uploads/2024/09/CDN-Report-–-Fool-Me-Once_-Russian-Influence-Operation-Doppelganger-Continues-on-X-and-Facebook-–-September-2024.pdf

Galeotti, M. (2019). The mythical "Gerasimov Doctrine" and the language of threat. *Critical Studies on Security, 7*(2), 157–161. https://doi.org/10.1080/21624887.2018.1441623

Hahn, W. (1998). *Symmetry as a developmental principle in nature and art*. World Scientific Publishing Company.

Janicki, M., & Władyka, W. (2023). *Symetryści. Jak się pomaga autokratycznej władzy* [*Symmetrists: How authoritarian power is enabled*]. Polityka Sp. z o.o.

Kanevskaya, N. (2014, June 24). How the Kremlin wields its soft power in France. *Radio Free Europe/Radio Liberty*. https://www.rferl.org/a/russia-soft-power-france/25433946.html

Kirchick, J. (2015, February 8). Anti-Nazi group secretly helping Kremlin rebuild Russian empire. *The Daily Beast*. https://www.thedailybeast.com/anti-nazi-group-secretly-helping-kremlin-rebuild-russian-empire/

Kurtzleben, D. (2017, March 17). Trump embraces one of Russia's favorite propaganda tactics: Whataboutism. *NPR*. https://www.npr.org/2017/03/17/520435073/trump-embraces-one-of-russias-favorite-propaganda-tactics-whataboutism

Lindemann, A. S. (2000). *Esau's tears: Modern anti-Semitism and the rise of the Jews*. Cambridge University Press.

Martina, M. (2019, March 18). China says Xinjiang has "boarding schools," not "concentration camps." *Reuters*. https://www.reuters.com/world/china-says-xinjiang-has-boarding-schools-not-concentration-camps-idUSKBN1QT1D5/

McIntyre, L. (2018). *Post-truth*. MIT Press.

Naím, M. (2022). *The revenge of power: How autocrats are reinventing politics for the twenty-first century*. St. Martin's Press.

Narsee, A., Negri, G., & others (Eds.). (2024). *Civic space report 2023: Hungary*. https://civic-forum.eu/wp-content/uploads/2023/03/Civic-Space-Report-2023-HUNGARY-European-Civic-Forum.pdf

Nichols, T. (2017). *The death of expertise: The campaign against established knowledge and why it matters*. Oxford University Press.

Novaya Gazeta (2025, February 5). Russia to reboot Soviet-era song contest as 'perversion-free' alternative to Eurovision.

https://novayagazeta.eu/articles/2025/02/05/russia-to-reboot-soviet-era-song-contest-as-perversion-free-alternative-to-eurovision-en-news

Oxford University Press. (n.d.). *Doppelgänger*. In *Oxford English Dictionary* (3rd ed.). Retrieved September 28, 2025, from https://www.oed.com/

Pew Research Center. (2019, October 10). *Partisan antipathy: More intense, more personal*. https://www.pewresearch.org/politics/2019/10/10/partisan-antipathy-more-intense-more-personal

Pomerantsev, P. (2014). *Nothing is true and everything is possible: The surreal heart of the new Russia*. PublicAffairs.

Pospielovsky, D. V. (1988). *A history of Soviet atheism in theory and practice, and the believer* (Vol. 2: *Soviet antireligious campaigns and persecutions*). St. Martin's Press.

Radio Free Europe/Radio Liberty. (2009, September 30). Report blames both sides, says Georgia started 2008 war (Reuters). https://www.rferl.org/a/Report_Blames_Both_Sides_Says_Georgia_Started_2008_War/1840381.html

Reiljan, A. (2020). "Fear and loathing across party lines" (also) in Europe: Affective polarization in European party systems. *European Journal of Political Research, 59*(2), 376–396.

Reuters. (2018, April 14). Russia's Lavrov says Skripals may have been poisoned by substance Russia never made. https://www.reuters.com/article/us-britain-russia-skripal-lavrov/russias-lavrov-says-skripals-may-have-been-poisoned-by-substance-russia-never-made-idUSKBN1HL17K/

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. Farrar, Straus and Giroux.

Rhodes, B. (2021). *After the fall: Being American in the world we've made*. Random House.

Sweney, M. (2018, April 9). BBC Radio 4 broke accuracy rules in Nigel Lawson climate change interview. *The Guardian*. https://www.theguardian.com/environment/2018/apr/09/bbc-radio-4-broke-impartiality-rules-in-nigel-lawson-climate-change-interview

Thompson, J. B. (1995). *The media and modernity: A social theory of the media*. Stanford University Press.

United Nations Human Rights Council. (2019, September 3). *Yemen: Collective failure, collective responsibility – UN expert report*. https://www.ohchr.org/en/press-releases/2019/09/yemen-collective-failure-collective-responsibility-un-expert-report

Van Dijk, T. A. (2008). *Discourse and context: A sociocognitive approach*. Cambridge University Press.

Van Eemeren, F., Garssen, B., & Meuffels, B. (2009). *Fallacies and judgments of reasonableness: Empirical research concerning the pragma-dialectical discussion rules*. Springer.

Wayback Machine, Internet Archive. (2024). Институт демократии и сотрудничества [Institute for Democracy and Cooperation]. https://web.archive.org/web/20091025180115/http://www.idc-europe.org/ru/

Welch, D. (2002). *The Third Reich: Politics and propaganda*. Routledge.

Zabecki, D. T. (2007). *Dönitz: A defense*. Merriam Press.

Zak, D. (2017, August 18). Whataboutism: The Cold War tactic, thawed by Putin, is brandished by Donald Trump. *The Washington Post*.

Zimmer, B. (2017, June 9). The roots of the "What about?" ploy. *The Wall Street Journal*.

ORIGINAL ARTICLE

DOI: 10.51480/1899-5101.18.3(41).820

Can Disinformation be Regulated? A Comprehensive Overview of the European Union's Pertaining Initiatives

Gergely Ferenc Lendvai

Ludovika University of Public Service

Tamás Attila Szikora

Ludovika University of Public Service

János Tamás Papp

Pázmány Péter Catholic University

Krzysztof Wasilewski

Koszalin University of Technology

Abstract: This paper critically examines the European Union's regulatory approaches to combating disinformation, focusing on the Code of Practice on Disinformation, Digital Services Act, European Media Freedom Act and regulation on transparency and targeting of political advertising. Through legal analysis and literature review, the study assesses the strengths and limitations of these frameworks in addressing disinformation, particularly within the context of advertising transparency and platform accountability. Key findings reveal that while the European Union has made significant strides toward improving transparency and holding platforms accountable, current measures remain largely voluntary and lack sufficient enforcement mechanisms. The paper concludes that regulation alone is insufficient to effectively eradicate disinformation, and proposes a more holistic approach, combining media literacy (critical thinking), cross-border collaboration and adaptive strategies.

Keywords: disinformation, platform regulation, European Union, transparency, fake news

INTRODUCTION

Can disinformation *truly* be regulated? Or are we fighting a losing battle? How *should* governments and international institutions distinguish between curbing harmful disinformation and protecting free speech? Can regulatory frameworks effectively "stem the tide of fake news" without inadvertently becoming tools of censorship (Gosztonyi, 2023)? Or are other policy means (also) needed, such as the promotion of quality journalism and information literacy, as opposed to or in addition to restrictive measures (Török, 2024)? These pressing questions lie at the heart of the present study, which aims to outline the disinformation regulation in the European Union via its ambitious attempts to combat fake news through legislative actions.

The primary objective of the paper is to review whether the European Union's regulatory initiatives, including the Code of Practice on Disinformation, Digital Services Act (DSA), European Media Freedom Act (EMFA) and Regulation on the transparency and targeting of political advertising (RPA) can address the pervasive influence of disinformation in modern society. As a premise, the paper briefly explores the origins and evolution of fake news, the various typologies thereof, and how these forms of information warfare have been weaponized in political and social contexts. The core of the paper lies in the assessment of the impact and effectiveness of the EU's legal frameworks, particularly in scrutinizing the advertising ecosystem that fuels disinformation, ensuring political ads are transparent, and enhancing the integrity of online platforms. By critically examining the legislation, this study aims to reveal the limitations of the current regulatory measures and their key similarities and differences. The paper is built on legal analyses of the instruments examined as well as a comprehensive literature review.

The study aims to contribute to the growing body of literature on disinformation governance. It also sets forth as a goal to enrich the existing scholarship by highlighting gaps in enforcement, the challenges of transnational disinformation and the need for more adaptive, forward-looking strategies. The paper positions itself within broader debates on media regulation, political communication and the safeguarding of democratic processes in the digital age, and invites researchers from the social sciences to take part in the discourse on the critical examination of policymaking concerning fake news within and outside the EU.

THE CONCEPT OF DISINFORMATION AND THE CATEGORIZATION OF "FAKE NEWS"

Though fake news as a concept is not novel, nor is the spreading thereof (cf. the spreading of false information during the World Wars (Barragán-Romero and Bellido-Pérez, 2019)), the term gained notoriety in 2016 before and during the US presidential election campaign (Van Duyn and Collier, 2019; Nordberg et al., 2020).

Research underscores that the proliferation of fake information, especially from then-candidate Donald Trump, caused a significant shift in the spreading of fake news; the quantity of fake information on social media grew rapidly, and the trust in media, as well as the identification of "real" news, deteriorated significantly (Van Duyn and Collier, 2019). The disinformation crisis has significantly influenced policymaking and decision-making processes, too. Bovet and Makse (2019) found that 25 per cent of 171 million tweets during a particular period contained or spread fake or biased news, affecting political preferences and election dynamics. The spread of false information peaked again during the COVID-19 pandemic, considered one of the most severe cases of disinformation (Shrestha and Spezzano, 2022; Palomino-Flores et al., 2024). Al-Zaman (2021) highlights that in India, COVID-19-related fake news not only targeted health issues but also politics and entertainment, exacerbating the pandemic's impact. This widespread misinformation led to lower vaccination rates, disregard for safety measures, and increased infections and deaths. The World Health Organization and experts have referred to this disinformation wave as the "infodemic" (Pagoto et al., 2023), noting its global reach and cross-continental effects.

It is imperative to highlight the typology of fake news, too. Although the term "fake news" is commonly used in both popular and academic discourse, its definition remains unclear (Wasilewski, 2021, p. 5). One of the most popular understandings of "fake news" associates it with "viral posts based on fictitious accounts made to look like news reports" (Tandoc et al., 2018, p. 2). Most media scholars agree that the contemporary understanding of "fake news" should focus on news fabrication and manipulation, along with propaganda (Brennen, 2017). Still others explain "fake news" as "information disorder" together with its related challenges, such as echo chambers (Wardle and Derakhshan, 2017). Moreover, as Wardle (2023) described the above issue, it is also critical to focus on the interrelational aspects of the information disorder as well, given that false information and the disseminator thereof rarely operate in "silos", e.g., someone who posts health misinformation may very well take part in the spreading of untrue information in other "disciplines" as well, such as in politics.

Dealing with "fake news" one must be aware of its various forms. According to the literature, at least 7 main types can be distinguished: satire (parody),

misleading content, fabricated content, false content, imposter content, click-bait, propaganda, conspiracy theories, and partisan content (Bąkowicz, 2019, p. 284–285). Satire aims to fool the recipients but has no intention to harm whereas misleading content purposefully frames an issue or individual. Imposter content, on the other hand, means that reliable sources are impersonated by third parties to cause deception. A similar goal aims to achieve fabricated content, which is false in its entirety. False connections might use true stories but give them headlines that do not support the content. In this way, false context shares genuine content with false contextual information. Finally, fabricated (manipulated) content uses reliable information and modifies it in order to achieve some political (or other) agenda (Wardle, 2017). Other forms of fake news demand much broader definitions, which go beyond the scope of this paper. However, it is important to remember that fake news – or disinformation – can take various forms, which is why the issue is so problematic (Broda & Strömbäck, 2024). Different types vary in their potential to deceive recipients.

INCREASING REGULATORY DEMANDS

IMMEDIATE HISTORY: THE DYNAMICS OF HOW THE YEAR 2016 SHIFTED THE DISCOURSE ON DISINFORMATION

Online platforms have become a prominent space for political communication, including the publication of political advertisements. The harmful influence of political ads on social networking sites gained significant attention following the 2016 US presidential election. While the dissemination of false claims was not new, the scale and impact during this election, along with the 2016 Brexit campaign, raised serious concerns. These events marked the beginning of a new era in political disinformation, further highlighted by the Italian and French presidential elections in 2017.

In itself, the appearance of untrue statements during political campaigns is by no means a recent phenomenon; however, the fact that the term "fake news" has gained considerable attention is due to the platform on which it is disseminated, namely social networking sites. The basic operating principles and structural design of these sites have made it easy to facilitate the spread of previously unknown amounts of information at almost unimaginable speed.

The real novelty was the fact that, in the absence of significant resources, practically anyone can deliver messages to the masses, even in a targeted way, without revealing their identity. In fact, the use of targeted advertising messages is an indisputable feature of social media campaigning. It is also clear that, in addition to the dangers and risks that are often highlighted, they also have

a number of positive benefits for democratic (decision-making) mechanisms. These include the ability to reach voters with messages that are relevant to them, to reach those who are difficult or almost impossible to reach through other channels, and to be effective, efficient and sometimes cost-effective for politicians. At the same time, it can benefit the public by leading to more diverse political campaigns and greater awareness of certain issues among voters (Borgesius et al., 2018; Dobber et al., 2019).

As a result of the issues outlined above, accountability has also faced numerous obstacles. One reason for this is that, unlike media service providers, online platforms (e.g., Facebook) do not bear editorial responsibility for content appearing on their interfaces. Under the DSA, the activities of these services consist primarily of storing and publicly disseminating content (Article 3(i)). Through recommendation systems, ranking by platforms, and content moderation, they cannot be considered completely neutral with regard to content published by users. However, since their activities in the preliminary compilation of content differ significantly from those of "traditional" media service providers, the question of what obligations can be imposed on platforms in order to protect democratic public discourse arises.

PROPOSED SOLUTION: INITIAL EFFORTS TOWARDS TRANSPARENCY

These events have drawn attention to the fact that paid political advertising is an effective tool in large-scale and coordinated disinformation campaigns. The lack of transparency and the publication itself not knowing the identity of the "client" behind each message made it easy to avoid prosecution. The transparency and accountability requirements for political advertising are intended to serve three main purposes: to prevent disinformation and foreign influence, to facilitate the emergence of opposing views in public discourse, and to create at least the same level of transparency in online political discourse as has long been the case for paid political advertising on television and radio (Wood, 2020).

Since the second half of the 2010s, serious efforts have been made to address this problem, seeking to find solutions to ensure credible information in the spirit of transparency. Initially, the European Commission's Communication (COM(2018) 236 final), summarizing the main findings and recommendations of the report and the report's own report on misinformation, published in March 2018 (A multi-dimensional approach to disinformation, 2018), declared the need for transparency in political advertising.

On September 12, 2018, the European Commission adopted a Recommendation on security issues related to the European Parliament elections in spring 2019, in the context of the election campaign and the conduct of the elections, primarily, of course, in the online space (C(2018) 5949). Among the measures to be taken

to ensure that voters are informed and to guarantee freedom of public debate, the Commission's Recommendation mainly set out requirements for transparency in political advertising published before the European Parliament elections and during the election campaign. While the expectations set out in these documents have gone some way toward improving the safe conduct of the 2019 European Parliament elections (free from external interference), the lack of binding regulation has not fully ensured a digital environment that provides the public with transparent and reliable, credible information (Kirk and Teeling, 2022).

SELF-REGULATION OF PLATFORMS

In response to the challenges outlined earlier, platforms like Facebook (now Meta) have taken steps to address these issues through their own regulatory initiatives. In 2018, Meta introduced new rules for political advertising, requiring each ad to display the name of the organization or individual that paid for it. (Dommett, 2014). While this promotes accountability, it carries risks: Leersen and colleagues point out that Facebook's data is often incomplete or inaccurate, potentially misleading journalists (the misleading nature of the advertisement deserves special mention because if the press uses data it considers reliable to inform the public about political issues, and this data later turns out to be inaccurate, the deceptive information can cause serious damage in a democratic public sphere). Moreover, this focus on transparency may divert attention from more obscure issues, like targeted advertising practices, that are less disclosed by the platform (Leerssen et al., 2023).

Other social platforms have not taken the most radical steps against political advertising: Twitter (now known as X) announced at the end of 2019 that it will ban political advertising on its platform in the future. The idea has been widely criticized, with the main criticism being that the decision is a disproportionate restriction on political communication, which is the most precious core of freedom of expression. Another relevant issue raised by the ban concerned the definition of political advertising, namely how to distinguish between purely political advertising and messages dealing with issues that affect a wide range of society and typically give rise to lively public debate.

THE EUROPEAN UNION AND DISINFORMATION

The European Union's Code of Practice on Disinformation was introduced in 2018. It was a manifestation of the EU's so-called "self-regulation" approach to the problem of disinformation (Wasilewski, 2021). This approach can be defined as "a type of voluntary initiative which enables economic operators, social

partners, non-governmental organizations or associations to adopt common guidelines amongst themselves and for themselves" (Ilves et al., 2016). Although some researchers question the effectiveness of the approach (Shattock, 2021), it is by far the most ambitious effort by UE to tackle disinformation. The introduction of the 2018 Code of Practice was part of a broader strategy to counter the rising threat of false information, particularly following the proliferation of digital platforms and the influence of disinformation campaigns on democratic processes, including elections. The Code of Practice has since become a key element of the EU's broader digital policy framework, and its impact warrants detailed academic evaluation. It must be noted, however, that the EU "self-regulation" approach was met with some heavy criticism. Among the critics were representatives of online platforms, advertisers, academics, media, and civil society organizations. Together they issued a statement criticizing the code for being too general in scope, as well as for its lack of a common approach (The Sounding Board's, 2018).

Following critical voices, the EU has sought to strengthen its regulatory framework through the DSA. The Act was proposed in 2020 with the intention of introducing binding obligations for online platforms, including measures to counter disinformation and mandates increased transparency and accountability. The introduction of the DSA manifested a serious change in the EU approach to disinformation, as it moved from voluntary commitments toward more enforceable rules (Nannini et al., 2024; Husovec, 2024). Recognizing the need for a stronger response to the growing threat of disinformation, in 2022, the EU established the Strengthened Code of Practice on Disinformation. Its preamble states that the member states "recognize their collective and individual accountabilities to work together to defund Disinformation in advertising and media across the following types of organizations and their respective" (Code of Conduct, 2022).

A COMPREHENSIVE EXAMINATION OF THE CODE

The Code contains 44 commitments and 128 specific measures and is structured into 10 sections. Table 1 presents the structure of the 2022 Code of Practice on Disinformation.

Table 1. The Structure of the 2022 Code of Practice on Disinformation

Section	Content
Preamble	The role of signatories in combating disinformation, safeguarding democratic processes, and maintaining a balance with fundamental rights like freedom of expression.
Scrutiny of Ad Placements	Focus on reducing the revenues of purveyors of disinformation by improving transparency in advertising and ensuring responsible ad placements.

Section	Content
Political Advertising	Transparency standards for political and issue ads, including labelling, verification, and repositories to ensure users understand the origins and purposes of such ads.
Integrity of Services	Conceptualization and countermeasures regarding manipulative behaviors, such as fake accounts or malicious use of AI systems, with clear transparency obligations.
Empowering Users	The outline of the provisions for users with tools and education to recognize disinformation, including media literacy initiatives and transparency in recommender systems.
Empowering the Research Community	Researchers' access to data for studying disinformation, with structured governance for sensitive data and cooperation between stakeholders.
Empowering the Fact-Checking Community	Collaboration with independent fact-checkers, ensuring they have the necessary resources, tools, and access to data to conduct their work effectively.
Transparency Centre	A central online "hub" where information about the implementation of the Code is made available to the public and regularly updated.
Permanent Task-Force	Establishment of a Task-force which is responsible for monitoring and reviewing the implementation of the Code and ensuring it adapts to technological and societal changes.
Monitoring of the Code	The process for continuous reporting, assessment, and improvements to ensure the Code's effectiveness in reducing disinformation.

The Preamble "sets the stage" by emphasizing the collective responsibility to fight disinformation, a menace that threatens democratic values through misinformation, influence operations, and foreign interference. It stresses the need for a careful balance between combating disinformation and upholding fundamental rights like free speech and privacy, while urging coordinated efforts from platforms, advertisers, researchers, and the European Commission. This creates the framework for a holistic strategy against disinformation.

Sections II and III dive deeper into the mechanisms that fuel disinformation – financial incentives and political manipulation. Section II tackles the economic engine behind disinformation by scrutinizing ad placements, ensuring that harmful actors are denied revenue streams through brand safety policies. Section III focuses on the transparency of political ads, which are frequently weaponized in disinformation campaigns, as demonstrated by Russian interference in elections (Espaliú-Berdud, 2024). As full disclosure of sponsors and public repositories is required for oversight, these sections aim to prevent political discourse from being distorted by hidden agendas. Both sections highlight the importance of collaboration between platforms, advertisers, and fact-checkers to curb disinformation's reach.

Section IV is of critical importance as it addresses the technological side of the issue, focusing on protecting the integrity of services. By cracking down on the use of fake accounts, bots, and deepfakes, this section aims to neutralize the tools that disinformation actors often deploy. The emphasis of this section on evolving safeguards ensures that as new threats arise, the response can adapt and remain effective. Building on these protective measures, the "Empowering

Triangle" in Sections V–VII takes a proactive approach. Section V arms users with the ability to detect and report false information, bolstering media literacy and encouraging informed content consumption. Section VI grants researchers access to crucial platform data, enabling them to analyze disinformation trends and assess the effectiveness of interventions, fostering cooperation between academia and civil society. Section VII supports fact-checkers with automated data access, ensuring that their findings are integrated into platform services. By empowering these groups – users, researchers, and fact-checkers – the Code creates a multi-layered defense system against disinformation. Finally, Section VIII ties it all together with the Transparency Centre, a public platform that ensures accountability and tracks the progress of these measures. By regularly updating with metrics and reports, the Transparency Centre fosters ongoing commitment from signatories and promotes a transparent fight against disinformation. This interconnected approach ensures that every aspect of the disinformation ecosystem is addressed, from economic incentives to technological manipulation, while empowering key stakeholders to act effectively.

THE DIGITAL SERVICES ACT

The Proposal for a Regulation on a Single Market for Digital Services (COM/2020/825 final) was introduced on December 15, 2020, and following a relatively swift negotiation process, the final version was approved by the European Parliament on July 5, 2022. The DSA adopts a tiered regulatory framework, where the obligations increase in accordance with the significance of the role played by the online intermediary. It defines four categories of service providers: online intermediary services, hosting services, online platforms, and very large online platforms (VLOPs). Large social networks, where disinformation circulates with potential systemic impacts, such as the erosion of trust in democracy and institutions, are classified as VLOPs.

With respect to disinformation, the DSA introduces a "co-regulatory" framework that permits service providers to voluntarily adopt codes of conduct aimed at mitigating the harmful effects of illegal content dissemination, as well as manipulative and abusive behaviors. Article 45(1) grants the Commission and the Board the authority to support and facilitate the creation of diverse codes of conduct intended to strengthen the implementation of the DSA. These codes primarily address two areas of concern: illegal content, as defined in Article 3(h), and the broader systemic risks outlined in Article 34. The key function of these codes of conduct is to offer detailed interpretations and enhancements to the existing legal framework. They encourage the adoption of voluntary standards that surpass statutory requirements, or, when necessary, the development of broader guidelines functioning as soft law.

While adherence to these codes is voluntary, the DSA notes that a service provider's refusal to adopt a code, without providing a satisfactory rationale, may be considered when evaluating its compliance with DSA obligations. In instances where specific systemic risks are identified under Article 34, the Commission may invite service providers, relevant authorities, civil society organizations, and other stakeholders to collaborate in the development of these codes. The Commission already expressed that the Code of Practice on Disinformation is set to evolve into a formal Code of Conduct (European Commission, 2024). In this regard, the Commission appears to adopt a rather broad interpretation of co-regulation. While the establishment of the Codes of Conduct was initially voluntary, their subsequent elevation to a status formally recognized under the DSA effectively integrates them into the regulatory framework. As a result, they function as a kind of Trojan horse, gradually transforming a soft law instrument into a form of hard law.

Currently, the protection against disinformation largely relies on the willingness of information society service providers to fulfill their duties of care regarding the content they distribute. This includes their ability to self-assess the "systemic risks" inherent in their activities and implement preventive actions, particularly through content moderation procedures. However, the DSA does not specifically define harmful content, including disinformation, nor does it mandate its removal (Leiser, 2023). Additionally, the effectiveness of these measures also depends on the capacity of relevant public authorities to monitor and enforce compliance with these obligations. Whether this approach will provide enough response to the challenges disinformation poses to democratic societies is still an open question (Vicente, 2023).

Requiring large online platforms to conduct their own risk assessments is logical, as they have exclusive access to user data and understand the risks and remedies of their services. It is equally reasonable for the Commission to expect these platforms to cover compliance costs under the DSA. However, ensuring the independence of private audit firms, which may face conflicts of interest, is crucial. There appears to be a lack of a specific supervisory framework for ensuring compliance, to address this, a broader principle of transparency and a strengthened co-regulatory, multistakeholder model are needed (Strowel, 2023). Also, there needs to be access to researchers who meet certain criteria, for the purpose of conducting research that helps identify and understand systemic risks, and help combat disinformation (Cauffman, 2021).

THE EUROPEAN MEDIA FREEDOM ACT

The European Democracy Action Plan, launched at the end of 2020, is a framework developed by the European Commission to strengthen democracy within the EU. Its objectives include safeguarding the integrity of elections and political advertising, enhancing transparency, supporting independent journalism, and improving the EU's capabilities in detecting and responding to disinformation. As part of this initiative, Ursula von der Leyen announced the proposal for the EMFA in 2021, which builds on the Audiovisual Media Services Directive and establishes various regulations for the media sector (European Commission, 2020). On April 11, 2024, the EU officially adopted the legislation which consolidates various areas of media regulation, including provisions related to public service media, cooperation among national regulatory authorities, transparency of media ownership, and the distribution of state advertising.

These actions target the distribution or access to media services provided by non-EU media outlets that aim at or reach EU audiences, regardless of the distribution method. Such services, particularly when subject to control by third countries, are scrutinized for posing threats to public security or contributing to the spread of disinformation and the serious risks associated with it. The recitals associated with the Article 17 (Recitals 47–49) highlight the specific practical expertise of media authorities in protecting the internal market from activities of media services originating outside the EU that target or reach audiences within the EU and may threaten or endanger public security. Such risks include systematic international campaigns involving foreign information manipulation and interference, aimed at destabilizing the EU or any Member State.

To address these threats, the legislation aims to coordinate national measures that can be adopted to counter risks to public security posed by media services originating outside the EU, or established outside but targeting EU audiences. In this context, the legislation proposes the creation of a criteria list by the European Board for Media Services, as established under the EMFA, to assist national regulatory authorities or bodies when a media service provider seeks jurisdiction in a Member State, or when a media service provider under a Member State's jurisdiction presents a serious public security risk.

REGULATION ON THE TRANSPARENCY AND TARGETING OF POLITICAL ADVERTISING

In its Communication on the European Democracy Action Plan, published in December 2020, the European Commission, recognizing the threats posed by online platforms and social media sites, in particular to the electoral process, expressed its determination to adopt a comprehensive set of rules to ensure the visibility and transparency of political advertising for the 2024 European

Parliament elections (COM(2020) 790 final). As a first element of this commitment, in November 2021, the Commission published a draft regulation on rules of the transparency and targeting of political advertising (COM(2021) 731 final). The need for legislation was already expressed by several stakeholders before the draft was presented, in particular in the light of research findings based on information available in social networking sites' advertising directories (Dommett and Bakir, 2020). The proposal was intended as a response to the fragmentation of regulation in the EU Member States, but the increasing role of social platforms in election campaigns has undoubtedly led to calls for common EU regulation. The RPA was finally adopted in spring 2024.

The legislator has recognized the potential for misinformation through political advertising, stressing that this can occur in particular when the message is not overtly political, comes from outside the EU or, for example, uses targeted techniques (Recital 4). In principle, as the title of the legislation also indicates, the RPA lays down transparency requirements for political advertising services, such as identification (Article 7), retention and transmission of specific data on advertisers and advertisements (Articles 9–10 and 16–17). The other main part of the legislation deals with the targeting techniques and ad-delivery techniques of online political advertising, requiring political advertisers and political advertising solvers to comply with certain data protection conditions when using targeting or ad-delivery techniques (Articles 18–20).

It is important to note that the RPA does not affect the substantive content of political advertising in the legislation of each Member State. Another principle of the RPA that concerns the issue of responsibility for content posted on platforms is the prohibition of a general monitoring obligation, similar to the provision in Article 8 of the DSA (Recital 54); social networking sites should "merely" provide a suitable reporting platform and mechanism for users to report political advertising that does not comply with the rules of the RPA, or the persons who commission it (Article 15).

CRITICAL EVALUATION

Current approaches to the regulation of information reflect both progress and challenges. On the one hand, the EU has taken some significant steps in tackling fakes news in the European public sphere by introducing and modifying the Code of Practice on Disinformation, as well as the DSA. Experts agree that those documents have provided frameworks to improve so-much needed transparency, especially in political advertising. Moreover, they hold platforms accountable and – to some extent – empower users and researchers to combat disinformation. On the other hand, there is consensus that actions taken by the EU are

not enough to eradicate disinformation from the public sphere. Self-regulation measures, such as the aforementioned Code of Practice, have established voluntary commitments from social media platforms, but they lack binding enforcement mechanisms. This is why there is a growing political push for the Code to move swiftly with its conversion into a Code of Conduct under the DSA (Article 45).

Unsurprisingly, critics argue that this voluntary nature has allowed Facebook, Amazon and other large tech companies to remain blurred in their efforts, limiting the impact of these measures. Part of the problem is the very nature of the EU, which has limited competences in areas, such as internet and media policies. These are traditionally left for member states to regulate.

The 2022 Russian invasion of Ukraine has proven that dissemination of fake news is part of the contemporary warfare that may cause as much harm as traditional weaponry. What is lacking, apart from coherent legislation at a national and European level is coordination between member states and the EU when it comes to tackling disinformation. Despite some progress in this matter, there is still much to do in order to establish a successful system.

To end the evaluation, we propose that the emergence of instruments in the EU's anti-disinformation agenda is not only a rational response mechanism but also a structural necessity arising from the dispersed nature of both the problem and the Union's competences. This issue stems from a rather trivial polemic; disinformation cuts across economic, communicative, and democratic spaces, while the EU's authority is – still – fragmented between internal market, audiovisual, and electoral domains. Hence, no single act could adequately address its complexity. The Code and the Regulations, therefore, represent interlocking layers of governance that together form a composite regulatory ecosystem. In this "onion-like" layering, the principle-related foundation, lies the Code as it enables voluntary cooperation between platforms, advertisers, and civil society under the Commission's guidance. The DSA transforms this "soft" coordination into enforceable obligations by embedding the principles of transparency, accountability, and the necessity to ensure users' safety within broader risk-management architecture while the EMFA intervenes upstream by reinforcing the independence and plurality of the media sector. Lastly, the RPA targets the electoral interface, curbing opaque micro-targeting and foreign influence in political advertising.

It is also crucial to mention that the connection among these instruments is thus functional and sequential. Each act occupies a distinct but complementary regulatory space, whether it be soft law/self-regulatory, horizontal, structural, or sector-specific, all collectively seeking to transform the EU's approach into a systemic resilience-building against disinformation. Subsequently, in essence, the multiplicity of legal norms also mirrors the multidimensionality of the threat. Where earlier regimes focused on takedowns and liability, the current framework emphasizes transparency, accountability, and coordination and

in this context, the Code-DSA-EMFA-RPA quart-partite nexus demonstrates the EU's progressive understanding that combating disinformation requires much more than a single prohibitive instrument and proposes a network of mutually reinforcing mechanisms addressing different stages of information production, dissemination, and influence.

CONCLUSION

To conclude, we invite the readers to revisit the initial question: Can disinformation be regulated?

As seen from our above analysis, despite the efforts made, these instruments are solely legal *tools* but not *solutions*. Constrained by their reactive nature and the evolving tactics of disinformation campaigns, these initiatives are designed to address "symptoms" rather than the systemic causes of disinformation. We propose that the solution does not lie *solely* in regulation but in a more holistic approach that does not just include but rather promotes strengthening media literacy, enhancing cross-border collaboration, and developing adaptive and progressive strategies to fight disinformation.

In order to substantiate the above, we wish to outline a few recommendations – which should also be taken as an invitation for other scholars, policy-makers, and stakeholders to find a common solution. Our premise, in terms of suggestions, is that the path forward tackling disinformation on the EU level requires a set of complementary mechanisms that translate the normative ambition into operational capacity. First, the Union should consider establishing an entity that manages information integrity with interdisciplinary lenses. Adding to this, secondly, the EU must invest more decisively in structured digital literacy and professional training. For this, we also propose allocating cross-border scholarly funds as well so that experts and scholars could work on region-specific and sector-specific issues. Furthermore, given the generous scholarly funding schemes in the EU (such as the European Research Council grant), we also propose that projects related to disinformation and media literacy should be obliged to perform practical outputs that can be implemented on local or EU-wide layers. Lasty, we recommend – in accordance with Wardle's (2023) line of thinking – that these projects can and shall not be executed in "silos".

While the change in the framework of public discourse has placed limits on the effectiveness of legal regulation, it has not questioned its necessity. However, while regulating platforms, at least as much emphasis should be placed on promoting digital literacy; without public awareness, legal regulation of platforms and strict enforcement of rules will not be able to produce visible results. Successfully combating disinformation is a necessity for democratic societies.

Effective regulation, then, must be part of a broader societal response that involves both proactive governance and empowered citizens, recognizing that the battle against disinformation is not a finite one, but an ongoing challenge.

REFERENCES

Al-Zaman, M. S. (2021). Covid-19-related social media fake news in India. *Journalism and Media*, *2*(1), 100–114. https://doi.org/10.3390/journalmedia2010007

Argemi, M., & Fine, G. A. (2018). Faked news: The politics of rumour in British World War II propaganda. *Journal of War and Culture Studies*, *12*(2), 176–193. https://doi.org/10.1080/17526272.2018.1495905

Barragán-Romero, A. I., & Bellido-Pérez, E. (2019). Fake News durante la Primera Guerra Mundial: Estudio de su representatividad en las portadas de la prensa española (ABC Madrid). *Historia y Comunicación Social*, *24*(2), 433–447. https://doi.org/10.5209/hics.66288

Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & De Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, *14*(1), 82–96. https://doi.org/10.18352/ulr.420

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, *10*, 7. https://doi.org/10.1038/s41467-018-07761-2

Brennen, B. (2017). Making sense of lies, deceptive propaganda, and fake news. *Journal of Media Ethics*, *32*(3), 179–181. https://doi.org/10.1080/23736992.2017.1331023

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), 139–166. https://doi.org/10.1080/23808985.2024.2323736

Cauffman, C., & Goanta, C. (2021). A new order: The Digital Services Act and consumer protection. *European Journal of Risk Regulation,* 12(4), 758–774. https://doi.org/10.1017/err.2021.8

Code of practice on disinformation. (2023, September 22). Disinfocode. https://disinfocode.eu/wp-content/uploads/2023/09/code-of-practice-on-disinformation-september-22-2023.pdf.

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Tackling online disinformation: a European Approach. (COM(2018) 236) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European democracy action plan. (COM(2020) 790 final) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A790%3AFIN

Dobber, T., Ó Fathaigh, R., & Zuiderveen Borgesius, F. J. (2019): The regulation of online political micro-targeting in Europe. *Internet Policy Review*, *8*(4), https://doi.org/10.14763/2019.4.1440

Dommett, K. (2023). The 2024 election will be fought on the ground, not by AI, *Political Insight, 14*(4), 4–6. doi.org/10.1177/20419058231218316a

Dommett, K., & Bakir, M. E. (2020): A transparent digital election campaign? The insights and significance of political advertising archives for debates on electoral regulation. *Parlamentary Affairs*, *73*(1), 208–224. https://doi.org/10.1093/pa/gsaa029

Espaliú-Berdud, C. (2024). The EU Code of Practice on Disinformation. *VISUAL REVIEW International Visual Culture Review / Revista Internacional De Cultura Visual*, *16*(2), 95–109. https://doi.org/10.62161/revvisual.v16.5217

European Commission (2018). A multi-dimensional approach to disinformation, Report of the independent High level Group on fake news and online disinformation, Directorate-General for Communication Networks, Content and Technology

(2018). A multi-dimensional approach to disinformation: report of the independent High level Group on fake news and online disinformation. Publications Office. https://data.europa.eu/doi/10.2759/739290

European Commission, (2018). Recommendation of 12 September 2018 on electoral cooperation networks, online transparency, protection against cybersecurity incidents and fighting disinformation campaigns in the context of elections to the European Parliament (C(2018) 5949) https://cyberpolicy.nask.pl/wp-content/uploads/2018/09/soteu2018-cybersecurity-elections-recommendation-5949_en.pdf

European Commission: Third Meeting of the European Board for Digital Services https://digital-strategy.ec.europa.eu/en/news/third-meeting-european-board-digital-services

Fowler, E. F., Franz, M. M., & Ridout, T. N. (2022). The challange of online advertising. In *Political advertising in the United States* (2nd ed., pp. 63–75). Routledge. https://doi.org/10.4324/9781003165712-4

Fowler, E. F., Franz, M. M., & Travis N. (2022). Buying and Targeting Political Advertising. In *Political advertising in the United States* (2nd ed., pp. 93–114). Routledge. https://doi.org/10.4324/9781003165712-6

Gosztonyi, G. (2023). Content Management or Censorship? In *Censorship from Plato to Social Media* (Ser. Law, Governance and Technology Series, vol 61., pp. 7–19). Springer. https://doi.org/10.1007/978-3-031-46529-1_2

Husovec, M. (2024). The Digital Services Act's red line: What the Commission can and cannot do about disinformation. *Journal of Media Law*, *16*(1), 47–56. https://doi.org/10.1080/17577632.2024.2362483

Ilves, L. K., Evans T. J., Cilluffo F. J., & Nadeau, A. A. (2016). European Union and NATOglobal cybersecurity challenges: A way forward. *PRISM*, 6(2), 126–141. http://www.jstor.org/stable/26470452

Kirk, N., & Teeling, L. (2021). A review of political advertising online during the 2019 European Elections and establishing future regulatory requirements in Ireland, *Irish Political Studies*, *37*(1) doi.org/10.1080/07907184.2021.1907888

Leerssen., P., Dobber, T., Helberger, & N., de Vreese, C. (2021). News from the ad archive: how journalists use the Facebook Ad Library to hold online advertising accountable. *Information, Communication and Society*, *26*(7), 1381–1400. https://doi.org/10.1080/1369118X.2021.2009002

Leiser, M. (2023). Reimagining digital governance: The EU's Digital Service Act and the fight against disinformation. http://dx.doi.org/10.2139/ssrn.4427493 7

Nannini, L., Bonel, E., Bassi, D., & Maggini, M. J. (2025). Beyond phase-in: assessing impacts on disinformation of the EU Digital Services Act. *AI and Ethics, 5.* https://doi.org/10.1007/s43681-024-00467-w

Pagoto, S. L., Palmer, L., & Horwitz-Willis, N. (2023). The Next Infodemic: Abortion Misinformation. *Journal of Medical Internet Research*, *25*, e42582. https://doi.org/10.2196/42582

Palomino-Flores, P., Cristi-López, R., & Paul, D. (2024). Unraveling the truth: investigating the spread of fake news on Facebook during the COVID-19 Crisis. In Ibáñez, D.B., Castro, L.M., Espinosa, A., Puentes-Rivera, I., & López-López, P.C. (eds), *Communication and Applied Technologies. ICOMTA 2023.* (Ser. Smart Innovation, Systems and Technologies, vol 375, pp. 223–233). https://doi.org/10.1007/978-981-99-7210-4_21

Proposal for a Regulation of the European Parliament and of the Council on the Transparency and Targeting of Political Advertising (COM/2021/731 final) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0731

Shattock, E. (2021). Self-regulation 2:0? A critical reflection of the European fight against disinformation. *Harvard Kennedy School Misinformation Review*, *2*(3). https://doi.org/10.37016/mr-2020-73

Shrestha, A., & Spezzano, F. (2021). Characterizing and predicting fake news spreaders in social networks. *International Journal of Data Science and Analytics*, *13*(4), 385–398. https://doi.org/10.1007/s41060-021-00291-z

Strowel, A., De Meyere, J. (2023). The Digital Services Act: transparency as an efficient tool to curb the spread of disinformation on online platforms? *Journal of Intellectual Property, Information Technology and Electronic Commerce Law, 14*(1), 66–82.

The Sounding Board's Unanimous Final Opinion on the so-called Code of Practice. (2018). https://www.euractiv.com/wpcontent/uploads/sites/2/2018/10/3OpinionoftheSoundingboard-1.pdf.

Török, B. (2024). Free speech principles to consider when restricting disinformation. *Információs Társadalom, 24(*2), 115–128. https://doi.org/10.22503/inftars.XXIV.2024.2.7

Van Duyn, E., & Collier, J. (2018). Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication & Society*, *22*(1), 29–48. https://doi.org/10.1080/15205436.2018.1511807

Vicente, D. M. (2023). Protection against disinformation on the internet: A Portuguese perspective. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law 14*(3), 453–461.

Warde, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe report DGI(2017)09.

Wardle, C. (2017). Fake news. It's complicated. January 27, 2019 from https://fi rstdraftnews. com/fake-news-complicated/.

Wardle, C. (2023). Misunderstanding misinformation. *Issues in Science and Technology*, *39*(3), 38–40. https://doi.org/10.58875/ZAUD1691

Wasilewski, K. (2021). Fake news and the Europeanization of cyberspace. *Polish Political Science Yearbook, 50*, 61–80. https://doi.org/10.15804/ppsy202153

Wood, A. K. (2020). Facilitating Accountability for Online Political Advertisements, *Ohio State Technology Law Journal, 16*(2), 520–557.

ORIGINAL ARTICLE

DOI: 10.51480/1899-5101.18.3(41).826

Unveiling the Many Faces of Fact-Checking: State of the Art of Academic Research on Information Correction.

Maia Klaassen

0009-0009-2219-9691
University of Tartu, Estonia

Marju Himma

0000-0001-9441-7162
University of Tartu, Estonia

Ianis Bucholtz

0000-0001-6538-5245
Vidzeme University of Applied Sciences, Latvia

Krista Lepik-Verliin

0000-0002-6576-4429
University of Tartu, Estonia

Abstract: Research on fact-checking, which began gaining traction around 2010, has seen a significant surge in interest, complicating efforts to overview the field comprehensively. Previous mappings of fact-checking research were often limited in scope and accessibility and lacked context in the post-infodemic era. Addressing these gaps, a systematic literature review was conducted on English-language academic publications from 2010 to 2023, encompassing 675 articles. This review delved into various aspects of fact-checking, such as infrastructure, practices, tools, genres, effectiveness, public perception, and the emerging role of AI. It highlighted the necessity for refined research methodologies, longitudinal effectiveness studies, and a deeper understanding of social context's impact on misinformation and fact-checking efficacy, particularly beyond the U.S. This comprehensive analysis offers invaluable insights into the evolving landscape of fact-checking and information verification.

Keywords: information verification, prebunking, debunking, fact-checking, information correction

INTRODUCTION

As fact-checking has become a central response to information disorders, academic research on this practice has grown exponentially. Scholarly publications on fact-checking began after 2010 (Lewandowsky et al., 2012; Schudson, 2013; Uscinski & Butler, 2013), with a surge post-2016 due to the US Presidential elections and Brexit. This expanded to practical issues like infrastructure and automation (Hammouchi & Ghongho, 2022; Malik et al., 2021), and human practices (Ceron et al., 2021; Dhawan et al., 2021). Audience perception studies highlight the importance of effectiveness (Hameleers, 2022; Tunçer & Tam, 2022).

According to the Reporters' Lab database, the number of fact-check outlets has increased significantly – from 11 sites in 2008 to 424 in 2022, which indicates a 3,854% increase. By 2023, the number of fact-checking organizations had dropped to 417[1]. This rapidly emerging field demands revision – a helicopter's view – on the topics and domains of academic research. There have been several funding initiatives to promote fact-checking as a tool for combating information disorder (The Fact-Checking Innovation Initiative.[2], AFP Fact Check[3], and European Digital Media Observatory Fact-Checking[4]). Systematic literature reviews (SLRs) map fact-checking from technological perspectives (Capuano et al., 2023; Sequete et al., 2020), emphasizing the need for MILs to build an informed public (Onifade, 2023; Dierickx et al., 2023). Yet, all this expanding body of work lacks systematic integration: studies vary in focus, terminology, and disciplinary grounding, making it difficult to grasp the field's conceptual coherence and cumulative knowledge.

The aim of this study is to chart the evolving landscape of fact-checking research and to uncover how academia has responded to the surge of misinformation and disinformation since 2010. Through a systematic review of 675 peer-reviewed publications, the study maps the main topics, domains, and conceptual frameworks shaping the scholarly understanding of fact-checking as both a journalistic and socio-technical practice.

While we acknowledge that in a rapidly emerging field such as fact-checking, many valuable insights appear in reports, policy briefs, and popular science materials published by fact-checking organizations and expert communities, these outputs are not always peer-reviewed and therefore vary in methodological transparency and reliability. To ensure the analytical consistency and scientific rigor of this systematic literature review, we limited our corpus to peer-reviewed academic publications indexed in major scholarly databases. This approach

---

[1] https://reporterslab.org/category/fact-checking/
[2] https://www.poynter.org/innovation-initiative/
[3] https://factcheck.afp.com/
[4] https://edmo.eu/fact-checking/

enables a critical assessment of how fact-checking is conceptualized and studied within academic scholarship, while recognizing that non-academic sources represent a dynamic and practice-oriented complement to this body of research.

Stemming from the problem-setting, we approach the empirical material with the following Research Question (RQ):

- RQ: What research topics and topical domains are studied in fact-checking research?

METHOD

Our SLR approach follows Kapantai et al. (2021) and Behutiye et al. (2017), based on Kitchenham and Charters (2007). We collected academic literature from Web of Science, EBSCO, and Scopus, focusing on peer-reviewed journal articles to ensure quality. The search terms included: ((“fact-check*” OR debunk* OR prebunk*) AND (journalism OR media OR disinformation OR misinformation OR user* OR tools OR automat* OR “fake news” OR “news verification”)) OR “fact-check* epistemolog*”. We limited the search to 2010–2023, English language, and journal or critical review publications.

Searching only for “fact-checking” and “journalism” or “media” will return thousands of hits. Therefore, it is necessary to set some limitations: publication dates between 2010–2023, English language, and journal or critical review publication type. Similarly to Nieminen and Rapeli (2019), we do not question the scientific quality of the excluded contributions or contributions in languages other than English, but we chose these parameters for the sake of systematic selection. Conducted in April 2023, the search included all entries until that date.

This query brought 1116 results on Web of Science, 3803 on EBSCO, and 1054 on Scopus. 4616 articles were uploaded to Rayyan, a customizable online software for conducting SLRs that enables easy literature screening in a team. Out of that initial sample, 49 duplicates were detected and deleted by the software automatically, and 2867 duplicates were inspected manually. Out of the remaining 1749 articles, there were 694 inclusions, 25 maybes, and 1031 exclusions in the first screening round out of three.

The three screening rounds were a part of the study protocol to ensure impartiality and intercoder reliability, which we valued “not as a measure of ‘objectivity’ but as a means of reflexively improving the analysis by provoking dialogue between researchers” (Campbell et al., 2013; Hruschka et al., 2004; O’Connor & Joffe, 2020). After the initial data purification was done by Coder 1, the second round was done by Coder 2, designed both as a control mechanism and another revenue for the inductive labeling of themes and topics. To arrive at a consensus

on the conflicting 'maybe' cases, Coder 1 and Coder 2 had a revision meeting: 675 articles were included in the sample, and 1074 articles were excluded.

Exclusions were based on focus (N=248), publication type (N=46), language (N=16), missing abstracts (N=6), low research design quality (N=6), and ethical considerations regarding academic freedom (N=8). We also excluded articles where the focus was not on fact-checking per se but which suggested fact-checking in their abstracts as a possible remedy without going into more detail (N=264). Other co-authors decided on the conflicting decisions remaining, Coder 3 (N=1) and Coder 4 (N=1).

After the exclusions, we juxtaposed and compared the themes and research categorization criteria. Applying logical processing, we introduced our topical map of research on fact-checking. Coder 1 and Coder 2 developed initial coding schemes inductively and coded the sample separately. After consultation with co-authors, the codes were re-checked to avoid potential connotational misunderstandings. Coder 1 and Coder 2 held meetings throughout the coding process, discussed and rephrased codes marked with different words but with the same meaning to both, and discussed connotational misunderstandings until there was no more conflict of opinion. The results were then combined into a comprehensive overview. It is important to note that many articles are marked with multiple labels, so the sum of the articles in the groups exceeds that of the total articles in the sample.

FINDINGS

The conceptualization of fact-checking is explored in 52 articles, focusing on disinformation mapping (n=27), fact-checking evolution (n=12), and epistemology (n=13). Mapping includes SLR articles and topical studies (Bekoulis et al., 2021; Caled & Silva, 2022; Shahzad & Khan, 2022). Core aspects are discussed in 335 articles, covering effectiveness (n=221), pre-bunking (n=13), correcting falsehoods (n=113), and flagging falsehoods (n=22). Fact-checking is seen as a public good (Berghel, 2017), conformation journalism (Lotero-Echeverri et al., 2018), and crisis communication (Krause et al., 2020; Tian & Ding, 2019; Wang et al., 2021). Yu et al. (2023) note interdisciplinary research on misinformation sharing.

Fact-checking domains are covered in 266 articles, including COVID-19 (n=145), politics (n=82), medicine (n=21), climate change (n=9), and armed conflicts (n=7). Regional studies are in 302 articles, focusing on the U.S. (n=61), Spain (n=38), Brazil (n=18), the UK (n=17), China (n=17), and India (n=12). Professional practices are examined in 145 articles, including fact-checkers' practices (n=70), journalists' verification (n=46), infrastructure (n=16), and perceptions (n=14).

Fact-checking and debunking provide a roadmap for combating information disorders (Eysenbach, 2020; Lewandowsky, 2021). Contexts are discussed in 250 articles, including social media (n=103), scientific fact-checking (n=100), and experiments (n=77). A coordinated approach is needed, combining authorities, academics, and fact-checkers (Furnémont & Rokša-Zubčević, 2021; Durach et al., 2020). Fact-checking outside media is analyzed in 247 articles, focusing on audience perceptions (n=107), human practices (n=73), and media literacy (n=59).

Naeem and Bhatti (2020) highlight librarians' roles, while Hoque et al. (2021) call for socio-technical solutions to disinformation. New approaches are explored in 151 articles, covering tools (n=94), automated fact-checking (n=50), and new technologies (n=9). European actions against disinformation are also described (García et al., 2019).

FACT-CHECKING AS A JOURNALISTIC PRACTICE

The driving force of fact-checking relates to the professional motives of journalists (Graves et al., 2016). However, professional ideals may clash with everyday practice: while journalists emphasize the importance of checking information, they sometimes publish information that turns out to be imprecise (Dumitru, 2021). Journalists do not have consistent routines for confirming information from social media and blogs, and significant uncertainty exists among them regarding the feasibility of verifying such content (Edwardsson et al., 2021). Thomson et al. (2022) stress that today, the naked eye cannot necessarily discern visual manipulations, and source credibility is also not enough because reputable journalists are also known to share manipulated or inaccurate user-generated information.

Deficiencies in journalistic verification can be observed when covering secluded, faraway countries to which journalists do not have direct access; in these cases, errors cascade from one media outlet to the next, and even after more information has become available, journalists do not necessarily feel the need to issue a correction because of the obscure nature of the subject (Seo, 2022). Journalists advocate for fellow professionals to adopt more stringent fact-checking methods, especially when journalism's watchdog role is essential (Shapals, 2018). Of the three strategies Sarelska and Jenkins (2022) have identified in how journalists responded to COVID-19 falsehoods, only the fact-checking approach directly challenges disinformation. On the other hand, a study shows that journalists employ specific combinations of traditional fact-checking skills, and in most cases, this is enough to adequately examine information (Himma-Kadakas & Ojamets, 2022).

Online disinformation has encouraged a shift in journalism practice from competition to collaboration (Graves & Konieczna, 2015; Smyrnaios et al., 2019).

Micallef et al. (2022) found that fact-checkers are overwhelmed as the work is labor intensive and feel they are failing to fulfill journalistic duties because their impact and reach are limited (Singer, 2020).

FACT-CHECKING AS A NEW GENRE

Studies reveal the emergence of fact-checking organizations globally, driven by various forces. Fact-checking Africa, for instance, targets news from legacy media (Cheruiyot & Ferrer-Conill, 2018). In Tunisia, initiatives lacked editorial guidance and training (Zammit et al., 2020). Tsek.ph in the Philippines was established by universities and newsrooms for elections (Chua & Soriano, 2020). Spain's public service media set up units responding to COVID-19 misinformation (Palomo & Sedano, 2021; Ufarte-Ruiz et al., 2020). StopFake in Ukraine counters Russian propaganda (Haigh et al., 2018). Verificado 2018 in Mexico was a collaborative effort before the elections (Martinez-Carrillo & Tamul, 2019). Critical factors for these initiatives include elections, foreign interference, and the pandemic. Amazeen (2018) discusses a global movement arising from public disempowerment, journalism's downturn, technological change, and socio-political discord.

Lelo (2022) highlights economic and editorial influences on fact-checkers, noting similarities to journalism. Ownership structures and business models vary, with differences between media group operations and journalist-led initiatives (Esteban-Navarro et al., 2021). Fact-checkers need skills in technology, visualization, and databases (Ruiz et al., 2021). Amazeen (2015) reports a consensus on information selection and verification, with fact-checkers viewing their work as more trustworthy and educational (Singer, 2020). Juneström (2021) identifies accessibility, trust-building, and pedagogy as crucial aspects. Kim and Buzzelli (2022) suggest an emerging institutional form of fact-checking.

Palau-Sampio (2018) finds differences in Latin American fact-checking practices, with room for improvement in sourcing and expert insight. Lim (2018) notes that ambiguous political language influences claim selection and judgment variability. Schudson (2013) emphasizes adherence to journalism standards, with differences in transparency and source use (Vu et al., 2022). Rodríguez-Pérez et al. (2021) report that Colombian journalists see fact-checking as restoring credibility, while in authoritarian Arab countries, fact-checkers practice self-censorship (Fakida, 2021). Jo et al. (2022) explore public funding for fact-checking in South Korea. Horowitz et al. (2021) discuss the public service media's role in countering misinformation. Social media interventions by news media are effective against health misinformation, though "truth scales" and partial refutations weaken this effect (Walter et al., 2020).

FACT-CHECKING ON SOCIAL MEDIA PLATFORMS

False information often spreads via social media, and platform responses are crucial. Cotter et al. (2022) found that Facebook, Twitter, and YouTube's approaches to COVID-19 misinformation were inadequate. Nissen et al. (2022) noted biases in fact-checks facilitated by Google and the Poynter Institute. Fact-checkers must adapt their work to different platforms for funding (Lelo, 2022) and debunk rival media for economic viability (Long et al., 2019), increasing administrative burdens and vulnerability for smaller organizations. Crowdsourcing is rare (Martinez-Carrillo & Tamul, 2019).

Warnings and corrective labels, or flagging, are perceived differently, and labels indicating government affiliation reduce political information sharing (Liang et al., 2022). Scharrer et al. (2022) found that warnings increased skepticism toward health misinformation. However, fact-checking labels do not alter credibility perceptions of single posts but do affect site evaluations (Oeldorf--Hirsch et al., 2020). Critical comments by users curb sharing false information more effectively than platform warnings (Colliander, 2019). Garrett and Poulsen (2019) found that neither fact-checkers nor peer warnings effectively persuaded users about political posts, but humor labels reduced reactance.

Flagging is still developing, with the potential to help navigate information environments while respecting freedom of expression (Morrow et al., 2022). Fact-checking amplification was studied from platform-specific and audience perspectives. Li and Chang (2022) identified source credibility, political views, and prior shares as factors in sharing fact-checks. People share fact-checks selectively based on political alignment (Shin & Thorson, 2017). Correction-sharing is less studied than misinformation-sharing, with insufficient research on influencing factors.

THE EFFECTIVENESS OF FACT-CHECKING

Studies on fact-checking effectiveness overlap with correcting false information, audience perceptions, and context. Bocharov and Demidov (2021) argue that fact-checking requires cultural competence and prioritizing facts. Effectiveness is limited by ideological divisions and audience reach (Robertson et al., 2020). Trust in the government aids rumor-disproving in China (Gao et al., 2022), but no significant effect was found in Germany (Radechovsky et al., 2019). Fact-checking in Zimbabwe, Namibia, and South Africa needs to be combined with media literacy and public health information (Mare & Munoriyarwa, 2022).

Confirmation bias and motivated reasoning affect fact-checking. Nyhan and Reifler (2010) noted the "backfire effect," but Wood and Porter (2018) and

Swire-Thompson et al. (2020) found little evidence for it. Cognitive processes in assessing truthfulness are crucial (Lewandowsky et al., 2012). Even if fact-checking stops people from believing false claims, it may not be able to change the attitudes toward the political figure that made them (Nyhan et al., 2020). Similarly, it has been shown to improve attitudes toward immigrants without affecting policy preferences (Choi et al., 2023). Dan and Dixon (2021) suggest targeting societal outcomes to reduce polarization.

Li and Wagner (2020) identify three audience types: uninformed, ambiguous, and misinformed, with only the uninformed likely to change their beliefs. Biltoft (2020) highlights gullibility and refusal to accept facts. Risk factors are over-confidence and distinguishing facts from opinions (Martinez-Costa et al., 2022; Merpert et al., 2018). Belief in others' susceptibility to misinformation increases verification efforts (Sun, 2022). Warnings about reputational risks have limited effect (Nyhan & Reifler, 2015; Ma et al., 2022).

Timing in fact-checking is critical. Early evidence publication and rumor detection support timely interventions (Allein et al., 2021; Liu & Rong, 2022; Ma et al., 2020). Prebunking and debunking effectiveness are debated (Tay et al., 2022; van der Linden et al., 2021; Arcos et al., 2022).

Innovative formats like cartoon characters and audio dramas enhance fact-checking transmission (Opgenhaffen, 2021; Winters et al., 2021). Fact-checking's impact depends on audience alignment, consistency, timing, and source credibility. Corrections do not fully eliminate misinformation, and beliefs do not entirely revert; political, marketing, and health topics resist corrections (Walter et al., 2020).

RETHINKING THE FUTURE OF FACT-CHECKS

Fact-checking is a professional and human practice of information verification. Studies combining fact-checking with Media and Information Literacies (MILs) emphasize individual responsibility for debunking false information with the most effective real-time corrections (Kim & Walker, 2020; Musi & Reed, 2022). Social learning models, crowdsourcing, and critical thinking are also explored (Sikder et al., 2020; Soprano et al., 2021; Stokes-Parish, 2022). Experiments show MIL interventions improve distinguishing evidence-based information from misinformation (Hameleers, 2022; Kuś & Barczyszyn-Madziarz, 2020). New media literacy influences media trust and fact-checking motivation (Lee et al., 2022; Shahzad & Khan, 2022).

Critical thinking and MILs are crucial in curbing misinformation. Educational approaches include games, performance art, and mobile apps (Basol et al., 2021; França et al., 2021). Journalists and fact-checkers teach these skills in various

settings (Cunliffe-Jones, 2020). Emotional framing and humor-based corrections are studied for effectiveness (Kim & Chen, 2022; Vraga et al., 2019).

Fact-checks serve as data archives for researchers, aiding in studies on information disorders (León et al., 2022; Musi & Reed, 2022). Fact-checking organizations' websites are used as research sources (Pathak et al., 2021). Researchers incorporate fact-checking in strategies against misinformation and conduct experiments (Li & Wagner, 2020; Bekoulis et al., 2022) — fact-checking 'archives' false narratives, highlighting the importance of quality fact-checks for re-usability.

Innovating fact-checking includes online tools supporting individual fact-checking (Berghel, 2017) and APIs (Meel & Vishwakarma, 2020), machine learning (Smith & Bastian, 2022) and network models (Tambuscio et al., 2018), assessing bias (Allen et al., 2021), AI-assisted (Grmuša & Prelog, 2021), and fully automated fact-checking (Parratt-Fernandez et al., 2021). Platform-specific innovations include debunking functions on X (King et al., 2021), WhatsApp changes (Reis et al., 2020), and Instagram correction testing (Vraga et al., 2020). Fact-checking tools like Google Fact Check Tools are reviewed (Alfoldiova, 2018).

DISCUSSION

This systematic literature review (SLR) aimed to map the landscape of fact-checking research, focusing on its conceptualization, core aspects, domains, regional practices, and innovations. The findings reveal a multifaceted field with significant implications for theory, practice, and policy. The conceptualization of fact-checking highlights its role in disinformation mapping, evolution, and epistemology (Bekoulis et al., 2021; Caled & Silva, 2022; Shahzad & Khan, 2022). Core aspects such as effectiveness, pre-bunking, correcting falsehoods, and flagging falsehoods are discussed in several contexts (Berghel, 2017; Krause et al., 2020; Wang et al., 2021), aligning with existing literature that underscores the interdisciplinary nature of fact-checking research (Yu et al., 2023).

Regional studies reveal diverse fact-checking practices, emphasizing the need to assess fact-checker performance across contexts (Amazeen, 2015; Nieminen & Sankari, 2021; Uscinski & Butler, 2013), as fact-checking effectiveness varies by country and organization. Fact-checking is influenced by internal and external forces, including technology platforms (Lelo, 2022; Nissen et al., 2022). Research designs need improvement, as the experiments on fact-checking effectiveness can lack real-life context. The quality of fact-checking should also be further examined, as it serves as a study tool for MILs and an archive of false narratives, and findings indicate a high reuse rate (Kim & Walker, 2020; Musi & Reed, 2022).

A coordinated approach involving authorities, academics, and fact-checkers is necessary (Furnémont & Rokša-Zubčević, 2021; Durach et al., 2020). Innovations

in fact-checking, including online tools, machine learning, AI-assisted, and fully automated fact-checking, are explored in newer articles (Meel & Vishwakarma, 2020; Smith & Bastian, 2022; Parratt-Fernandez et al., 2021). However, there is a gap in research highlighting who, how, and how often these new tools will be used. Future research should address the need for longitudinal and qualitative studies and insights from non-Western contexts.

The SLR methodology provides a comprehensive overview of the field but has limitations. The exclusion of non-English articles and the focus on peer-reviewed journals may have omitted relevant studies. Additionally, the rapid evolution of fact-checking practices means that some recent developments may not be fully captured. It should be noted that a large body of work only mentions fact-checking in passing as a buzzword while offering few new insights into the practice or genre itself.

CONCLUSIONS

This systematic literature review provides a comprehensive mapping of academic research on fact-checking as a journalistic, technological, and socio-cultural practice. We analyzed 675 peer-reviewed publications across three major databases. The results reveal that fact-checking has evolved from a journalistic activity into a multifaceted field intersecting communication, technology, political science, and psychology. Research has expanded from conceptual and epistemological debates to applied domains such as health communication, political campaigning, crisis communication, and the COVID-19 infodemic. Thematically, fact-checking studies focus on effectiveness, correction and pre-bunking strategies, professional practices, and audience perceptions, while new trends emerge in automation, artificial intelligence, and platform-specific verification tools.

These findings underscore that fact-checking is no longer solely a journalistic correction mechanism but a key component of contemporary information governance and media literacy. Understanding how fact-checking operates across domains helps researchers and practitioners to evaluate its societal impact, inform policy initiatives, and design more effective interventions to mitigate misinformation. Mapping the academic landscape also exposes the unevenness of global research – most studies originate from Western contexts – and the dominance of experimental and survey-based approaches, which limits understanding of cultural, institutional, and ethical dimensions of fact-checking.

While this review focuses on peer-reviewed academic literature to ensure analytical rigor and comparability, we acknowledge that a large and influential body of knowledge also exists outside academia in the form of reports, policy briefs, and popular science materials published by fact-checking organizations

and expert communities. Future research should systematically analyze these sources to capture the practical evolution of fact-checking, its policy relevance, and its educational potential within Media and Information Literacy frameworks. Combining academic and practitioner perspectives will offer a more holistic understanding of fact-checking as both a scholarly field and a professional practice essential to democratic information ecosystems.

Acknowledgments and funding: The design, implementation, and publishing of this research was a part of the EU-funded project BECID (ID 118471).

REFERENCES

Alfoldiova, A. (2018). Google is Introducing a New Feature for Fact Checking. *Media Literacy And Academic Research*, *1*(2), 78–79. https://www.ceeol.com/search/article-detail?id=730595

Allein, L., Augenstein, I., & Moens, M.-F. (2021). Time-aware evidence ranking for fact-checking. *Journal of Web Semantics*, *71*. https://doi.org/10.1016/j.websem.2021.100663

Amazeen, M. (2018). Practitioner perceptions: Critical junctures and the global emergence and challenges of fact-checking. *International Communication Gazette*, *81*(6), 541–561. https://doi.org/10.1177/1748048518817674

Amazeen, M. A. (2015). Checking the Fact-Checkers in 2008: Predicting Political Ad Scrutiny and Assessing Consistency. *Journal of Political Marketing*, *15*(4), 433–464. https://doi.org/10.1080/15377857.2014.959691

Arcos, R., Gertrudix, M., Arribas, C. & Cardarilli, M. (2022). Responses To Digital Disinformation As Part Of Hybrid Threats: A Systematic Review On The Effects Of Disinformation And The Effectiveness Of Fact-Checking/Debunking. *Open Res Europe 2022*, 2:8 https://doi.org/10.12688/openreseurope.14088.1

Baptista, J., Jeronimo, P., Pineiro-Naval, V., & Gradim, A. (2022). Elections and fact-checking in Portugal: The case of the 2019 and 2022 legislative elections. *Profesional De La Informacion*, *31*(6). https://doi.org/10.3145/epi.2022.nov.11

Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. van der. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, *8*. https://doi.org/10.1177/20539517211013868

Behutiye, W. N., Rodríguez, P., Oivo, M., & Tosun, A. (2017). Analyzing the concept of technical debt in the context of agile software development: A Systematic Literature Review. *Information and Software Technology*, *82*, 139–158. https://doi.org/10.1016/j.infsof.2016.10.004

Bekoulis, G., Papagiannopoulou, C., & Deligiannis, N. (2022). A Review on Fact Extraction and Verification. *ACM Computing Surveys*, *54*(1), 1–35. https://doi.org/10.1145/3485127

Biltoft, C. N. (2020). The anatomy of credulity and incredulity: A hermeneutics of misinformation. *Harvard Kennedy School Misinformation Review*, *1*(2). https://doi.org/10.37016/mr-2020-016

Bocharov, A. B., & Demidov, M. O. (2021). Fact-checking Technology in the Fight Against 'Information Debris': Problems and Prospects. *Управленческое Консультирование*, *12*, 102–111. https://doi.org/10.22394/1726-1139-2020-12-102-111

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods & Research*, *42*(3). https://doi.org/10.1177/0049124113500475

Ceron, W., Gruszynski Sanseverino, G., de-Lima-Santos, M.-F., & Quiles, M. G. (2021). COVID-19 fake news diffusion across Latin America. *Social Network Analysis and Mining*, *11*(1), 47. https://doi.org/10.1007/s13278-021-00753-z

Cheruiyot, D., & Ferrer-Conill, R. (2018). 'FACT-CHECKING AFRICA' Epistemologies, data and the expansion of journalistic discourse. *Digital Journalism*, *6*(8), 964–975. https://doi.org/10.1080/21670811.2018.1493940

Chua, Y., & Soriano, J. (2020). Electoral Disinformation: Looking Through the Lens of Tsek.ph Fact Checks. *Plaridel*, *17*(1), 285–295. https://www.plarideljournal.org/wp-content/uploads/2020/06/2020-01-Chua-and-Soriano.pdf

Choi, D., Oh, H., Chun, S., Kwon, T., & Han, J. (2023). Preventing rumor spread with deep learning. Expert Systems with Applications, 197. https://doi.org/10.1016/j.eswa.2022.116688

Colliander, J. (2019). 'This is fake news': Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media. *Computers in Human Behavior*, *97*, 202–215. https://doi.org/10.1016/j.chb.2019.03.032

Cotter, K., DeCook, J., & Kanthawala, S. (2022). Fact-Checking the Crisis: COVID-19, Infodemics, and the Platformization of Truth. *Social Media + Society*, *8*(1). https://doi.org/10.1177/20563051211069048

Cunliffe-Jones, P. (2020). From Church and Mosque to WhatsApp-Africa Check's Holistic Approach to Countering 'Fake News.' *Political Quarterly*, *91*(3), 596–599. https://doi.org/10.1111/1467-923X.12899

Dan, V., & Dixon, G. N. (2021). Fighting the Infodemic on Two Fronts: Reducing False Beliefs Without Increasing Polarization. *Science Communication*, *43*(5), 674–682. https://doi.org/10.1177/10755470211020411

Dhawan, D., Bekalu, M., Pinnamaneni, R., McCloud, R., & Viswanath, K. (2021). COVID-19 News and Misinformation: Do They Matter for Public Health Prevention? *Journal of Health Communication*, *26*(11), 799–808. https://doi.org/10.1080/10810730.2021.2010841

Dumitru, E.-A. (2021). Is 'Letting the Truth Get in the Way of a Good Story' Enough? Journalists' Perception on Fake News. *Journal of Media Research*, *14*(3), 63–79. https://www.ceeol.com/search/article-detail?id=1000916

Durach, F., Bargaoanu, A., & Nastasiu, C. (2020). Tackling Disinformation: EU Regulation of the Digital Space. *Romanian Journal Of European Affairs*, *20*(1), 5–20. https://www.ceeol.com/search/article-detail?id=859431

Edwardsson, M., Al-Saqaf, W., & Nygren, G. (2021). Verification of Digital Sources in Swedish Newsrooms—A Technical Issue or a Question of Newsroom Culture? *Journalism Practice*. https://doi.org/10.1080/17512786.2021.2004200

Esteban-Navarro, M., Nogales-Bocio, A., Garcia-Madurga, M., & Morte-Nadal, T. (2021). Spanish Fact-Checking Services: An Approach to Their Business Models. *Publications*, *9*(3). https://doi.org/10.3390/publications9030038

Eysenbach, G. (2020). How to fight an infodemic: The four pillars of infodemic management. *Journal of Medical Internet Research*, *22*(6). https://doi.org/10.2196/21820

Fakida, A. (2021). Political fact-checking in the Middle East: What news can be verified in the Arab world? *Open Information Science*, *5*(1), 124–139. https://doi.org/10.1515/opis-2020-0117

Fischer, F., Carow, F., & Gillitzer, S. (2021). Humor and Fear—Two Sides of the Same Coin?: Experimental Evidence on Humor Appeals in Health Communication Related to Childhood Vaccination. *Frontiers in Public Health*, *9*, 649507. https://doi.org/10.3389/fpubh.2021.649507

França, G. S., Ribeiro, R. C., Soares, L. R., Calmoni, J., França, G. B. de, & Brito, P. E. (2021). The Flat Earth satire: Using science theater to debunk absurd theories. *Geoscience Communication*, *4*, 297–301. https://doi.org/10.5194/gc-4-297-2021

Gao, H., Guo, D., Yin, H., Wu, J., Cao, Z., & Li, L. (2022). Strategies and effectiveness of the Chinese government debunking COVID-19 rumors on Sina Weibo: Evaluating from emotions. *Journal Of Applied Communication Research*, *50*(6), 632–654. https://doi.org/10.1080/00909882.2022.2144409

García, X., Vizoso, Á., & Pérez-Seijo, S. (2019). Verification initiatives in the scenario of misinformation. Actants for integrated plans with multi-level strategies. *Brazilian Journalism Research*, *15*, 614–635. https://doi.org/10.25200/BJR.v15n3.2019.1215

García-Marín, D. (2022). Viralizing the truth: Predictive factors of fact-checkers' engagement on TikTok. *Profesional de la Informacion*, *31*(2). https://doi.org/10.3145/epi.2022.mar.10

Garrett, R., & Poulsen, S. (2019). Flagging Facebook Falsehoods: Self-Identified Humor Warnings Outperform Fact Checker and Peer Warnings. *Journal Of Computer-Mediated Communication*, *24*(5), 240–258. https://doi.org/10.1093/jcmc/zmz012

Gertrudix, M., Arcos, R., Cardarilli, M., & Arribas, C. (2022). Responses to digital disinformation as part of hybrid threats: A systematic review on the effects of disinformation and the effectiveness of fact-checking/debunking [version 1; peer review: 2 approved]. *Open Research Europe*, *2*. https://doi.org/10.12688/openreseurope.14088.1

González, M. S., Gonzales, H. M. S., & Moreno, J. M. (2022). Editorial innovation on social networks of the #CoronavirusFactCheck Alliance Hispanic verifiers: Contents and vision of those responsible. *Revista Latina de Comunicacion Social*, 2022(80), 135–161. https://doi.org/10.4185/RLCS-2022-1535

Graves, L., & Konieczna, M. (2015). Sharing the News: Journalistic Collaboration as Field Repair. *International Journal Of Communication*, *9*, 1966–1984. https://ijoc.org/index.php/ijoc/article/view/3381

Graves, L., Nyhan, B., & Reifler, J. (2016). Understanding Innovations in Journalistic Practice: A Field Experiment Examining Motivations for Fact-Checking. *Journal Of Communication*, *66*(1), 102–138. https://doi.org/10.1111/jcom.12198

Grmuša, T., & Prelog, L. (2021). The Role of New Technologies in Combatting Fake News –Experiences and Challenges of Croatian Media Organisations. *Medijske Studije*, *11*(22), 62–80. https://doi.org/10.20901/ms.11.22.4

Haigh, M., Haigh, T., & Kozak, N. (2018). Stopping Fake News: The work practices of peer-to-peer counter propaganda. *Journalism Studies*, *19*(14), 2062–2087. https://doi.org/10.1080/1461670X.2017.1316681

Hameleers, M. (2022). Separating Truth From Lies: Comparing The Effects Of News Media Literacy Interventions And Fact-Checkers In Response To Political Misinformation In The US And Netherlands. *Information, Communication & Society*, *25*(1), 110–126. https://doi.org/10.1080/1369118X.2020.1764603

Hammouchi, H., & Ghogho, M. (2022). Evidence-Aware Multilingual Fake News Detection. *IEEE Access*, *10*, 116808–116818. https://doi.org/10.1109/ACCESS.2022.3220690

Himma-Kadakas, M., & Ojamets, I. (2022). Debunking False Information: Investigating Journalists' Fact--Checking Skills. *Digital Journalism*, *10*(5), 866–887. https://doi.org/10.1080/21670811.2022.2043173

Hoque, M., Ferdous, M., Khan, M., & Tarkoma, S. (2021). Real, Forged or Deep Fake? Enabling the Ground Truth on the Internet. *IEEE Access*, *9*, 160471–160484. https://doi.org/10.1109/ACCESS.2021.3131517

Hornoiu, D. (2019). Some considerations on fake news detection. *Analele Universitatii Ovidius Constanta, Seria Filologie*, *30*(2), 336–344. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090885811&partnerID=40&md5=f658ad53f3fda74c96f5f569c2f47e21

Horowitz, M., Cushion, S., Dragomir, M., Manjon, S., & Pantti, M. (2022). A Framework for Assessing the Role of Public Service Media Organizations in Countering Disinformation. *Digital Journalism*, *10*(5), 843–865. https://doi.org/10.1080/21670811.2021.1987948

Hruschka, D. J., Schwartz, D., St.John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research. *Field Methods*, *16*(3), 307–331. https://doi.org/10.1177/1525822X04266540

Hylton-Kong, T., Steiner, M. J., Bailey, A., Palazzi, M., & Gallo, M. F. (2021). Debunking myths about contraceptive safety among women in Kingston, Jamaica: Pilot randomized controlled trial. *Contraception*, *103*(5), 356–360.

Jiang, X., & Zhang, J. (2016). A text visualization method for cross-domain research topic mining. *Journal of Visualization*, *19*, 561–576. https://doi.org/10.1007/s12650-015-0323-9

Jo, H., Park, S., Shin, D., Shin, J., & Lee, C. (2022). Estimating cost of fighting against fake news during catastrophic situations. *Telematics And Informatics*, *66*. https://doi.org/10.1016/j.tele.2021.101734

Juneström, A. (2021). An emerging genre of contemporary fact-checking. *Journal of Documentation*, *77*(2), 501–517. https://www.emerald.com/insight/content/doi/10.1108/JD-05-2020-0083/full/html

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A Systematic Literature Review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, 23(5), 1301–1326. https://doi.org/10.1177/1461444820959296

Katsaounidou, A., Vrysis, L., Kotsakis, R., Dimoulas, C., & Veglis, A. (2019). MAthE the Game: A Serious Game for Education and Training in News Verification. *Education Sciences*, *9*(2), 155–155. https://doi.org/10.3390/educsci9020155

Kim, S., & Chen, K. (2022). The use of emotions in conspiracy and debunking videos to engage publics on YouTube. *New Media & Society*. https://doi.org/10.1177/14614448221105877

King, K., Wang, B., Escobari, D., & Oraby, T. (2021). Dynamic Effects of Falsehoods and Corrections on Social Media: A Theoretical Modeling and Empirical Evidence. *Journal Of Management Information Systems*, *38*(4), 989–1010. https://doi.org/10.1080/07421222.2021.1990611

Kitchenham, B., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in software engineering. *Technical report*, EBSE Technical Report EBSE-2007-01.

Krause, N. M., Freiling, I., Beets, B., & Brossard, D. (2020). Fact-checking as risk communication: The multi-layered risk of misinformation in times of COVID-19. *Journal of Risk Research*, *23*(7), 1052–1059. https://doi.org/10.1080/13669877.2020.1756385

Kuś, M., & Barczyszyn-Madziarz, P. (2020). Fact-checking initiatives as promoters of media and information literacy: The case of Poland. *Central European Journal of Communication*, *13*, 249–265.

Lamprou, E., Antonopoulos, N., Anomeritou, I., & Apostolou, C. (2021). Characteristics of Fake News and Misinformation in Greece: The Rise of New Crowdsourcing-Based Journalistic Fact-Checking Models. *Journalism and Media*, *2*(25), 417–439. https://doi.org/10.3390/journalmedia2030025

Lee, E., Lee, T., & Lee, B. (2022). Understanding the role of new media literacy in the diffusion of unverified information during the COVID-19 pandemic. *New Media & Society*. https://doi.org/10.1177/14614448221130955

Lelo, T. (2022). The Rise of the Brazilian Fact-checking Movement: Between Economic Sustainability and Editorial Independence. *Journalism Studies*, *23*(9), 1077–1095. https://doi.org/10.1080/1461670X.2022.2069588

Lelo, T. V. (2022). Truth and hegemony in the journalistic strategies to deter disinformation. *Comunicacao Midia e Consumo*, *19*(55), 250–269. https://doi.org/10.18568/cmc.v19i55.2637

León, B. L., Martínez-Costa, M.-P., Salaverría, R., López-Goñi, I. (2022). Health and science-related disinformation on COVID-19: A content analysis of hoaxes identified by fact-checkers in Spain. *PLoS ONE*, 17(4). https://doi.org/10.1371/journal.pone.0265995

Lewandowsky, S. (2021). Climate Change Disinformation and How to Combat It. *Annual Review Of Public Health*, 42, 1–21. https://doi.org/10.1146/annurev-publhealth-090419-102409

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Liang, F., Zhu, Q., & Li, G. (2022). The Effects of Flagging Propaganda Sources on News Sharing: Quasi-Experimental Evidence from Twitter. *International Journal Of Press-Politics*. https://doi.org/10.1177/19401612221086905

Linder, R., Mohseni, S., Yang, F., Pentyala, S. K., Ragan, E. D., & Hu, X. B. (2021). How level of explanation detail affects human performance in interpretable intelligent systems: A study on explainable fact checking. *Applied AI Letters*, 2(4). https://doi.org/10.1002/ail2.49

Liu, M., & Rong, L. (2022). An online multi-dimensional opinion dynamic model with misinformation diffusion in emergency events. *Journal of Information Science*, *48*(5), 640–659. https://doi.org/10.1177/0165551520977430

Long, N., Richardson, M., & Stahler, F. (2019). Media, fake news, and debunking. *Economic Record*, *95*(310), 312–324. https://doi.org/10.1111/1475-4932.12487

Lotero-Echeverri, G., Romero-Rodríguez, L. M., & Pérez-Rodríguez, M. A. (2018). Fact-checking vs. Fake news: Confirmation journalism as a tool of media literacy against misinformation. *Index Comunicación*, *8*(2), 295–316.

Ma, J., Gao, W., Joty, S., & Wong, K. (2020). An Attention-based Rumor Detection Model with Tree-structured Recursive Neural Networks. *Acm Transactions On Intelligent Systems AndTechnology*, *11*(4). https://doi.org/10.1145/3391250

Ma, S., Bergan, D., Ahn, S., Carnahan, D., Gimby, N., McGraw, J., & Virtue, I. (2022). Fact-checking as a deterrent? A conceptual replication of the influence of fact-checking on the sharing of misinformation by political elites. *Human Communication Research*. https://doi.org/10.1093/hcr/hqac031

Malik, A., Khan, M. L., & Quan-Haase, A. (2021). Public health agencies outreach through Instagram during the COVID-19 pandemic: Crisis and Emergency Risk Communication perspective. *International Journal of Disaster Risk Reduction*, *61*, 102346. https://doi.org/10.1016/j.ijdrr.2021.102346

Mare, A., & Munoriyarwa, A. (2022). Guardians of truth? Fact-checking the 'disinfodemic' in Southern Africa during the COVID-19 pandemic. *Journal Of African Media Studies*, *14*(1), 63–79. https://doi.org/10.1386/jams_00065_1

Martinez-Carrillo, N., & Tamul, D. (2019). (Re)constructing Professional Journalistic Practice in Mexico: Verificado's Marketing of Legitimacy, Collaboration, and Pop Culture in Fact-Checking the 2018 Elections. *International Journal Of Communication*, *13*, 2596-2618. https://ijoc.org/index.php/ijoc/article/view/10933

Martinez-Costa, M., Lopez-Pan, F., Buslon, N., & Salaverria, R. (2022). Nobody-fools-me perception: Influence of Age and Education on Overconfidence About Spotting Disinformation. *Journalism Practice*. https://doi.org/10.1080/17512786.2022.2135128

Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, *153*, 112986. https://doi.org/10.1016/j.eswa.2019.112986

Merpert, A., Furman, M., Anauati, M. V., Zommer, L., & Taylor, I. (2018). Is That Even Checkable? An Experimental Study in Identifying Checkable Statements in Political Discourse. *Communication Research Reports*, *35*(1), 48–57. https://doi.org/10.1080/08824096.2017.1366303

Musi, E., & Reed, C. (2022). From fallacies to semi-fake news: Improving the identification of misinformation triggers across digital media. *Discourse & Society*, *33*(3), 349–370. https://doi.org/10.1177/09579265221076609

Nabożny, A., Balcerzak, B., Morzy, M., Wierzbicki, A., Savov, P., & Warpechowski, K. (2022). Improving medical experts' efficiency of misinformation detection: An exploratory study. *World Wide Web: Internet and Web Information Systems*, 1–26. https://doi.org/10.1007/s11280-022-01084-5

Naeem, S. B., & Bhatti, R. (2020). The Covid-19 'infodemic': A new front for information professionals. *Health Information and Libraries Journal*, *37*(3), 233–239. https://doi.org/10.1111/hir.12311

Nieminen, S., & Rapeli, L. (2019). Fighting misperceptions and doubting journalists' objectivity: A review of fact-checking literature. *Political Studies Review*, *17*(3), 296–309. https://doi.org/10.1177/1478929918786852

Nieminen, S., & Sankari, V. (2021). Checking PolitiFact's Fact-Checks. *Journalism Studies*, *22*(3), 358–378. https://doi.org/10.1080/1461670X.2021.1873818

Nissen, I., Walter, J., Charquero-Ballester, M., & Bechmann, A. (2022). Digital Infrastructures of COVID-19 Misinformation: A New Conceptual and Analytical Perspective on Fact-Checking. *Digital Journalism*, *10*(5), 738–760. https://doi.org/10.1080/21670811.2022.2026795

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Nyhan, B., & Reifler, J. (2015). The Effect of Fact-Checking on Elites: A Field Experiment on U.S. State Legislators. *American Journal of Political Science*, *59*(3), 628–640. https://doi.org/10.1111/ajps.12162

Nyhan, B., Porter, E., Reifler, J., & Wood, T. (2020). Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Political Behavior*, *42*(3), 939–960. https://doi.org/10.1007/s11109-019-09528-x

O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, *19*. https://doi.org/10.1177/1609406919899220

Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A., & Boyle, M. (2020). The Ineffectiveness of Fact-checking Labels on News Memes and Articles. *Mass Communication And Society*, *23*(5), 682–704. https://doi.org/10.1080/15205436.2020.1733613

Opgenhaffen, M. (2021). Fact-Checking Interventions on Social Media Using Cartoon Figures: Lessons Learned from 'the Tooties.' *Digital Journalism*, *10*(5), 888–911. https://doi.org/10.1080/21670811.2021.2011758

Palau-Sampio, D. (2018). Fact-checking and scrutiny of power: Supervision of public discourses in new media platforms from Latin America. *Communication & Society-Spain*, *31*(3), 347–365. https://doi.org/10.15581/003.31.3.347-365

Palomo, B., & Sedano, J. (2018). WhatsApp as a verification tool for fake news. The case of B de Bulo. *Revista Latina de Comunicacion Social*, *2018*(73), 1384–1397. https://doi.org/10.4185/RLCS-2018-1312

Palomo, B., & Sedano, J. (2021). Cross-Media Alliances to Stop Disinformation: A Real Solution? *Media & Communication*, *9*(1), 239–250. https://www.researchgate.net/publication/349794727_Cross-Media_Alliances_to_Stop_Disinformation_A_Real_Solution

Parratt-Fernandez, S., Mayoral-Sanchez, J., & Mera-Fernandez, M. (2021). The application of artificial intelligence to journalism: An analysis of academic production. *Profesional De La Informacion*, *30*(3). https://doi.org/10.3145/epi.2021.may.17

Pathak, A., Srihari, R., & Natu, N. (2021). Disinformation: Analysis and identification. *Computational And Mathematical Organization Theory*, *27*(3), 357–375. https://doi.org/10.1007/s10588-021-09336-x

Perez-Curiel, C., Ruas-Araujo, J., & Rivas-de-Roca, R. (2022). When Politicians Meet Experts: Disinformation on Twitter About Covid-19 Vaccination. *Media And Communication*, *10*(2), 157–168. https://doi.org/10.17645/mac.v10i2.4955

Radechovsky, J., Berger, P., & Wolling, J. (2019). Nothing's gonna change my world—Or do journalistic clarifications help against rumors? *Scm Studies In Communication And Media*, *8*(4), 497–522. https://doi.org/10.5771/2192-4007-2019-4-497

Reis, J. C. S., Melo, P., Garimella, K., & Benevenuto, F. (2020). Can WhatsApp benefit from debunked fact-checked stories to reduce misinformation? *Harvard Kennedy School Misinformation Review*, *1*(5). https://doi.org/10.37016/mr-2020-035

Ribes, D., Hoffman, A., Slota, S., & Bowker, G. (2019). The logic of domains. *Social Studies of Science*, pp. 49, 281–309. https://doi.org/10.1177/0306312719849709.

Robertson, C. T., Mourão, R. R., & Thorson, E. (2020). Who Uses Fact-Checking Sites? The Impact of Demographics, Political Antecedents, and Media Use on Fact-Checking Site Awareness, Attitudes, and Behavior. *International Journal of Press/Politics*, *25*(2), 217–237. https://doi.org/10.1177/1940161219898055

Rodríguez-Pérez, C., Paniagua-Rojano, F. J., & Magallón-Rosa, R. (2021). Debunking Political Disinformation through Journalists' Perceptions: An Analysis of Colombia's Fact-Checking News Practices. *Media and Communication*, *9*(1), 264–275. https://doi.org/10.17645/mac.v9i1.3374

Roozenbeek, J., Linden, S. V. D., & Nygren, T. (2020). Prebunking interventions based on 'inoculation' theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*, *1*(2). https://doi.org/10.37016//mr-2020-008

Ruiz, M., Rubio, L., & Verdu, F. (2021). Fight Fake News. The Professional Profile Of The Information Verifier In Spain. *Vivat Academia*, *155*, 265–294. https://doi.org/10.15178/va.2022.155.e1365

Sarelska, D., & Jenkins, J. (2022). Truth on Demand: Influences on How Journalists in Italy, Spain, and Bulgaria Responded to Covid-19 Misinformation and Disinformation. *Journalism Practice*. https://doi.org/10.1080/17512786.2022.2153075

Scharrer, L., Pape, V., & Stadtler, M. (2022). Watch Out: Fake! How Warning Labels Affect Laypeople's Evaluation of Simplified Scientific Misinformation. *Discourse Processes*, 59(8), 575–590. https://doi.org/10.1080/0163853X.2022.2096364

Schudson, M. (2013). Reluctant Stewards: Journalism in a Democratic Society. *Daedalus*, *142*(2), 159–176. https://doi.org/10.1162/DAED_a_00210

Schuetz, S., Sykes, T., & Venkatesh, V. (2021). Combating COVID-19 fake news on social media through fact checking: Antecedents and consequences. *European Journal Of Information Systems*, *30*(4), 376–388. https://doi.org/10.1080/0960085X.2021.1895682

Sencan, I., & Soydal, I. (2023). Fact-Checking Behaviors Of Undergraduate Students. *Information Development*. https://doi.org/10.1177/02666669231156861

Seo, S. (2022). Why the media gets it wrong when it comes to North Korea: Cases of 'dead' North Koreans in the Kim Jong-un era. *Journalism*. https://doi.org/10.1177/14648849221105924

Shahzad, K., & Khan, S. (2022). Relationship between new media literacy (NML) and web-based fake news epidemic control: A systematic literature review. *Global Knowledge Memory And Communication*. https://doi.org/10.1108/GKMC-08-2022-0197

Sharma, U., Pandey, P., & Kumar, S. (2022). A Transformer-Based Model for Evaluation of Information Relevance in Online Social-Media: A Case Study of Covid-19 Media Posts. *New Generation Computing*, *40*(4), 1029–1052. https://doi.org/10.1007/s00354-021-00151-1

Shin, J., & Thorson, K. (2017). Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. *Journal of Communication*, *67*(2), 233–255. https://doi.org/10.1111/jcom.12284

Sikder, O., Smith, R., Vivo, P., & Livan, G. (2020). A minimalistic model of bias, polarization and misinformation in social networks. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-62085-w

Singer, J. (2020). Border patrol: The rise and role of fact-checkers and their challenge to journalists' normative boundaries. *Journalism*, *22*(8), 1929–1946. https://doi.org/10.1177/1464884920933137

Sivasankari, S., & Vadivu, G. (2021). Tracing the fake news propagation path using social network analysis. *Soft Computing: A Fusion of Foundations, Methodologies and Applications*, 1–9. https://doi.org/10.1007/s00500-021-06043-2

Smith, J. H., & Bastian, N. D. (2022). A ranked solution for social media fact checking using epidemic spread modeling. *Information Sciences*, *589*, 550–563. https://doi.org/10.1016/j.ins.2022.01.002

Smyrnaios, N., Chauvet, S., & Marty, E. (2019). Journalistic Collaboration as a Response to Online Disinformation. *Sur Le Journalisme, About Journalism, Sobre Jornalismo*, 8(1), 68–81. http://www.surlejournalisme.com/rev

Soprano, M., Roitero, K., La Barbera, D., Ceolin, D., Spina, D., Mizzaro, S., & Demartini, G. (2021). The many dimensions of truthfulness: Crowdsourcing misinformation assessments on a multidimensional scale. *Information Processing & Management*, *58*(6). https://doi.org/10.1016/j.ipm.2021.102710

Stokes-Parish, J. (2022). Navigating credibility of online information during COVID-19: Using mnemonics to empower the public to spot red flags in health information online. Journal of Medical Internet Research, 24(6), e38269. https://doi.org/10.2196/38269

Sun, Y. (2022). Verification Upon Exposure to COVID-19 Misinformation: Predictors, Outcomes, and the Mediating Role of Verification. *Science Communication*, *44*(3), 261–291. https://doi.org/10.1177/10755470221088927

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the Backfire Effect: Measurement and Design Considerations. *Journal of Applied Research in Memory and Cognition*, *9*(3), 286–299. https://doi.org/10.1016/j.jarmac.2020.06.006

Tay, L. Q., Hurlstone, M. J., Kurz, T., & Ecker, U. K. H. (2022). A comparison of prebunking and debunking interventions for implied versus explicit misinformation. *British Journal of Psychology*, *113*(3), 591–607. https://doi.org/10.1111/bjop.12551

Tebaldi, C., & Nygreen, K. (2022). Opening or Impasse? Critical Media Literacy Pedagogy in a Posttruth Era. *Cultural Studies/Critical Methodologies*, *22*(2), 143–153.

Thanarak, R., & Ruangnapakul, N. (2021). Civic-Technology: Lesson Learned From Cofact, A Fact--Checking Platform In Thailand. *Journal Of Engineering Science And Technology*, 16, 137–143.

Thomson, T., Angus, D., Dootson, P., Hurcombe, E., & Smith, A. (2022). Visual Mis/disinformation in Journalism and Public Communications: Current Verification Practices, Challenges, and Future Opportunities. *Journalism Practice*, *16*(5), 938–962. https://doi.org/10.1080/17512786.2020.1832139

Thornhill, C., Meeus, Q., Peperkamp, J., & Berendt, B. (2019). A Digital Nudge to Counter Confirmation Bias. *Frontiers In Big Data*, *2*. https://doi.org/10.3389/fdata.2019.00011

Tian, Y., & Ding, X. (2019). Rumor spreading model with considering debunking behavior in emergencies. *Applied Mathematics & Computation*, *363*(C). https://doi.org/DOI: 10.1016/j.amc.2019.124599

Tunçer, S., & Tam, M. S. (2022). The COVID-19 Infodemic: Misinformation About Health on Social Media in Istanbul. *Türkiye İletişim Araştırmaları Dergisi*, *40*, 340–358. https://doi.org/10.17829/turcom.1050696

Ufarte-Ruiz, M., Galletero-Campos, B., & Lopez-Cepeda, A. (2020). Fact-Checking, a Public Service Value in the Face of the Hoaxes of the Healthcare Crisis. *Tripodos*, *47*, 87–103. https://www.raco.cat/index.php/Tripodos/article/view/377176

Uscinski, J. E., & Butler, R. W. (2013). The Epistemology of Fact Checking. *Critical Review*, *25*(2), 162–180. https://doi.org/10.1080/08913811.2013.843872

van der Linden, S., Roozenbeek, J., Maertens, R., Basol, M., Kacha, O., Rathje, S., & Traberg, C. S. (2021). How Can Psychological Science Help Counter the Spread of Fake News? *Spanish Journal Of Psychology*, *24*. https://doi.org/10.1017/SJP.2021.23

Vraga, E., Kim, S., & Cook, J. (2019). Testing Logic-based and Humor-based Corrections for Science, Health, and Political Misinformation on Social Media. *Journal Of Broadcasting & Electronic Media*, *63*(3), 393–414. https://doi.org/10.1080/08838151.2019.1653102

Vraga, E., Kim, S., Cook, J., & Bode, L. (2020). Testing the Effectiveness of Correction Placement and Type on Instagram. *International Journal Of Press-Politics*, 25(4), 632–652. https://doi.org/10.1177/1940161220919082

Vu, H., Baines, A., & Nguyen, N. (2022). Fact-checking Climate Change: An Analysis of Claims and Verification Practices by Fact-checkers in Four Countries. *Journalism & Mass Communication Quarterly*. https://doi.org/10.1177/10776990221138058

Walter, N., Cohen, J., Holbert, R., & Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, *37*(3), 350–375. https://doi.org/10.1080/10584609.2019.1668894

Winters, M., Oppenheim, B., Sengeh, P., Jalloh, M. B., Webber, N., Pratt, S. A., Leigh, B., Molsted--Alvesson, H., Zeebari, Z., Sundberg, C. J., Jalloh, M. F., & Nordenstedt, H. (2021). Debunking highly prevalent health misinformation using audio dramas delivered by WhatsApp: Evidence from a randomised controlled trial in Sierra Leone. *BMJ Global Health*, 6(11). https://doi.org/10.1136/bmjgh-2021-006954

Wintersieck, A. L. (2017). Debating the Truth: The Impact of Fact-Checking During Electoral Debates. *American Politics Research*, *45*(2), 304–331. https://doi.org/10.1177/1532673X16686555

Wood, T., & Porter, E. (2018). The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence. *Political Behavior*, *41*(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y

Young, D. L G., Holbert, R. L., & Jamieson, K. H. (2014). Successful Practices For The Strategic Use Of Political Parody And Satire: Lessons From The P6 Symposium And The 2012 Election Campaign. *American Behavioral Scientist*, *58*(9), 1111–1130. Https://Doi.Org/10.1177/0002764213506213

Yu, W., Payton, B., Sun, M., Jia, W., & Huang, G. (2023). Toward an integrated framework for misinformation and correction sharing: A systematic review across domains. *New Media & Society*. https://doi.org/10.1177/14614448221116569

Zhou, X. Y., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, *53*(5). https://doi.org/10.1145/3395046

DOI: 10.51480/1899-5101.18.3(41).819

Towards a Multi-Level Model of Resilience to Information Disorders: A Systematic Literature Review

Dmytro Iarovyi

0000-0002-5848-7785
Vytautas Magnus University, Lithuania;
Kyiv School of Economics, Ukraine

Sten Torpan

0000-0003-2052-092X
University of Tartu, Estonia

Ragne Kõuts-Klemm

0000-0001-9277-6700
University of Tartu, Estonia

Kristina Juraitė

0000-0002-7521-8560
Vytautas Magnus University, Lithuania

Abstract: The article attempts, based on academic sources, to develop a systematic framework of the concepts that define resilience against information disorders – specifically, the factors of resilience and the social practices that support it. To achieve this, a domain-based systematic literature review (SLR) was conducted, drawing on an interdisciplinary selection of articles from peer-reviewed journals. The article describes and categorizes factors that either facilitate or undermine resilience against information disorders, organizing them into a three-dimensional model comprising macro-level, meso-level, and micro-level factors. The findings also emphasize existing academic achievements in recognizing the diverse approaches to enhancing resilience, some of which extend beyond the communication sphere. Given the growing challenges posed by disinformation and information manipulation, this review offers an evidence-based foundation for both theoretical exploration and practical interventions.

Keywords: resilience, vulnerability, information disorders, disinformation, systematic literature review

INTRODUCTION

Resilience has become a prominent concept within the social sciences for the past few decades, describing how societies and individuals cope with crises and disruptions. Earlier theories of social change – evolutionary, cyclical, or structural (Schneider, 1976; Bourdieu & Coleman, 1991; Parsons, 1968) – focused on gradual transformations, whereas resilience draws attention to sudden shocks that can divert societies from their expected developmental paths.

Following Obrist et al. (2010), we view resilience as a value-laden construct shaped by its political and institutional context. In rapidly changing environments, it cannot be separated from structural conditions or power relations. Lorenz (2013) similarly emphasizes that resilience is relational and processual: social systems are continuously changing, and, echoing Heraclitus' assertion that "no man ever steps in the same river twice", they never return to the same state after major shocks. Maclean et al. (2017) and Moya & Goenechea (2022) also argue that resilience is not an end-state but an ongoing adaptive process shaped by learning, governance, and institutional arrangements.

Consequently, resilience should be conceptualized not as a static attribute but as a dynamic interaction between systems and their environments, raising questions about where resilience is constructed: at the micro level through individual and community capacities, or at the macro level through institutions and structures. These questions are particularly relevant in the field of information and communication, where technological change, the attention economy, unmoderated platforms, and advances in AI produce new and unforeseen vulnerabilities. In a world organized through networked communication, many threats stem from information disorders, making societal resilience in this domain an urgent research topic.

Among the plethora of academic and policy papers, there is still a recognizable lack of studies that focus on reducing the impact of false, misleading and manipulative information in real-world environments (Roozenbeek et al., 2022). Some scholars acknowledge the relatively slow pace of research on information disorders in the context of global transformational changes (Kozyreva et al., 2022).

These are the issues we aim to address in the present study. Our goal is to explore the concept of societal resilience within the field of information and communication, focusing on the factors conditioning resilience and vulnerability to information disorders, and the tools used to maintain it. We do not focus on approaches to societal resilience conceptualization, as we find little to no gap in this area. Instead, this review aims to clarify types of vulnerabilities that emerge alongside resilience factors and to assess how these elements can be organized into a multi-level conceptual model. Thus, our research questions are: (1) What factors are identified in the literature as either weakening or strengthening resilience

to information disorders? (2) How do the identified vulnerability factors interact with resilience-enhancing mechanisms across macro-, meso-, and micro-levels? And (3) what social practices are observed for sustaining this type of resilience? The outcome of our research, as we see it, are the frameworks that systematize both the contributing factors and the social practice used when approaching societal resilience against information disorders.

Our approach relies on two complementary bodies of literature: one focused on resilience and the other on vulnerabilities. This dual strategy enables capturing sources that address both the "positive" (resilience) and "negative" (vulnerability) dimensions of adaptive capacity.

The review is designed to serve multiple audiences. For scholars in communication studies, it provides a synthesis of existing research and a conceptual framework that can guide future work. For policymakers, educators, and practitioners, it offers insights into how resilience can be developed and supported in practice. Theoretically, this article contributes to interdisciplinary debates by providing perspectives from politics, technology, and communication.

As this review was being completed, several other attempts to synthesize the field emerged. Rød et al. (2025) offer an interesting framework for combatting disinformation constructed through thematical blocks rather than levels (legal, educational, political and governance, psychological, and technological). Ferraz & Hellal's (2024) approach resilience from an information-science perspective, yet emphasizing more system robustness and technical safeguards, in their review of informational resilience. Ricard et al. (2025) also engage in the systematic overview of information disorder mechanisms, providing a model of information disorder as a five-stage socio-material process, mapping vulnerabilities, agents, tactics, and impacts across the information environment. These studies confirm the need for conceptual consolidation, yet they remain oriented toward specific disciplinary lenses (security studies, information science, or modelling approaches). The present review complements these efforts by offering a multi-level framework grounded in communication and social-science research.

FRAMING THE CHALLENGE: RESILIENCE TO INFORMATION DISORDERS

Recognizing the relational nature of resilience allows us to examine the specific threats and vulnerabilities within information and communication. In this domain, resilience is shaped by the evolving digital environment, the dynamics of mis- and disinformation, and the structural conditions that enable or constrain adaptive responses. Treating resilience as multi-level, dynamic, and threat-specific (Lorenz, 2013) is therefore essential.

Clarifying what constitutes these threats requires addressing the fragmented terminology surrounding misleading content. The academic vocabulary is highly inconsistent, with overlapping terms shaped by different disciplines and historical contexts, complicating both scholarly and policy efforts (Pantazi et al., 2021). Although resolving these debates lies beyond our scope, a working definition is necessary.

Several attempts to systematize terminology exist. Kapantai et al. (2021) map taxonomies of false or misleading information, drawing on concepts such as "information pollution" (Meel & Vishwakarma, 2019) and "information disorders" (Wardle & Derakhshan, 2017), ultimately adopting the European Commission High-level expert group definition of "disinformation" as false or misleading information intentionally promoted to cause harm or profit (HLEG, 2018). Their taxonomy includes "fake news", "false information", and "disinformation", but notably excludes "misinformation".

Wardle & Derakhshan (2017), however, argue that "information disorders" offers a broader systemic perspective by shifting attention from content to the wider ecosystem in which it circulates. Given this study's interdisciplinary scope, we follow Golob et al. (2021) in using "information disorders" as an overarching term, as different forms of misleading content – such as misinformation, disinformation, and propaganda – often operate through similar mechanisms and require comparable resilience strategies.

Before analysing the factors of resilience against information disorders, it is important to clarify how resilience in this field is understood in relation to its dynamics and vector.

When discussing resilience to information disorders in the context of its dynamics, scholars generally identify two macro approaches: paternalistic and adaptive (Moussa et al., 2022; Romanova et al., 2020). The *paternalistic paradigm* assumes a more conservative approach of "bouncing back", returning to the status quo (Buzzanell, 2010), and the *adaptive* entails "moving forward", adjusting to information threats, and adapting to those challenges (Kõuts-Klemm et al., 2022). The former relies on the involvement of authorities and regulatory development, and the latter emphasizes citizen participation and empowering individuals to prevent the spread of information disorders (Lazer et al., 2018, as cited by Golob et al., 2021).

Resilience to information disorders can also be interpreted through the lens of the "freedom-security dilemma", a continuum between public security and public freedom. In asymmetric information warfare, this dilemma becomes salient as democracies respond to autocratic influence by balancing *security-oriented physical resilience* measures (such as censorship or blocking sources) with *freedom-oriented cognitive resilience* that is more inclined towards developing critical thinking. Physical and cognitive resilience should not necessarily

be dichotomic; they can complement one another. Nonetheless, policymakers frequently frame them as opposing paradigms, where prioritizing one is seen as jeopardizing the other (Liagusha & Iarovyi, 2025).

Since the paternalistic approach to resilience is typically associated with public security, and the adaptive approach with public freedom (Romanova et al., 2020), it may seem intuitive to align physical resilience with paternalism and cognitive resilience with adaptation, placing both on the freedom-security continuum. However, this categorization oversimplifies their interaction. The freedom-security distinction itself is contradictory: regulatory measures may simultaneously be justified as protecting free speech (Shadmy, 2022) and criticized for restricting it (Kozyreva et al., 2022).

In the next section, resilience is approached from vulnerability perspective, and based on existing studies in information and communication, we outline the levels of its analysis.

LEVELS OF RESILIENCE

Building on the discussion of resilience and its paradigms, we now turn to the concept of vulnerability as its counterpart. In social sciences, vulnerability can be defined as an internal risk factor within a subject or system exposed to a hazard, reflecting an intrinsic susceptibility to harm (Singh et al., 2014). In this sense, resilience and vulnerability function as opposite ends of a continuum.

To locate vulnerability and resilience analytically, we must also clarify the levels of analysis. Hall and Lamont (2013) distinguish between macro, organisational, and individual resilience. While the organisational level – including actors that deliberately amplify information disorders – remains important, analysing it in depth lies beyond the scope of this article. This dimension has been extensively examined elsewhere (e.g., Dawson & Innes, 2019). In this study, we focus on the macro-societal and individual levels, where information flows, everyday practices, and structural conditions most directly shape societal resilience to information disorders.

At the same time, resilience cannot be reduced to a macro-micro distinction. Research in information and communication consistently shows that these levels are connected through intermediary processes and structures. Bjola & Papadakis (2020) base their findings on Habermas' concept of the public sphere, and regarding the consumption of disinformation, they explain that on the micro level, within everyday use, people act as both producers/distributors of disinformation (often unintentionally) and its consumers. At the macro level, disinformation serves as an incentive for decision-makers. The topics infiltrate

the macrosphere from below, as emotion-driven and algorithmically amplified disinformation undermines the norm-sustaining function of the public sphere.

Moussa et al. (2022) similarly observe that integrating macro variables matter, but do not deterministically shape individual resilience, as people deploy diverse strategies with varying effectiveness. Boulianne et al. (2022) argue that developing resilience requires examining how macro- and micro-level indicators relate to one another, noting methodological distinctions between exposure, awareness, and sharing of disinformation.

An important contribution comes from Humprecht and colleagues, whose comparative studies of online resilience highlight macro-level variation. Humprecht et al. (2020) conceptualize resilience as a collective characteristic with individual-level implications and identify seven key macro indicators: societal polarization, populist communication, trust in news, strength of public media, media fragmentation, advertising market size, and social media use. Their later work stresses that these indicators are context-dependent and shaped by political and informational environments (Humprecht et al., 2023). For instance, consensus-oriented democracies may experience less impact from extreme ideologies, whereas resilience in countries with populist dominance may be negatively affected; the trust in traditional media may have been weakened during the pandemic, and resilience to disinformation may vary depending on the topics being discussed. Thus, strong macro-level strategies do not automatically translate into individual-level resilience (Bjola & Papadakis, 2020, as cited in Kõuts-Klemm et al., 2022).

Consequently, the questions arise about how countries and separate individuals become resilient or vulnerable to information disorders, and whether other factors contribute to explaining these variations.

Resilience is also a discursive phenomenon, understood as public power to identify the risk, where public discourses in terms of perceptions of risk and relevant beliefs, knowledge and experiences are of great importance (Obrist et al., 2010).

Recognizing this interplay across levels, this systematic literature review develops a three-dimensional approach to resilience and vulnerability to information disorders. These dimensions encompass macro-level contextual factors, meso-level discursive practices, and micro-level individual characteristics. We propose this model as a useful analytical framework for understanding how resilience operates within the field of information and communication.

SYSTEMATIC LITERATURE REVIEW

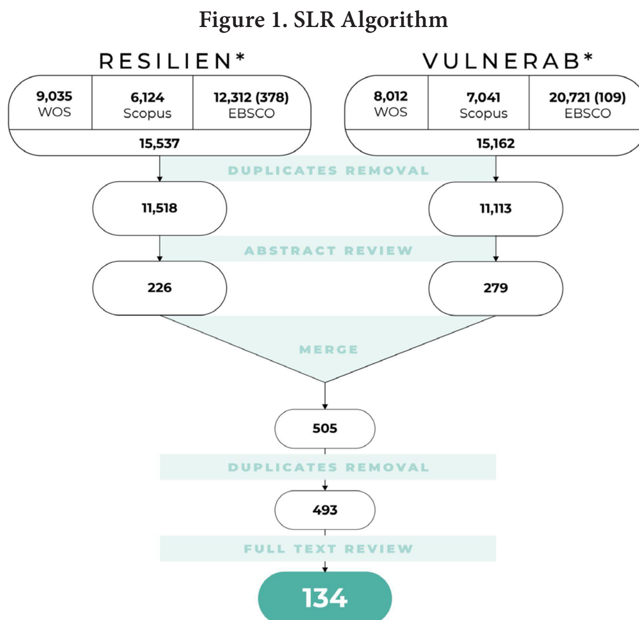
The choice of the method (a systematic literature review – SLR) was guided by the research team's recognition that the topic is evolving and requires systematization (Paul & Barari, 2022). This is a domain-based conceptual review focused on developing a conceptual model and proposing future research directions within the field (ibid.).

The review timeline covers from 2010 to May 2023, when the article extraction took place. The choice of 2010 as the starting point aimed to capture the most recent findings and served as a precaution, as most articles used for the review were published after the mid-2010s. A search for relevant concepts prior to 2010 found no direct connections to the information and communication field.

Following established qualitative SLR guidelines and prioritising rigor (Paul & Barari, 2022), we included peer-reviewed English-language articles from journals indexed in Web of Science or Scopus, supplemented by EBSCO to ensure strong coverage of the social sciences. To capture the field from multiple angles, we conducted two separate searches using the terms "resilien*" and "vulnerab*" combined with keywords related to information disorders. At this stage, we applied only broad exclusions, omitting fields clearly outside social sciences, information, and communication to avoid missing potentially relevant work (see Annex 2, 3 for the details).

This search yielded the following results (see Figure 1).

Figure 1. SLR Algorithm



For the EBSCO results, where field-specific filtering was not possible, we manually selected items relevant to information disorders, yielding 378 articles for "resilien*" and 109 for "vulnerab*."

All retrieved items were then imported into the Rayyan platform (15,537 for "resilien*" and 15,162 for "vulnerab*"). After automatic duplicate detection at a 95% similarity threshold and manual verification, the datasets were reduced to 11,518 and 11,113 items, respectively.

We next conducted an abstract screening to remove articles unrelated to resilience or vulnerability in the information and communication domain. Following the approach of Loecherbach et al. (2020), we excluded studies from other fields that use similar terminology but address different conceptual problems. While subjective judgement is inherent to SLRs, the interdisciplinary background of our international team strengthens the reliability and replicability of the selection.

As "resilience" and "vulnerability" generate large cross-disciplinary datasets, we applied structured exclusion criteria to ensure conceptual relevance to information disorders. This step is standard in domain-focused SLRs to reduce false positives arising from overlapping terminology. A full list of excluded subject areas is available upon request. After this screening, 226 "resilien*" items and 279 "vulnerab*" items remained.

These sets were merged, with 12 duplicates removed, resulting in 493 articles. Each was independently reviewed by two scholars, and items were excluded only when both reviewers agreed they did not meet the predefined criteria (see Annex 2). The final systematic literature review includes 134 peer-reviewed English-language articles directly addressing resilience and vulnerability to information disorders within the communication field (Annex 1).

RESULTS

The following section presents the results of our SLR, categorizing the factors across three levels.

MACRO-LEVEL: CONTEXTUAL FACTORS

For the purposes of the model, we largely draw on the factors identified by Humprecht et al. (2020), which we consider seminal. Since the focus of the SLR extends beyond online disinformation, we have incorporated additional factors into the analysis.

One group of macro-level factors relates to the political environment. Humprecht et al. (2020) measure levels of *political polarization* and *populist communication*, which are conceptually interconnected. Some scholars associate political polarization with *social inequality*, which fosters conflict, damages the information

climate, and exacerbates divisions between elites and the public, as well as within society itself, creating fertile ground for the spread of misleading information (Bargaoanu & Radu, 2018; Loos et al., 2018; Nguyễn et al., 2023; Pottle, 2022; Scheffer et al., 2022). Freedman et al. (2021) also argue that greater *societal equality* promotes solidarity, making it more difficult for information disorders to destabilize society.

Political polarization intensifies during *electoral processes* (Badrinathan, 2021; Filipec, 2019), *through the interventions by external actors* (Bjola & Papadakis, 2020; Cook, 2022; Disha et al., 2023; Dolan, 2022; Manwaring & Holloway, 2023; Rosca, 2018), and *during crises* such as the COVID-19 pandemic (Dunleavy et al., 2022; Vériter et al., 2020).

Humprecht et al. (2020) also emphasize media-related factors, such as *trust in media*, the *strength of public service media*, and *audience fragmentation*. These factors interact with the political environment (e.g., high levels of populism can undermine traditional media credibility and general media trust) and may either bolster or undermine resilience to information disorders (yet the authors note about the possible contradictory effects, considering the exposure to disinformation and its sharing).

Other media-related factors influencing resilience include *the quality of media legislation* aimed at preventing disorders (Filipec, 2019; Rosca, 2018; Vériter et al., 2020), *media market monopolization* (Lebid et al., 2021; Rosca, 2018), *the role of media literacy education and critical thinking* in society (Filipec, 2019; McDougall & Rega, 2022), *the professionalism and training of media and journalism professionals* (Disha & Rustemi, 2022; Rosca, 2018), and the *status (decline or development) of the mainstream media*, including at the local level (Jerónimo & Esparza, 2022).

Trust plays a crucial role in both media and political environments, shaping how individuals engage with information; as McDougall (2019) notes, it is "a key discursive marker" in media literacy. Low trust creates openings for malicious actors and increases vulnerability to disinformation (Bjola & Papadakis, 2020; Disha & Rustemi, 2022; Reeves, 2022; Rožukalne et al., 2022; Wagnsson, 2020), and can prompt individuals to seek negative content about actors they distrust (Klebba & Winter, 2023), while higher trust in news media reduces exposure (Boulianne et al., 2022). Yet excessive trust also carries risks: uncritical acceptance of information heightens susceptibility to misleading content (Collado et al., 2020; Zimmerman et al., 2022) and rumours (Indu & Thampi, 2021), reinforced by truth bias – the tendency to assume claims are true (Claudia, 2022; Drotner, 2020; Deinla et al., 2022). Group-based trust further weakens scrutiny in hostile environments (Reeves, 2022), and longstanding conformity-based trust patterns, established as early as the Neolithic era (Walsh, 2021) may increase vulnerability. Context matters: trust in science can reduce susceptibility

to COVID-19 disinformation (Roozenbeek et al., 2020) but can also heighten receptivity to pseudo-science (Albarracín, 2022; O'Brien et al., 2021).

In summary, macro-level factors are about the context in which the information exists. Moving on to the meso-level, we focus on the discursive practices and patterns, examining *how* information is produced, shared, and consumed.

MESO-LEVEL: CHARACTERISTICS OF THE ENVIRONMENT

The meso-level factors relate to the characteristics of the information itself and how it is conveyed and disseminated.

The model by Humprecht et al. (2020) highlights the distinct role of social media alongside political and traditional media factors. Because our framework treats discursive practices as a separate level of analysis, we place social media influences at the meso-level. This is particularly relevant in democracies, where media pluralism means that neither traditional nor digital platforms dominate individuals' media diets.

Social media significantly accelerates the spread of information disorders. Humprecht et al. (2021, 2023) highlight social media market size, the volume of malicious content, and repeated exposure to disinformation. Boulianne et al. (2022) note that higher social media use correlates with increased awareness, exposure to, and sharing of misleading content. While greater access to information helps individuals find useful content, it also increases exposure to misleading information (Pérez-Escoda, 2022). These factors contribute significantly to information disorders, extending beyond social media to mainstream media.

Social media platforms, with weaker gatekeeping and lower filtering capabilities, often allow misleading content to go undetected (Badrinathan, 2021; Christiano, 2022; Puebla-Martínez et al., 2021; Yan & Pan, 2023). It happens due to the overwhelming volume of information, which hampers filtering processes (Kirby et al., 2018; Yang & Tian, 2021; Zimmerman et al., 2022) and distracts consumers (Fritts & Cabrera, 2022; Kozyreva et al., 2020). Even a consideration of sharing disinformation can reduce the resilience of individuals to its content (Epstein et al., 2023).

As the "price" of sharing is now lower, the speed of information flow increases (Salvi et al., 2021; Vériter et al., 2020), and misleading content often spreads faster than accurate information (Filipec, 2019; Kyza et al., 2020; et al., 2021). This speed is driven by the desire for novelty, sensation, or entertainment (Caled & Silva, 2022; Kozyreva et al., 2020), and by monetization (Fritts & Cabrera, 2022; Nguyễn et al., 2023). Rapid dissemination reduces the effectiveness of interventions, as disinformation often goes viral before it can be debunked (Stein et al., 2023), and the combination of volume and speed encourages "vertical

reading" (Tripodi et al., 2023) – quick, surface-level processing that amplifies emotional responses and susceptibility to influence (Claudia, 2022; Christiano, 2022; Kozyreva et al., 2020; Petukhov & Chuprakova, 2014; Reeves, 2022). Due to the perceived closeness and immediacy of social media, people may also tend to over-rely on it as an information source, often without engaging in sufficient fact-checking (Caled & Silva, 2022; Zimmerman et al., 2022), which exacerbates the above-mentioned issue of excessive trust.

The repetitive exposure to information, an attribute of social media, has been discussed as far as since the classic studies of Lasswell. Repetitive exposure fosters the creation of echo chambers (Caled & Silva, 2022; Deinla et al., 2022; Di Mascio et al., 2021; Greer et al., 2022), reinforcing what is known as the "*illusory truth effect*", where repeated statements are perceived as true regardless of their veracity (Loos et al., 2018; Millar, 2022). Another feature of that is the echo-chamber, a place of *continued influence*. Even after disinformation is debunked, it often persists in memory, sometimes more so than the source of the information itself (Millar, 2022; Mlejnková, 2022), especially in online spaces (Millar, 2022; Yang & Tian, 2021).

There are technological advancements that are making information disorders more "efficient": *algorithmization* of content, which allows for highly targeted dissemination, reaching selective groups (Bjola & Papadakis, 2020; Christiano, 2022; Kertysova, 2018; Humprecht et al., 2020; Mlejnková, 2022), *use of bots and trolls*, or *AI-driven solutions* such as deepfakes, which simplify the information production (Bjola & Papadakis, 2020; Humprecht et al., 2020; Miyamoto, 2021).

Other features making information disorders more believable, include the information being emotionally charged, which contributes to the radicalization (Bjola & Papadakis, 2020), confusing narrative, which creates a sense of frustration (Nagler et al., 2023), and well-elaborated misleading content, which is perceived as more believable (Greer, et al., 2022).

These practices demonstrate how the content and structure of information disorders influence the public and its perception. However, understanding the full scope of resilience also requires examining how individuals process this information.

MICRO-LEVEL: INDIVIDUAL CHARACTERISTICS

Demographic characteristics such as age, education, and gender may appear to be straightforward predictors of resilience or vulnerability. However, although several studies address these factors, we chose not to adopt a strictly demographic approach. Demographics alone do not capture the complexity of individual variability, and generalizations lead to risk oversimplification. Demographic groups

are not homogeneous (Loos et al., 2018; Moussa et al., 2022), and cross-country studies show substantial contextual variation.

Findings on education, for example, are mixed: many studies report a positive correlation between higher education and resilience (Badrinathan, 2021; Golob et al., 2021; Lai et al., 2020; Rodríguez-Pérez & Canel, 2023; Seo et al., 2021), while others find little or no effect (Boulianne et al., 2022; Claudia, 2022).

Similarly, the role of gender is inconsistent – while Boulianne et al. (2022) argue that being female is rarely significant in predicting awareness of, exposure to, and sharing of misinformation, Rodríguez-Pérez & Canel (2023) suggest that women may show weaker resilience. Golob et al. (2021) suggest that despite higher meta-reflexivity, women may have fewer opportunities to fact-check content due to domestic responsibilities.

Age also presents mixed results. Older adults having higher information literacy and lower trust in mainstream media are more mentioned as more affected by fake news (Anspach & Carlson (2022), while younger people may be more vulnerable due to greater exposure to information (Craig & Vijaykumar, 2023; Loos et al., 2018; Miyamoto, 2021). Increased exposure can worsen attitudes toward political institutions (Stankutė-Søsted, 2019) and make younger people more likely to share disinformation (Deinla et al., 2022; Greer et al., 2022; Jo et al., 2022). Conversely, older adults may overestimate their immunity to disinformation, despite weaker media literacy (Boulianne et al., 2022; Claudia, 2022; Golob et al., 2021; Kozyreva et al., 2020; Miyamoto, 2021; Rodríguez-Pérez & Canel, 2023; Moore & Hancock, 2022; Zainab et al., 2022).

Because demographic characteristics rarely act independently and instead intersect with other social factors, we do not treat them as standalone predictors. Our micro-level analysis focuses on individual traits that shape resilience or vulnerability to information disorders, including information-processing capacities, perceptual and evaluative biases, affective dispositions, and value orientations.

The first and most widely discussed group – *information-processing capacities* – can be linked to media and information literacy, though we treat it as a broader construct. It encompasses both cognitive skills and knowledge-based dispositions, forming a coherent set of practices that determine an individual's ability to process information effectively.

Motivated reasoning (confirmation bias) leads individuals to seek information that reinforces their prior attitudes (Caled & Silva, 2022; Collado et al., 2020; Humprecht et al., 2023; Klebba & Winter, 2023; Levine, 2022; Millar, 2022; Pantazi et al., 2021; Pehlivanoglu et al., 2022; Pérez-Escoda, 2022; Scherer & Pennycook, 2020; Yan & Pan, 2023). It is often driven by cognitive rigidity and binary thinking – core components of socio-cognitive polarization (Klebba & Winter, 2023; Salve et al., 2021). Experiments show that exposure to diverse

ideological content improves the ability to detect disinformation while ideologically segregated environments reduce it (Stein et al., 2023).

Conversely, *analytical reasoning and critical thinking* strengthen resilience (Bjola & Papadakis, 2020). Epistemic humility supports critical evaluation (Billingsley & Heyes, 2023), and analytical skills improve information processing (Rodríguez-Pérez & Canel, 2023). Critical engagement with media fosters scepticism and detection of manipulative structures (Olariu, 2022; Pehlivanoglu et al., 2022) and encourages fact-checking (Golob et al., 2021; Yan & Pan, 2023). Political knowledge may also enhance resilience (Mader et al., 2022), though findings remain mixed (Nyhan & Zeitzoff, 2018).

Perceptual and evaluative biases also undermine resilience. Excessive *self-confidence* creates a false sense of immunity to manipulation (Benková, 2019; Manwaring & Holloway, 2023; McDougall, 2019; Pérez-Escoda, 2022; Rodríguez-Pérez & Canel, 2023), while third-person perception – the belief that others are more vulnerable than oneself – similarly distorts risk assessment (Altay & Acerbi, 2023; Loos et al., 2018; van der Meer et al., 2023; Yang & Tian, 2021). This effect is amplified by social media use (Yang & Tian, 2021) and mobile news consumption (Liu & Huang, 2020).

There are factors that, although do not necessarily affect all individuals in the same way (Pierre, 2020), may hinder the resilience to information disorders: for instance, *paranormal beliefs/magical thinking* (Blanco et al., 2015; March & Springer, 2019), or *conspiracy beliefs* (Mader et al., 2022).

Feelings and emotions, though short-lived, significantly shape how individuals engage with information; affective dispositions therefore matter. Therefore, the affective dispositions matter. F*ear and anxiety* limit rational perception (Liu & Huang, 2020; Yang & Tian, 2021), pushing individuals toward information-seeking behaviour and spreading any information they come across (Salvi et al., 2021), which makes them more vulnerable to phishing attacks (Nugraha et al., 2020). The third-person effect is also used as a coping mechanism to manage anxiety (Liu & Huang, 2020). Other factors include *traumatic experiences* (Nguyễn et al., 2023, *high levels of neuroticism or extraversion* (Lai et al., 2020), such personality traits as *Machiavellianism* (March & Springer, 2019), or having *schizotypal, paranoid, or histrionic* personality traits (Escolà-Gascón et al., 2023),

The *value orientations* that can mediate exposure and reaction to information disorders are the political, ideological, or other group-based identities that frame the interpretation.

Group identity plays a crucial role in shaping the above-described motivated reasoning. The social congruence in disinformation perception occurs when individuals accept messages aligned with beliefs shared by their social networks, leading to selective exposure (Levine, 2022), encouraging dissatisfied information consumers to seek alternative sources that push specific agendas (Mlejnková,

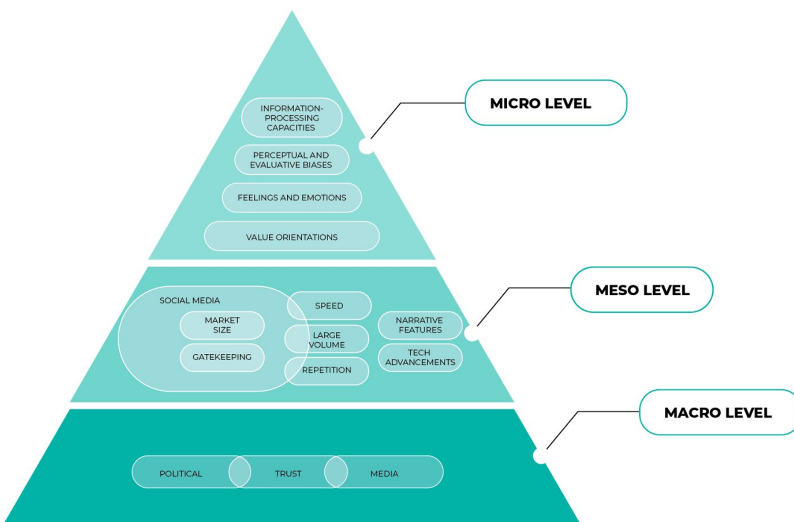
2022). When disinformation is linked to one's social group, perceived affiliation further increases vulnerability (Bost et al., 2010; Seo et al., 2021).

Vulnerable groups, such as *counterpublics*, with oppositional identities and discourses, are central to understanding resilience to information disorders. They function as parallel discursive spaces where marginalized or oppositional groups construct and amplify alternative narratives that challenge mainstream interpretations (Bjola & Papadakis, 2020). Krishna (2021) writes about the *lacuna publics* – knowledge-deficient, extreme activist groups whose radical views and limited factual grounding make them particularly susceptible to disinformation.

Political attitudes also influence vulnerability. This may stem from strong partisan attachment (Bragarnich, 2022; Disha et al., 2023; Disha & Rustemi, 2022; Kozyreva et al., 2020) or identification with specific parties or leaders (Clementson, 2018). Ideological beliefs shape how macro-level factors affect individuals (Boulianne et al., 2022), and several studies find that right-wing orientations are associated with greater susceptibility to disinformation – through biased information selection (Deinla et al., 2022; Klebba & Winter, 2023; Rodríguez--Pérez & Canel, 2023), scepticism toward fact-checking (Çömlekçi, 2022; Lyons et al., 2020), or reduced concern about disinformation (Rodríguez-Virgili et al., 2021). However, this relationship is not universal: some research finds no consistent ideological effect (Altay et al., 2023), and Baptista & Gradim (2022) argue that motivated reasoning occurs across the political spectrum.

Also, as a result, there has been an effort to group the factors of vulnerability listed above in a three-dimensional model (see Figure 2).

Figure 2. Factors of Vulnerability/Resilience to Information Disorders



SOCIAL PRACTICES OF RESILIENCE TO INFORMATION DISORDERS

Resilience emerges from the interaction of multi-level dynamics. Macro-level structures – political regimes, regulations, and media – shape information flows while continuously evolving through societal action. At the meso-level, changing modes of information circulation can either strengthen resilience or amplify vulnerabilities. Micro-level resilience is reflected in individual habits and patterns, yet these are shaped by the surrounding environment; even well-equipped individuals are influenced by social norms.

The choice of strategy for addressing information disorders is determined by context, resources, and vulnerable groups targeted by the interventions (Janmohamed et al., 2021; Klebba & Winter, 2023; Mader et al., 2022; Melo--Pfeifer & Gertz, 2022).

The fundamental recommendation for sustaining resilience is using the *proactive approach* – acting in advance to prevent the spread of disinformation, for instance through sustained exposure to counterevidence (Scheffer et al., 2022). It may take the form of inoculation, or prebunking rather than debunking (Kozyreva et al., 2022; Roozenbeek et al., 2022), to equip individuals with reliable sources, introduce well-crafted facts, and provide alternative narratives instead of merely refuting false claims (Bjola & Papadakis, 2020). It allows incorporating warning messages that trigger threat perception and is particularly effective when individuals are actively engaged in generating their own counterarguments (Habro et al., 2020; Lewandowsky & Van Der Linden, 2021; Roozenbeek et al., 2022). *Participatory practices* are useful in general, as content co-creation also improves mental health (Bruinenberg et al., 2021) which allows people to integrate knowledge into daily activities (Drotner (2020), decreases polarization and established common ground (Leurs et al., 2018).

The macro-level efforts are accomplished through the *regulatory and policy-based instruments* that address disinformation through the rule of law, such as banning the sources of disinformation (Albarracín, 2022; Hall, 2017). While some perspectives suggest that focusing solely on preventing exposure is not effective, and that resilience should instead concentrate on minimizing the impact of information disorders (Boulianne et al., 2022), we do not witness an academic consensus on the "unnecessity" of such interventions. These practices remain reasonable, especially in situations of crises, such as war, pandemics, or the public's low capacity for critical information analysis (Golob et al., 2021).

The *cooperative practices* include citizen-led initiatives (Hall, 2017), or government-established anti-disinformation bodies and regulators (Lebid et al., 2021; Rosca, 2018). The joint efforts of multiple actors is essential to achieve full potential (Buzzanell, 2010; Rosca, 2018). Cooperation takes on the forms of promoting justice, equality, and fairness, fostering an ethos of inquiry in society (Pottle, 2022), addressing divisions between groups through dialogue (Nguyễn et al.,

2023), enabling diverse representation to address injustice (Christiano, 2022), or fostering a motivated interest in civic affairs (Golob et al., 2021; Filipec, 2019).

One challenge is reaching *vulnerable groups* who are mistrustful of government, science, and media. Engaging them requires identifying relatable role models from different social groups who can resonate with these communities (Rožukalne et al., 2021). Building trust can also involve bringing in individuals with shared identities (Reeves, 2022) or offering information in the native languages of minority communities (Nguyễn et al., 2023; Rosca, 2018). Crowdsourcing, using a "wisdom of crowds", can serve as an effective tool for fact-checking through applications where users assess the accuracy of information (Claudia, 2022). There are limitations related to these groups: while fact-checking is a widely used tool to combat information disorders (Caled & Silva, 2022; Claudia, 2022; Çömlekçi, 2022; Fritts & Cabrera, 2022; Rožukalne et al., 2021; Rodríguez-Pérez & Canel, 2022; Yang & Tian, 2021), some studies suggest that if scepticism is linked to low trust (Reeves, 2022), strong pre-existing beliefs (Kyza et al., 2020; Millar, 2022), or ideological biases (Fritts & Cabrera, 2022; Lyons et al., 2020), fact-checking can reinforce, rather than diminish, those beliefs.

When discussing citizen engagement in resilience to information disorders, *educational practices* are closely linked to cognitive resilience, emphasizing the development of attitudes rather than merely providing information. These approaches focus on "teaching to fish" rather than "giving a fish" (McDougall, 2019), fostering self-organization, adaptability, trust development, and an emphasis on education (Romanova et al., 2020).

Starting from early education, media and information literacy enhances the individuals' ability to respond to manipulative tactics (Christiano, 2022; De Leyn et al., 2022; Dumitru, 2020; Frolova et al., 2022), raise awareness of vulnerabilities in distinguishing truth from falsehood (Rodríguez-Pérez & Canel, 2023), address biases, and promote critical content evaluation (Caled & Silva, 2022). They also encourage values that diminish the impact of malicious actors (Bjola & Papadakis, 2020), empowering individuals to make informed decisions and take active roles in society (Golob et al., 2021; McDougall, 2019).

For media and information literacy to be effective, it must be integrated into the general curriculum and reflect real-world contexts (Burns et al., 2023; Golob et al., 2021; McDougall, 2019; Moussa et al., 2022). Programs should adopt a multifaceted approach, including emotional resilience, behavioral vigilance, and cognitive skills (Nugraha et al., 2020). These efforts should extend beyond younger generations and be tailored to diverse audiences (Lee, 2018; Loos et al., 2018; McDougall, 2019; Pérez-Escoda, 2022; Yang & Tian, 2021). The backfire effect remains a potential challenge – while people are vulnerable to misleading content, they can also be resistant to accurate information or corrections (Pantazi et al., 2021) – in the process of training some people may increase distrust in both

false and truthful news, leading to misidentification of true information as false (Moore & Hancock, 2022).

Finally, the development of *high-quality journalism and media* is another crucial component (Caled & Silva, 2022; Vériter et al., 2020; Yang & Tian, 2021). Journalism reinforces its gatekeeping role to better manage the spread of "deviant information" (Hopp & Ferrucci, 2020). This responsibility underscores the need for intermedia cooperation, interaction, and experience exchange among journalists and media outlets (Çömlekçi, 2022; Golob et al., 2021; Jerónimo & Esparza, 2022; Kyza et al., 2020; McDougall, 2019; Rosca, 2018; Tan, 2022).

DISCUSSION AND FUTURE OUTLOOK

Considering factors of resilience and vulnerability to information disorders, we developed a three-dimensional model addressing macro, meso and micro levels.

The model is not claimed to be holistic or universally comprehensive in systematizing factors of resilience or vulnerability to information disorders. Nevertheless, this model would facilitate further efforts in framing and analysing resilience to information disorders in the social sciences, offering an overview of the measures used to sustain such resilience.

Building on the findings of this review, several avenues for future research emerge. First, the three-dimensional model requires empirical validation: future studies should reveal which of these factors exert the strongest predictive weight in real-world settings. Second, cross-national research is needed to determine how different social and political environments condition resilience and vulnerability, and which indicators are transferable across contexts. Third, experimental designs could assess the durability of specific interventions and evaluate how these tools operate within different population groups. Addressing these questions would allow the model presented in this review to evolve into an analytical toolkit for communication research and policy design.

Acknowledgement: The research was supported by the Estonian Research Council under the grant agreement No PRG2555 "Reinforcing Societal Resilience through Securitization" and conducted as part of the Baltic Engagement Centre for Combating Information Disorders (BECID) project.

REFERENCES

Bjola, C., & Papadakis, K. (2021). Digital propaganda, counterpublics, and the disruption of the public sphere: The Finnish approach to building digital resilience. In *The World Information War* (pp. 186–213). Routledge.

Boulianne, S., Tenove, C., & Buffie, J. (2022). Complicating the resilience model: a four-country study about misinformation. *Media and Communication*, *10*(3), 169–182. https://doi.org/10.17645/mac.v10i3.5346

Bourdieu, P. and Coleman, J. S. (Eds.). (1991). *Social Theory for a Changing Society*. Westview Press, Russell Sage Foundation.

Buzzanell, P. M. (2010). Resilience: Talking, resisting, and imagining new normalcies into being. *Journal of Communication*, *60*(1), 1–14. https://doi.org/10.1111/j.1460-2466.2009.01469.x

Dawson, A., & Innes, M. (2019). How Russia's internet research agency built its disinformation campaign. *The Political Quarterly*, *90*(2), 245–256. https://doi.org/10.1111/1467-923X.12690

Ferraz, T. V., & Helal, D. H. (2024). Informational resilience: a literature review. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, *22*, e024009.

Golob, T., Makarovič, M., & Rek, M. (2021). Meta-reflexivity for resilience against disinformation. *Comunicar*, *29*(66), 107–118. https://www.revistacomunicar.com/ojs/index.php/comunicar/article/view/C66-2021-09

Hall, P. A., & Lamont, M. (Eds.). (2013). *Social resilience in the neoliberal era*. Cambridge University Press.

HLEG (2018) A multi-dimensional approach to disinformation: report of the independent high-level group (HLEG) on fake news and online disinformation. *European Commission. Publications Office of the European Union*. Available at: https://blog.wan-ifra.org/sites/default/files/field_blog_entry_file/HLEGReportonFakeNewsandOnlineDisinformation.pdf

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. *The International Journal of Press/Politics*, *25*(3), 493–516. https://doi.org/10.1177/1940161219900126

Humprecht, E., Esser, F., Aelst, P. V., Staender, A., & Morosoli, S. (2023). The sharing of disinformation in cross-national comparison: Analyzing patterns of resilience. *Information, Communication & Society*, *26*(7), 1342–1362. https://doi.org/10.1080/1369118X.2021.2006744

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, *23*(5), 1301–1326. https://doi.org/10.1177/1461444820959296

Kõuts-Klemm, R., Rožukalne, A., & Jastramskis, D. (2022). Resilience of national media systems: Baltic media in the global network environment. *Journal of Baltic Studies*, *53*(4), 543–564. https://doi.org/10.1080/01629778.2022.2103162

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens versus the internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest*, *21*(3), 103–156. https://doi.org/10.1177/1529100620946707

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., … & Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094-1096. https://www.science.org/doi/10.1126/science.aao2998

Liagusha, A., & Iarovyi, D. (2025). Memes, freedom, and resilience to information disorders: Information warfare between democracies and autocracies. *Social Sciences & Humanities Open*, *11*, 101247. https://doi.org/10.1016/j.ssaho.2024.101247

Loecherbach, F., Moeller, J., Trilling, D., & van Atteveldt, W. (2020). The unified framework of media diversity: A systematic literature review. *Digital Journalism, 8*(5), 605–642. https://doi.org/10.1080/21670811.2020.1764374

Lorenz, D. F. (2013). The diversity of resilience: contributions from a social science perspective. *Natural hazards*, *67*(1), 7–24. https://doi.org/10.1007/s11069-010-9654-y

Maclean, K., Ross, H., Cuthill, M., & Witt, B. (2017). Converging disciplinary understandings of social aspects of resilience. *Journal of Environmental Planning and Management, 60*(3), 519–537. https://doi.org/10.1080/09640568.2016.1162706

Meel, P. & Vishwakarma, D. (2019). Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications (153)*, 112986.

Mlejnková, P. (2022). Issues of resilience to cyber-enabled psychological and information operations. *Vojenské rozhledy*, *31*(1), 38–50.

Moussa, M. B., Radwan, A. F., & Zaid, B. (2022). Resilience to Covid-Misinformation Among Youth in a Paternalistic Context: The Case of the UAE. *American Communication Journal*, *24*(1).

Moya, J., & Goenechea, M. (2022). An approach to the unified conceptualization, definition, and characterization of social resilience. *International Journal of Environmental Research and Public Health*, 19(9), 5746. https://www.mdpi.com/1660-4601/19/9/5746

Obrist, B., Pfeiffer, C., & Henley, R. (2010). Multi-layered social resilience: A new approach in mitigation research. *Progress in Development Studies*, *10*(4), 283–293. https://doi.org/10.1177/146499340901000402

Pantazi, M., Hale, S., & Klein, O. (2021). Social and cognitive aspects of the vulnerability to political misinformation. *Political Psychology*, *42*, 267–304. https://doi.org/10.1111/pops.12797

Parsons, T. (1968). On the concept of value-commitments. *Sociological Inquiry, 38*(2), 135–160.

Paul, J., & Barari, M. (2022). Meta-analysis and traditional systematic literature reviews—What, why, when, where, and how?. *Psychology & Marketing*, *39*(6), 1099–1115. https://doi.org/10.1002/mar.21657

Ricard, J., Yañez, I., & Hora, L. (2025). A Framework for Information Disorder: Modeling Mechanisms and Implications Based on a Systematic Literature Review. *arXiv preprint arXiv:2504.12537.*

Rød, B., Pursiainen, C., & Eklund, N. (2025). Combatting Disinformation–How Do We Create Resilient Societies? Literature Review and Analytical Framework. *European Journal for Security Research*, 1–43. https://doi.org/10.1007/s41125-025-00105-4

Romanova, T. A., Sokolov, N. I., & Kolotaev, Y. Y. (2020). Disinformation (fake news, propaganda) as a threat to resilience: approaches used in the EU and its member state Lithuania. *Baltic region*, *12*(1), 53–67. https://balticregion.kantiana.ru/en/jour/4418/13008/

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science advances*, *8*(34), eabo6254. https://www.science.org/doi/10.1126/sciadv.abo6254

Schneider, L. (1976). *Classical Theories of Social Change*. General Learning Press.

Shadmy, T. (2022). Content Traffic Regulation: A Democratic Framework for Addressing Misinformation. *JURIMETRICS J.*, *63*, 1–10.

Singh, S. R., Eghdami, M. R., & Singh, S. (2014). The concept of social vulnerability: A review from disasters perspectives. *International Journal of Interdisciplinary and Multidisciplinary Studies, 1*(6), 71–82.

Wardle, C. & Derakhshan, H. (2017) Information disorder: toward an interdisciplinary framework for research and policy making. *Council of Europe report DGI 09, 31 October. Brussels: Council of Europe*.

ANNEX 1. SOURCES USED FOR THE SYSTEMATIC LITERATURE REVIEW

Albarracín, D. (2022). Processes of persuasion and social influence in conspiracy beliefs. *Current Opinion in Psychology*, *48*, 101463.

Altay, S., & Acerbi, A. (2023). People believe misinformation is a threat because they assume others are gullible. *New Media & Society*, *26*(11), 6440–6461.

Altay, S., Nielsen, R. K., & Fletcher, R. (2023). News can help! The impact of news media and digital platforms on awareness of and belief in misinformation. *The International Journal of Press/Politics*, *29*(2), 459–484.

Anspach, N. M., & Carlson, T. N. (2022). Not who you think? Exposure and vulnerability to misinformation. *New Media & Society*, *26*(8), 4847–4866.

Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in India. *American Political Science Review*, *115*(4), 1325–1341.

Baptista, J. P., & Gradim, A. (2022). Who believes in fake news? Identification of political (a) symmetries. *Social Sciences*, *11*(10), 460.

Bargaoanu, A., & Radu, L. (2018). Fake news or disinformation 2.0? Some insights into Romanians' digital behaviour. *Romanian Journal of European Affairs*, *18*, 24–38.

Benková, Z. (2019). Media education and religion in Slovak republic. *European Journal of Science and Theology*, *15*(2), 159–170.

Billingsley, B., & Heyes, J. M. (2023). Preparing students to engage with science-and technology-related misinformation: The role of epistemic insight. *The Curriculum Journal*, *34*(2), 335–351.

Bjola, C., & Papadakis, K. (2021). Digital propaganda, counterpublics, and the disruption of the public sphere: The Finnish approach to building digital resilience. In *The World Information War* (pp. 186–213). Routledge.

Blanco, F., Barberia, I., & Matute, H. (2015). Individuals who believe in the paranormal expose themselves to biased information and develop more causal illusions than nonbelievers in the laboratory. *PloS one*, *10*(7), e0131378.

Bost, P. R., Prunier, S. G., & Piper, A. J. (2010). Relations of familiarity with reasoning strategies in conspiracy beliefs. *Psychological reports*, *107*(2), 593–602.

Boulianne, S., Tenove, C., & Buffie, J. (2022). Complicating the resilience model: a four-country study about misinformation. *Media and Communication*, *10*(3), 169–182.

Bragarnich, R. (2022). Observations on psychic vulnerability to media dissemination of false political-ideological messages (fake news). *Journal of Analytical Psychology*, *67*(2), 455–467.

Bruinenberg, H., Sprenger, S., Omerović, E., & Leurs, K. (2021). Practicing critical media literacy education with/for young migrants: Lessons learned from a participatory action research project. *International Communication Gazette*, *83*(1), 26–47.

Burns, A., Lobry de Bruyn, L., & Wilson, S. C. (2023). A Rubric Approach to Assessing Information Literacy Competency in Tertiary Curricula. *Journal of University Teaching and Learning Practice*, *20*(1), 10.

Buzzanell, P. M. (2010). Resilience: Talking, resisting, and imagining new normalcies into being. *Journal of Communication*, *60*(1), 1–14.

Caled, D., & Silva, M. J. (2022). Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. *Journal of Computational Social Science*, *5*(1), 123–159.

Collado, Z. C., Basco, A. J. M., & Sison, A. A. (2020). Falling victims to online disinformation among young Filipino people: Is human mind to blame?. *Cognition, Brain, Behavior*, *24*(2), 75–91.

Christiano, T. (2022). Algorithms, manipulation, and democracy. *Canadian Journal of Philosophy*, *52*(1), 109–124.

Claudia, M. E. (2022). Psychological effects of fake news-literature review. *Journal of Educational Sciences & Psychology*, *12*(2), 95–103.

Clementson, D. E. (2018). Susceptibility to deception in a political news interview: Effects of identification, perceived cooperativeness, and ingroup vulnerability. *Communication Studies*, *69*(5), 522–544.

Çömlekçi, M. F. (2022). Why do fact-checking organizations go beyond fact-checking? A Leap toward media and information literacy education. *International Journal of Communication*, *16*, 4563–4583.

Cook, S. (2022). Countering Beijing's Media Manipulation. *Journal of Democracy*, *33*(1), 116–130.

Cook, J., Ecker, U. K., Trecek-King, M., Schade, G., Jeffers-Tracy, K., Fessmann, J., Kim, S.,Kinkead, D., Orr, M., Vraga, E., Roberts, K. & McDowell, J. (2023). The cranky uncle game—Combining humor and gamification to build student resilience against climate misinformation. *Environmental Education Research*, *29*(4), 607–623.

Craig, M., & Vijaykumar, S. (2023). One dose is not enough: the beneficial effect of corrective COVID-19 information is diminished if followed by misinformation. *Social Media+ Society*, *9*(2). https://doi.org/10.1177/2056305123116129

De Leyn, T., Waeterloos, C., De Wolf, R., Vanhaelewyn, B., Ponnet, K., & De Marez, L. (2022). Teenagers' reflections on media literacy initiatives at school and everyday media literacy discourses. *Journal of children and media*, *16*(2), 221–239.

Deinla, I. B., Mendoza, G. A. S., Ballar, K. J., & Yap, J. K. (2022). The link between fake news susceptibility and political polarization of the youth in the Philippines. *Asian Journal of Political Science*, *30*(2), 160–181.

Di Mascio, F., Barbieri, M., Natalini, A., & Selva, D. (2021). Covid-19 and the information crisis of liberal democracies: Insights from anti-disinformation action in Italy and EU. *Partecipazione e conflitto*, *14*(1), 221–240.

Disha, E. P., Halili, A., & Rustemi, A. (2023). Vulnerability to disinformation in relation to political affiliation in North Macedonia. *Media and Communication*, *11*(2), 42–52.

Disha, E., & Rustemi, A. (2022). The Relation Between Political Indifference and Resilience to Disinformation. *Media Literacy and Academic Research*, *5*(2), 164–177.

Dolan, C. J. (2022). Hybrid warfare in the western Balkans: how structural vulnerability attracts maligned powers and hostile influence. *SEEU review*, *17*(1), 3–25.

Drotner, K. (2020). Minimizing Knowledge Scepticism–Resourcing Students through Media and Information Literacy. *European Review*, 28(S1), 56–66.

Dumitru, E. A. (2020). Testing children and adolescents' ability to identify fake news: a combined design of quasi-experiment and group discussions. *Societies, 10*(3), 71.

Dunleavy, V. O., Ahn, R., Mayo, D., & Grace, L. D. (2022). Addressing COVID-19 Misinformation and Resiliency Among Latinos Living With HIV: Formative Research Guiding the Latinos Unidos Microgame Intervention. *The American Behavioral Scientist*.

Epstein, Z., Sirlin, N., Arechar, A., Pennycook, G., & Rand, D. (2023). The social media context interferes with truth discernment. *Science Advances*, *9*(9), eabo6169.

Escolà-Gascón, Á., Dagnall, N., Denovan, A., Drinkwater, K., & Diez-Bosch, M. (2023). Who falls for fake news? Psychological and clinical profiling evidence of fake news consumers. *Personality and Individual Differences*, *200*, 111893.

Filipec, O. (2019). Towards a disinformation resilient society?: The experience of the Czech republic. *Cosmopolitan Civil Societies: an Interdisciplinary Journal*, *11*(1), 1–26.

Freedman, J., Gjørv, G. H., & Razakamaharavo, V. (2021). Identity, stability, Hybrid Threats and Disinformation. *ICONO 14, Revista de comunicación y tecnologías emergentes*, *19*(1), 38–69.

Fritts, M., & Cabrera, F. (2022). Online Misinformation and "Phantom Patterns": Epistemic Exploitation in the Era of Big Data. *The Southern Journal of Philosophy*, *60*(1), 57–87.

Frolova, E., Rogach, O., & Tyurikov, A. (2022). Student's Media Competence: New Opportunities to Counteract Information Manipulations In Network Interactions. *Media Education (Mediaobrazovanie)*, *3*, 380–389.

Golob, T., Makarovič, M., & Rek, M. (2021). Meta-reflexivity for resilience against disinformation. *Comunicar*, *29*(66), 107–118.

Greer, J., Fitzgerald, K., & Vijaykumar, S. (2022). Narrative elaboration makes misinformation and corrective information regarding COVID-19 more believable. *BMC Research Notes*, *15*(1), 1–5.

Habro, I., Vovchuk, L., & Shevchuk, O. (2020). Informational and Psychological Influence on Student Youth inthe Conditions of the Information-Psychological War. *Journal of Educational and Social Research*, *10*(1), 56–61.

Hall, H. K. (2017). The new voice of America: Countering foreign propaganda and disinformation act. *First Amendment Studies*, *51*(2), 49–61.

He, L., Yang, H., Xiong, X., & Lai, K. (2019). Online rumor transmission among younger and older adults. *Sage open*, *9*(3), 2158244019876273.

Hodson, J., & Gosse, C. (2022). Lost in the information maze: Understanding the multiple drivers of QAnon online conspiracy theories. *Canadian Journal of Communication, 47*(4). https://doi.org/10.3138/cjc.2022-07-2

Hopp, T., & Ferrucci, P. (2020). A spherical rendering of deviant information resilience. *Journalism & Mass Communication Quarterly*, *97*(2), 492–508.

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. *The international journal of press/politics*, *25*(3), 493–516.

Humprecht, E., Esser, F., Aelst, P. V., Staender, A., & Morosoli, S. (2023). The sharing of disinformation in cross-national comparison: Analyzing patterns of resilience. *Information, Communication & Society*, *26*(7), 1342–1362.

Indu, V., & Thampi, S. M. (2021). A psychologically-inspired fuzzy-based approach for user personality prediction in rumor propagation across social networks. *Journal of Intelligent & Fuzzy Systems*, *41*(5), 5425–5439.

Jackson, N. J. (2021). The Canadian government's response to foreign disinformation: Rhetoric, stated policy intentions, and practices. *International Journal*, *76*(4), 544–563.

Janmohamed, K., Walter, N., Nyhan, K., Khoshnood, K., Tucker, J. D., Sangngam, N., Altice, F., Ding, Q,,, Wong, A., Schwitzky, Z., Bauch, C., Choudhury, M., Papakyriakopoulos, O., & Kumar, N. (2021). Interventions to mitigate COVID-19 misinformation: a systematic review and meta-analysis. *Journal of Health Communication*, *26*(12), 846–857.

Jerónimo, P., & Esparza, M. S. (2022). Disinformation at a Local Level: An Emerging Discussion. *Publications*, *10*(2), 15.

Jo, H., Yang, F., & Yan, Q. (2022). Spreaders vs victims: the nuanced relationship between age and misinformation via FoMO and digital literacy in different cultures. *New Media & Society*, *26*(9), 5169–5194.

Juurvee, I., & Arold, U. (2021). Psychological Defence and Cyber Security: Two Integral Parts of Estonia's Comprehensive Approach for Countering Hybrid Threats. *ICONO 14, Revista de comunicación y tecnologías emergentes*, *19*(1), 70–94.

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New media & society*, *23*(5), 1301–1326.

Kertysova, K. (2018). Artificial intelligence and disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered. *Security and Human Rights*, *29*(1-4), 55–81.

Kirby, R., Valaskova, K., Kolencik, J., & Kubala, P. (2018). Online habits of the fake news audience: The vulnerabilities of Internet users to manipulations by malevolent participants. *Geopolitics, History and International Relations*, *10*(2), 44–50.

Klebba, L. J., & Winter, S. (2023). Crisis alert:(Dis) information selection and sharing in the COVID-19 pandemic. *Communications*, *49*(2), 318–338. https://doi.org/10.1515/commun-2022-0020

Kõuts-Klemm, R., Rožukalne, A., & Jastramskis, D. (2022). Resilience of national media systems: Baltic media in the global network environment. *Journal of Baltic Studies*, *53*(4), 543–564.

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens versus the internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest*, *21*(3), 103–156.

Krishna, A. (2021). Lacuna publics: Advancing a typology of disinformation-susceptible publics using the motivation-attitude-knowledge framework. *Journal of Public Relations Research*, *33*(2), 63–85.

Kyza, E. A. (2020). Combating misinformation online: re-imagining social media for policy-making. *Internet Policy Review, 9(4).* https://doi.org/10.14763/2020.4.1514

Lai, K., Xiong, X., Jiang, X., Sun, M., & He, L. (2020). Who falls for rumor? Influence of personality traits on false rumor belief. *Personality and Individual Differences*, *152*, 109520.

Lebid, A. E., Nazarov, M. S., & Shevchenko, N. A. (2021). Information Resilience and Information Security as Indicators of the Level of Development of Information and Media Literacy. *International Journal of Media and Information Literacy*, *6*(2), 157–166.

Lebid, A. E., Vashyst, K. M., & Nazarov, M. S. (2022). Information resilience as a means of countering the socio-psychological strategies of information wars. *International Journal of Media and Information Literacy*, *7*(1), 157–166.

Lee, N. M. (2018). Fake news, phishing, and fraud: a call for research on digital media literacy education beyond the classroom. *Communication Education*, *67*(4), 460–466.

Leurs, K., Omerović, E., Bruinenberg, H., & Sprenger, S. (2018). Critical media literacy through making media: A key to participation for young migrants?. *Communications*, *43*(3), 427–450.

Levine, T. R. (2022). Truth-default theory and the psychology of lying and deception detection. *Current Opinion in Psychology*, *47*, 101380.

Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, *32*(2), 348–384.

Liu, P. L., & Huang, L. V. (2020). Digital disinformation about COVID-19 and the third-person effect: examining the channel differences and negative emotional outcomes. *Cyberpsychology, Behavior, and Social Networking*, *23*(11), 789–793.

Loos, E., Ivan, L., & Leu, D. (2018). Save the Pacific Northwest tree octopus": a hoax revisited. Or: How vulnerable are school children to fake news? *Information and Learning Sciences, 119* (9/10), 514–528.

Lyons, B., Mérola, V., Reifler, J., & Stoeckel, F. (2020). How politics shape views toward fact-checking: Evidence from six European countries. *The International Journal of Press/Politics*, *25*(3), 469–492.

Mader, M., Marinov, N., & Schoen, H. (2022). Foreign Anti-Mainstream Propaganda and Democratic Publics. *Comparative Political Studies*, *55*(10), 1732–1764.

Mälksoo, M. (2020). Countering hybrid warfare as ontological security management: the emerging practices of the EU and NATO. In *Ontological Insecurity in the European Union* (pp. 126–144). Routledge.

Manwaring, R., & Holloway, J. (2023). Resilience to cyber-enabled foreign interference: citizen understanding and threat perceptions. *Defence Studies*, *23*(2), 334–357.

March, E., & Springer, J. (2019). Belief in conspiracy theories: The predictive role of schizotypy, Machiavellianism, and primary psychopathy. *PloS one*, *14*(12), e0225964.

Mehrad, J., Eftekhar, Z., & Goltaji, M. (2020). Vaccinating users against the hypodermic needle theory of social media: Libraries and improving media literacy. *International Journal of Information Science and Management (IJISM)*, *18*(1), 17–24.

McDougall, J. (2019). Media literacy versus fake news: critical thinking, resilience and civic engagement. *Media Studies*, *10*(19), 29–45.

McDougall, J., & Rega, I. (2022). Beyond solutionism: Differently motivating media literacy. *Media and Communication*, *10*(4), 267–276.

McGrew, S., & Chinoy, I. (2022). Fighting misinformation in college: students learn to search and evaluate online information through flexible modules. *Information and Learning Sciences*, *123*(1/2), 45–64.

Melo-Pfeifer, S., & Gertz, H. D. (2022). Transforming disinformation on minorities into a pedagogical resource: towards a critical intercultural news literacy. *Media and Communication*, *10*(4), 338–346.

Millar, B. (2022). Epistemic obligations and free speech. *Analytic Philosophy*, *65*(2), 203–222.

Miyamoto, I. (2021). Disinformation: policy responses to building citizen resiliency. *Connections: The Quarterly Journal*, *20*(2), 47–55.

Mlejnková, P. (2022). Issues of resilience to cyber-enabled psychological and information operations. *Vojenské rozhledy*, *31*(1), 38–50.

Moore, R. C., & Hancock, J. T. (2022). A digital media literacy intervention for older adults improves resilience to fake news. *Scientific reports*, *12*(1), 6008.

Moussa, M. B., Radwan, A. F., & Zaid, B. (2022). Resilience to Covid-Misinformation Among Youth in a Paternalistic Context: The Case of the UAE. *American Communication Journal*, *24*(1).

Nagler, R. H., Vogel, R. I., Rothman, A. J., Yzer, M. C., & Gollust, S. E. (2023). Vulnerability to the effects of conflicting health information: Testing the moderating roles of trust in news media and research literacy. *Health Education & Behavior*, *50*(2), 224–233.

Nyhan, B., & Zeitzoff, T. (2018). Conspiracy and misperception belief in the Middle East and North Africa. *The Journal of Politics*, *80*(4), 1400–1404.

Nguyễn, S., Moran, R. E., Nguyen, T. A., & Bui, L. (2023). "We Never Really Talked About politics": Race and Ethnicity as Foundational Forces Structuring Information Disorder Within the Vietnamese Diaspora. *Political Communication*, *40*(4), 415–439. https://doi.org/10.1080/10584609.2023.2201940

Nugraha, M. A., Banglali, N. P., Abraham, J., Ali, M. M., & Andangsari, E. W. (2020). Insights on Media Literacy and Social Engineering Vulnerability Predictors: Lifelong Learning Gravity. *Cypriot Journal of Educational Sciences*, *15*(5), 955–975.

O'Brien, T. C., Palmer, R., & Albarracin, D. (2021). Misplaced trust: When trust in science fosters belief in pseudoscience and the benefits of critical evaluation. *Journal of Experimental Social Psychology*, *96*, 104184.

Olariu, O. (2022). Critical Thinking as Dynamic Shield against Media Deception. Exploring Connections between the Analytical Mind and Detecting Disinformation Techniques and Logical Fallacies in Journalistic Production. *LOGOS, UNIVERSALITY, MENTALITY, EDUCATION, NOVELTY. Section Social Sciences*, *11*(1), 29–57.

Pantazi, M., Hale, S., & Klein, O. (2021). Social and cognitive aspects of the vulnerability to political misinformation. *Political Psychology*, *42*, 267–304.

Pehlivanoglu, D., Lighthall, N. R., Lin, T., Chi, K. J., Polk, R., Perez, E., Cahill, B. S., & Ebner, N. C. (2022). Aging in an "infodemic": The role of analytical reasoning, affect, and news consumption frequency on news veracity detection. *Journal of Experimental Psychology: Applied*. *28*(3), 468–485. https://doi.org/10.1037/xap0000426

Pérez-Escoda, A. (2022). Infodemic and Fake News Turning Shift for Media: Distrust among University Students. *Information*, *13*(11), 523.

Petukhov, A. Y., & Chuprakova, N. S. (2014). Development of vulnerability to manipulations in the consciousness of adolescents with the help of modern virtual-communicative means. *Ayer*, *1*, 49–55.

Pierre, J. M. (2020). Mistrust and misinformation: A two-component, socio-epistemic model of belief in conspiracy theories. *Journal of Social and Political Psychology*, *8*(2), 617–641.

Pottle, J. (2022). “The Greatest Flood of Mass Suggestion”: John Dewey, Propaganda, and Epistemic Costs of Social Organization. *The Journal of Politics*, *84*(3), 1515–1527.

Puebla-Martínez, B., Navarro-Sierra, N., & Alcolea-Díaz, G. (2021). Methodological proposal for the detection of the composing elements of vulnerability regarding disinformation. *Publications*, *9*(4), 44.

Rechtik, M., & Mareš, M. (2021). Russian disinformation threat: comparative case study of Czech and Slovak approaches. *Journal of Comparative Politics*, *14*(1), 4–19.

Reeves, J. (2022). A Theological Engagement with the Science of Science Skepticism. *TheoLogica: An International Journal for Philosophy of Religion and Philosophical Theology*, *6*(1), 197–214.

Rodríguez-Pérez, C., & Canel, M. J. (2023). Exploring European citizens’ resilience to misinformation: Media Legitimacy and Media Trust as predictive variables. *Media and Communication*, *11*(2), 30–41.

Rodríguez-Virgili, J., Serrano-Puche, J., & Fernández, C. B. (2021). Digital disinformation and preventive actions: Perceptions of users from Argentina, Chile, and Spain. *Media and Communication, 9*(1), 323–337.

Romanova, T. A., Sokolov, N. I., & Kolotaev, Y. Y. (2020). Disinformation (fake news, propaganda) as a threat to resilience: approaches used in the EU and its member state Lithuania. *Baltic region*, *12*(1), 53–67.

Rosca, A. (2018). Media Security Structural Indicators: The Case of Moldova. *Demokratizatsiya: The Journal of Post-Soviet Democratization*, *26*(3), 365–400.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., van der Bles, A. M. & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science*, *7*(10), 201199.

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science advances*, *8*(34). https://www.science.org/doi/10.1126/sciadv.abo6254

Rožukalne, A., Kleinberga, V., Tīfentāle, A., & Strode, I. (2022). What Is the Flag We Rally Around? Trust in Information Sources at the Outset of the COVID-19 Pandemic in Latvia. *Social Sciences*, *11*(3), 123.

Rožukalne, A., Murinska, S., & Tīfentāle, A. (2021). Is COVID-19 an ‘ordinary flu’that benefits politicians? Perception of pandemic disinformation in Latvia. *Communication Today*, *12*(2), 69–83.

Salvi, C., Iannello, P., McClay, M., Rago, S., Dunsmoor, J. E., & Antonietti, A. (2021). Going viral: How fear, socio-cognitive polarization and problem-solving influence fake news detection and proliferation during COVID-19 pandemic. *Frontiers in Communication*, *5*, 127.

Scharrer, L., Pape, V., & Stadtler, M. (2022). Watch oation. *Discourse Processes*, *59*(8), 575–590.

Scheffer, M., Borsboom, D., Nieuwenhuis, S., & Westley, F. (2022). Belief traps: Tackling the inertia of harmful beliefs. *Proceedings of the National Academy of Sciences*, *119*(32), e2203149119.

Scherer, L. D., & Pennycook, G. (2020). Who is susceptible to online health misinformation?. *American Journal of Public Health*, *110*(3), 276–277.

Sengl, M., & Heinke, E. (2023). Teaching Journalism Literacy in Schools: The Role of Media Companies as Media Educators in Germany. *Media and Communication*, *11*(2), 53–63.

Seo, H., Blomberg, M., Altschwager, D., & Vu, H. T. (2021). Vulnerable populations and misinformation: A mixed-methods approach to underserved older adults’ online information assessment. *New Media & Society*, *23*(7), 2012–2033.

Shadmy, T. (2022). Content Traffic Regulation: A Democratic Framework for Addressing Misinformation. *JURIMETRICS J.*, *63*, 1–10.

Stankutė-Søsted, R. (2019). The Susceptibility of Lithuanian Youth to Information Attacks: the Elaboration Likelihood Model and Presumable Attack Topics. *Lithuanian Annual Strategic Review*, *17*(1), 335–359.

Stein, J., Frey, V., & van de Rijt, A. (2023). Realtime user ratings as a strategy for combatting misinformation: an experimental study. *Scientific reports*, *13*(1), 1626.

Tan, C. (2022). The curious case of regulating false news on Google. *Computer Law & Security Review*, *46*, 105738.

Tripodi, F. B., Stevenson, J. A., Slama, R., & Reich, J. (2023). Libraries combating disinformation: From the front line to the long game. *The Library Quarterly*, *93*(1), 48–71.

Tsene, L. (2022). Using comics as a media literacy tool for marginalised groups: The case of Athens Comics Library. *Media and Communication*, *10*(4), 289–293.

van der Meer, T. G., Brosius, A., & Hameleers, M. (2023). The role of media use and misinformation perceptions in optimistic bias and third-person perceptions in times of high media dependency: Evidence from four countries in the first stage of the COVID-19 pandemic. *Mass Communication and Society*, *26*(3), 438–462.

Ventsel, A., Hansson, S., Rickberg, M., & Madisson, M. L. (2023). Building Resilience Against Hostile Information Influence Activities: How a New Media Literacy Learning Platform Was Developed for the Estonian Defense Forces. *Armed Forces & Society*, 0095327X231163265.

Vériter, S. L., Bjola, C., & Koops, J. A. (2020). Tackling COVID-19 disinformation: Internal and external challenges for the European Union. *The Hague Journal of Diplomacy*, *15*(4), 569–582.

Vernier, M., Cárcamo, L., & Scheihing, E. (2018). Critical thinking of young citizens towards news headlines in Chile. *Comunicar*, 54, v. XXVI, 101–110.

Walsh, R. L. (2021). An Inquisition for Propaganda and Mass Deception: Deposing the Neolithic Mind. *Frontiers in Communication*, *6*, 636292.

Wagnsson, C. (2020). What is at stake in the information sphere? Anxieties about malign information influence among ordinary Swedes. *European Security*, *29*(4), 397–415.

Wang, Y., Gao, S., & Gao, W. (2022). Investigating dynamic relations between factual information and misinformation: Empirical studies of tweets related to prevention measures during COVID-19. *Journal of Contingencies and Crisis Management*, *30*(4), 427–439.

Yan, W., & Pan, Z. (2023). Believing and Sharing False News on Social Media: The Role of News Presentation, Epistemic Motives, and Deliberative Thinking. *Media Psychology*, *26*(6), 743–766. https://doi.org/10.1080/15213269.2023.2208363

Yang, J., & Tian, Y. (2021). "Others are more vulnerable to fake news than I Am": Third-person effect of COVID-19 fake news on social media users. *Computers in Human Behavior*, *125*, 106950.

Zainab, S., Mansoor, M., & Bukhari, M. (2022). Media Literacy and Senior Adults' navigation to social media: Practices, Participation and Challenges. *Observatorio (OBS*)*, *16*.

Zimmerman, T., Njeri, M., Khader, M., & Allen, J. (2022). Default to truth in information behavior: a proposed framework for understanding vulnerability to deceptive information. *Information and Learning Sciences*, *123*(1/2), 111–126.

ANNEX 2. SEARCH PROTOCOL AND INCLUSION/EXCLUSION CRITERIA

- Timeline: from 2010 to May 2023.
- Sources: peer-reviewed articles in English within the Web of Science, Scopus, and EBSCO databases. Articles from fields unrelated to social science were excluded.

"RESILIENCE" SEARCH

- A total of 9,035 articles were found in Web of Science, 6,124 in Scopus, and 12,312 in EBSCO.
- For EBSCO, manual selection was further applied to include only articles related to the information and communication field within the context of media literacy, disinformation, and other forms of deviant information. This narrowed it down to 378 articles in the dataset.
- 15,537 articles uploaded to Rayyan, where automatic duplicate detection was applied with a threshold of 95% similarity.
- Manual removal of duplicates resulted in 4,019 duplicates removed and the identification of 2 non-duplicates, leaving 11,518 articles for further analysis.
- 10,870 were deemed irrelevant to the topic of resilience to disinformation and other information disorders, leaving 226 articles for further study.
- Articles were excluded if their subject classification belonged to domains where resilience and vulnerability refer to physical, biological, ecological, medical, or engineering processes rather than information flows, communication practices, or socio-political dynamics. Thus, records from engineering, computer science hardware/software, environmental and life sciences, medicine and clinical research, physics, chemistry, agriculture, and related technical fields were filtered out.

"VULNERABILITY" SEARCH

- A total of 8,012 articles were found in Web of Science, 7,041 in Scopus, and 20,721 in EBSCO.
- For EBSCO to include only articles related to the information and communication field within the context of media literacy, disinformation, and other forms of deviant information, resulting in 109 articles in the dataset.
- 15,162 articles were uploaded to Rayyan, where automatic duplicate detection was applied with a 95% similarity threshold.

- Manual removal of duplicates resulted in 4,007 duplicates removed and the identification of 2 non-duplicates, leaving 11,113 articles for further analysis.
- 10,834 were deemed irrelevant to the topic of resilience to disinformation and other information disorders, leaving 279 articles for further study.
- Articles were excluded if their subject classification belonged to domains where resilience and vulnerability refer to physical, biological, ecological, medical, or engineering processes rather than information flows, communication practices, or socio-political dynamics. Thus, records from engineering, computer science hardware/software, environmental and life sciences, medicine and clinical research, physics, chemistry, agriculture, and related technical fields were filtered out.

MERGING THE SUBSETS AND REMOVING THE IRRELEVANT ARTICLES

- 505 articles were placed in the dataset.
- 12 removed as duplicates (present in both the "resilience" and "vulnerability" datasets).
- 493 articles in the dataset subsequently studied by the research team.
- The aim was to focus on two sets of articles:
- Vulnerabilities – articles that describe and categorize factors contributing to susceptibility to information disorders.
- Resilience – articles that explore measures or social practices aimed at bolstering resistance against information disorders.

To achieve that, each article was reviewed by two scholars to make a decision, and if both independently decided to exclude it based on certain criteria, the article was excluded.

- the items are not from peer-reviewed journals (16);
- language is other than English (7);
- not accessible (2);
- different focus: education, digital skills and digital divide, personal development with no impact on resilience to information disorders (87);
- different focus: articles on the mental health issues with no impact on resilience to information disorders (58);
- different focus: health care communication, attitude to vaccination, health literacy with no impact on resilience to information disorders (39);
- different focus: political and security issues with no impact on resilience to information disorders (39);

- different focus: technological issues with no impact on resilience to information disorders (34);
- different focus: economics and marketing (19);
- different focus: philosophical and ethical issues with no impact on resilience to information disorders (15);
- different focus: journalism profession and personality with no impact on resilience to information disorders (14);
- too narrow: electoral risks (7), censorship (6), natural science (3), popular culture (1), online terrorism (3), comparison of the platforms (2), legal context for deep fakes (1), critique in social media (1), experts' map (1), questionable content (1), scepticism to the science (1), detection of disinformation (1), scam studies (1).

ANNEX 3. KEY WORDS FOR THE SEARCH

- "resil*" OR "vulnerab*"

AND

- disinformation
- misinformation
- malinformation
- propagand*
- communicat*
- narrative*
- media
- "deviant information"
- "information disorder*"
- hoax*
- rumor*
- rumour*
- fals*
- "false information"
- "false claim*"
- "false news"
- "fake information"
- "fake claim*"
- "fake news"
- "artificial fact*"
- manipulat*
- deceiv*
- decept*
- obfuscat*
- distort*
- disrupt*
- "information litera*"
- "information educat*"
- "information skill*"
- "media litera*"
- "media educat*"
- "media skill*"
- "digital* litera*"
- "digital educat*"
- "digital skill*"
- "critical thinking"
- "critical skill*"
- "fact check*"
- "fact-check*"
- "information threat*"
- "media threat*"
- "digital threat*"
- "information risk*"
- "media risk*"
- "digital risk*"
- "information challeng*"
- "media challeng*"
- "digital challeng*"

BOOK REVIEW

DOI: 10.51480/1899-5101.18.3(41).830

JAN STASIEŃKO (2022), *MEDIA TECHNOLOGIES AND POSTHUMAN INTIMACY*, NEW YORK: BLOOMSBURY, PAGES. 304, ISBN: 978-1-5013-8051-8

This review provides a description and recommendation of Jan Stasieńko's monograph, which is a deeply critical project. Media technologies comprised in the title serve not only for systematizing the description of mechanisms and influences but also as a frame of reference for reflecting on the language, scope, utility, and consequences of studying the human in media communication. These reflections, supported by a wealth of visuals, are conducted from a posthumanist perspective, which requires careful consideration. Consequently, the review emphasizes the theoretical layer over factual elements.

To date, posthumanism as a project of a research perspective has only been referenced once in CEJC (Kalpokas, 2020). This is surprising when taking into consideration the long tradition of a trend formed half a century ago under the influence of the critical paradigm (in the versions of R. Barthes and M. Foucault) but also in a certain opposition to technological determinism, both well-known to media studies reflection. The criticality of posthumanism appears to remain neglected potentially due to its challenge to established epistemological and ontological categories, frequently perceived as constant and strict, particularly within the empirical and deterministic paradigm of large-scale communication studies and the concurrent development of transhumanism.

In contrast to transhumanism, which focuses on the empirically tangible deterministic technologicality of transforming the human being as a species, posthumanism emphasizes individual subjectivity (not only human) as a product of discursive relationality and intimacy between humans and the *other being*. In the context of posthumanism, the concepts of human subjectivity, integrity, and rationality are not strict but rather emerge and are made tangible through discursive relationships with non-human beings and those recognized as subjective beings, namely products of nature and technology.

Previous posthumanist studies have concentrated on two categories of other beings, namely creations of nature (subjectified animals and plants) and near-human creations of technology (autonomous machines). Stasieńko proposes a new line of research, which is particularly relevant from a communicological perspective. The author identifies the primary objective of his work as an endeavor to delineate the third category of otherness—information, and to investigate the

subjective relationships between media technology users and the meanings and interpretations derived from information creations.

The incorporation of the principles of neomaterialism, which challenges conventional dichotomies between subject and object and representation-materiality opposition (pp.14–15), is one of the foundations of this approach. In the posthumanist perspective, human actions are always considered part of complex and dynamic structures, referred to as assemblages, apparatuses, or, as Stasieńko proposes, vehicles. It appears that the concept of a media vehicle most effectively represents the dynamics of communication processes between participating beings. Employing the terminology of programming theories, Stasieńko characterizes these beings as 'instances', i.e., occurrences of objects representing particular classes. The process of instantiation, or 'becoming' of a being assigned to a class, is a discursive phenomenon in which otherness and subjectivity are constructed as resulting from a dynamic power relationship. In this model of analysis, human, medium, and information become instances.

This research project can be interpreted as an attempt to answer numerous questions. These comprise: how and what kind of people do we become in our relationships with information? To what degree do we control it? By what power and what do media and the meanings created in them do to us? And where do these media vehicles in which we disintegrate our subjectivity take us?

The functioning of technology is an indispensable aspect of the media vehicle, as its mechanisms establish the frameworks and conditions of the relationship between the user of the medium and information, positioning it within a specific context within the apparatus. Under the influence of technology, blurring of the boundaries of the human subject and also its construction and deconstruction occurs in distinct ways and with varying effects. 'Being or not being posthuman, being or not being a subject, is determined by the position of the being, understood as its place in the sense of a physical or conceptually construed space occupied by beings that differ in ontological status' (p.21).

The author conceptualizes this process as an 'agreement protocol' initiated by a specific consent, namely the adoption of a dialogical, open, and sensitive position (p.22). In another section of the text, the author presents the same criteria employed as conditions for the acceptance of the posthumanist perspective: 'it is I, the human being, who become aware of the presence of a non-human other, for whom I take responsibility. (…) I will shape my 'self 'so as to be capable of connection with the 'other,' and find a plane of understanding or ontological integration' (p.18).

The model for analyzing these processes is based on the notion of the apparatus, initially proposed by M. Foucault and G. Agamben. This concept exceeds the technological dimension, functioning as a metaphor for the systemic power structure of a given discourse. Nonetheless, this analysis lacks comprehensive

contextualization of media discourse. Stasieńko's work delves more into the cybernetic aspects of the system than the social ones, exploring a discourse in the interactional dimension rather than sociocultural, examining power not merely as a means of establishing and enforcing norms but as a tool for controlling the relationships between subjects. This approach entails conceptualizing power relationships through a discursive perspective based on symbolic and normative orders—not at a micro-scale but the level of private (intimate) mini-discourses, where within an unstable balance of power and dialogicity balancing between cooperation and competition, negotiations of boundaries between the human and the non-human (informational) are held.

The author employs the normative model of subjectivity formation contained within Lacan's theory of psychoanalysis, subsequently critiqued by Deleuze and Guattari, who described protocols of destabilization and disintegration. In reference to these concepts, Stasieńko highlights that the human user's interaction with information has a formatting and destabilizing dimension on their subjectivity (which, in the metalanguage of his research, is reflected in a balance between psychoanalysis and schizoanalysis). This enables Stasieńko to extend the critical instrumentation of posthumanism to encompass concepts such as desire, embodiment, and representation, particularly in the context of visibility and visuality (pp.8–11).

In the study of communication as a relationship, it is challenging to investigate phenomena at a deeper (or closer) level than the micro-level, specifically in the context of establishing and transcending intimacy. When considering the methodology of analysis which guides to the author's unique perspective on posthumanism, this category assumes particular significance. Reconsideration of the conditions of subjectivity at the empirical level entails an investigation of individual communicative acts and the factors directly establishing them. What is the catalyst for this phenomenon? Feasibly, this is most aptly (and subtly) expressed by the author when he discusses the search for a theory of 'users' affect vis-à-vis the technologies' (p.15), which encompasses the digitized body or embodied information (the first vehicle of 'affect vis-à-vis the technologies' was cinema).

In his analyses, Stasieńko considers the occurrence in various technological vehicles of selected individual and private micro-discourses between users and the meanings generated by a given information creation. In these instances, he identifies protocols of intimacy and emotional bond as particularly noteworthy; they construct the dimension of intimate relationships as either a ritualized structure of symbolic or physical violence or as a source of pleasure (including sexual). Another significant aspect of intimacy is theatricality, which can be defined as performative and ritualized world-making (p.13).

The extensive empirical section of this monograph is comprised of materials primarily collected for the purpose of 'intimate anthropology' of the creative

process and consumption (p.13). Such materials include making-of documentation (photographs and audio recordings from the production set, animatics, concept art, and interviews with creators and actors) and confirmations regarding the reception of media messages (blogs and recorded and published statements of media audiences and users).

Stasieńko employs examples from the digital age, applying his version of media archaeology (thereby diverging from the foundations of technological determinism) to suggest that the protocols of intimacy toward the message emerged considerably earlier. The selected examples (e.g., peep shows or a teleprompter) become metametaphors—the forms and uses of particular technologies serve as metaphors for user-meaning relationship in Stasieńko's research instrumentation.

One of the author's motivations for adopting a posthumanist perspective was his aspiration to embrace historical phenomena and reorder the historical account of media. Concurrently, an investigation of the more remote past facilitated the identification of the historical sources of the concept of posthumanism and enabled the formulation of a personal interpretation of it, which is a highly reflexive project.

Reviewing Jan Stasieńko's work, it is feasible to identify several fundamental aspects which can be understood as a critical theory of communication. Initially, on numerous levels, it exemplifies critical philosophy in the post-Kantian sense, challenging and verifying a multitude of rigid assumptions associated with traditional humanities. This is particularly evident when it identifies posthumanist phenomena as manifestations of a critical stance toward anthropocentrism. Subsequently, in the analysis of micro-discourses, power relationships are identified in a manner consistent with Foucault's conceptualization as immanent to the relationship rather than determined by technology itself. Finally, it is characterized by a profound reflexivity. In formulating his research program and method of analysis, Stasieńko engaged in a productive exchange with the assumptions of diverse forms of posthumanism and their language. Due to this, he developed a set of concepts which, as in the case of not strict, relational, and processual categories, is based on multi-level metaphors and an exceptionally dense terminological network, only partially reconstructed in this review.

It is important to notice that the reflexive complexity and multi-levelness of conceptual structures may give rise to ambiguity and a degree of blurring between the discrete references pertaining to the metalanguage of analysis and the language of the subject. The posthumanist researcher talks about the posthumanist human-subject; the method is described as posthumanist and the object of analysis is sometimes presented as posthumanist; and finally, the posthumanist research perspective is simultaneously a dialogical, open, and sensitive posthumanist position of the subject. It is possible to become confused.

Nonetheless, the author's thesis regarding self-reflexive criticality as a defensive strategy against despair and anxiety in the face of the loss of anthropocentric integrity, offering the prospect of redefining humanity may be accepted: 'the measure of our humanism lies in understanding what posthumanism is in all its diverse types and contexts' (p.227).

Michał Otrocki
UNIVERSITY OF LOWER SILESIA, WROCŁAW, POLAND

REFERENCES

Kalpokas, I. (2021). Agglomerations, Relationality, and In-betweenness: Re-learning to Research Agency in Digital Communication. *Central European Journal of Communication*, 13 3(27), 426–440. https://doi.org/10.51480/1899-5101.13.3(27).7

DOI: 10.51480/1899-5101.18.3(41).893

CAREY JEWITT, SARA PRICE, KERSTIN LEDER MACKLEY, NIKOLETA YIANNOUTSOU, DOUGLAS ATKINSON (2025). *INTERDISCIPLINARY INSIGHTS FOR DIGITAL TOUCH COMMUNICATION*. HEIDELBERG: SPRINGER, PAGES: 131, ISBN: 978-3-030-24564-1

The reviewed book explores the growing field of digital touch communication, emphasizing its social, technological, and ethical implications. Published by Springer in 2025, this book is part of the *Human–Computer Interaction Series* and is a product of the *InTouch* project. The authors provide an in-depth analysis of how digital touch is transforming human interaction, focusing on emerging technologies such as virtual reality (VR), haptics, and sensor-based communication (pp. 1–6).

As digital communication evolves beyond traditional visual and auditory modalities, the book argues that touch is becoming a critical element of interaction. The authors highlight how advancements in haptic feedback, remote touch, and multisensory communication are reshaping human experiences across different fields, from healthcare to virtual collaboration. This perspective situates the book within a broader discussion on the future of human-computer interaction and its societal impacts.

This review evaluates the book's structure, key arguments, and contributions to the field. It assesses how effectively the authors present their interdisciplinary approach and whether the book successfully bridges the gap between theoretical exploration and practical application. Additionally, the review considers the book's relevance to academics, designers, and policymakers interested in the ethical dimensions of digital touch communication.

The book is organized into eight chapters, beginning with an introduction to digital touch communication. The first chapter establishes the significance of touch in human communication and its transition to the digital realm. The authors introduce key concepts such as "digital touch," "haptic communication," and "remote tactile interaction" (pp. 1–6). These foundational ideas set the stage for understanding how touch-based interactions have evolved from direct physical contact to technologically mediated experiences, providing a framework for the discussions in subsequent chapters.

The second chapter discusses interdisciplinary approaches to studying digital touch, highlighting the multimodal and sensory methodologies employed in research (pp. 23–35). The authors emphasize the need for interdisciplinary

collaboration, incorporating insights from sociology, psychology, and human-computer interaction. The chapter presents a framework that integrates multimodality and sensory ethnography, arguing that these approaches enable a richer understanding of digital touch experiences. "We understand the physical, material, and sensory aspects of touch as a part of when and how touch-based resources are taken up (or excluded) and how they can shape—or are shaped by—people to become semiotic resources" (p. 24).

The third chapter maps out various applications of digital touch technology, examining human-human, human-robot, and human-object interactions (pp. 39–50). The authors present case studies showcasing how haptic technology is being used in diverse sectors, such as remote healthcare, social robotics, and immersive virtual environments. The chapter also highlights emerging and speculative touch technologies that push the boundaries of how humans interact with the digital world. "We look beyond our everyday interaction with touch screens, to focus on emergent and semi-speculative touch technologies that 'want us' and 'make' us want to be able to touch and feel objects in new ways: from tangibles, wearables, haptics for virtual reality, through to the tactile internet of skin" (p. 40).

The fourth chapter explores the role of social norms in shaping digital touch practices (pp. 57–69). The discussion centers on cultural differences, ethical concerns, and user acceptance of touch-based technologies. The authors argue that social norms around touch are neither static nor universal but evolve through societal and technological changes. "Social norms are in a state of continual flux, tension, and negotiation pulled across these sites of life, they are (simultaneously) fluid and fixed" (p. 59).

The fifth chapter addresses the psychological and emotional aspects of touch presence, absence, and connection (pp. 73–86). The authors examine how digital touch can bridge emotional gaps in remote communication, referencing studies on parental bonding and long-distance relationships. They discuss the "social shaping" perspective, which acknowledges that technology and social relations mutually influence each other, leading to the emergence of new ways to conceptualize presence and absence (p. 74).

The sixth chapter introduces the concept of "sociotechnical imaginaries," illustrating how digital touch is envisioned and designed within various societal and technological frameworks (pp. 89–104). The chapter defines sociotechnical imaginaries as "collectively held and performed visions of desirable futures… animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology" (p. 90).

The seventh chapter delves into ethical considerations surrounding digital touch, including consent, trust, and data privacy (pp. 107–119). The authors emphasize that "ethical touch technologies" must be designed with a focus

on human values, taking into account issues of agency, harm, ownership, and transparency (p. 108). They explore how mediated touch can create complex ethical dilemmas, particularly in remote personal touch communication and virtual touch, where the sender and receiver may not have equal control over the interaction. The discussion also highlights the challenges of ensuring consent in digital touch interactions, where "touch does not have to be synchronous, reciprocal, or bidirectional" (p. 116). Furthermore, the book raises critical questions about the collection and use of tactile data, stressing the need for policies that safeguard privacy and prevent misuse of sensitive touch-based information.

Finally, the eighth chapter provides a thought-provoking conclusion by synthesizing key insights from previous discussions and offering a roadmap for future research and design in the field of digital touch communication (pp. 123–131). The chapter underscores the importance of adopting a socially grounded perspective in understanding digital touch, emphasizing that „recognizing and benefiting from the potentials for difference and cultural flexibility" is essential in the development of inclusive and effective digital touch technologies. The authors reflect on the evolving landscape of touch-based interactions, highlighting the ways in which social norms, ethical considerations, and sociotechnical imaginaries shape the design and implementation of these technologies. This perspective is crucial as digital touch moves beyond the experimental phase and becomes integrated into everyday human-computer interactions.

Furthermore, the chapter presents an emergent research and design framework that serves as a guide for interdisciplinary collaboration in digital touch communication. The authors advocate for methodological approaches that are multimodal and interdisciplinary, ensuring that technological innovations align with real-world social and cultural dynamics. They state that "the thematic directions, emergent ideas, and provisional framework… highlight both the speculative and emergent character of digital touch communication," signaling the need for continued exploration and iteration in this field. The chapter concludes with a call for further research that considers ethical dimensions, user experiences, and the broader implications of digital touch in various domains, from healthcare to education and social connectivity. By providing these insights, the authors not only summarize the book's key contributions but also inspire future inquiries into the complex interplay between technology, touch, and human interaction.

The book's strength lies in its interdisciplinary approach, effectively integrating theoretical and practical perspectives. The authors draw from diverse fields, including cognitive science, user experience research, and social theory, to provide a holistic understanding of digital touch communication. The inclusion of case studies, such as *In Touch with Baby* (pp. 84–86) and *Remote Contact*

Exhibition (pp. 81–82), enriches the discussion by offering real-world applications of digital touch.

However, one limitation of the book is its predominantly Western perspective. While the authors acknowledge cultural variations in touch norms (pp. 57–59), a more in-depth exploration of non-Western perspectives on digital touch could enhance the book's global relevance. Additionally, although the book provides a comprehensive overview of ethical concerns, it could benefit from more concrete policy recommendations for governing digital touch technologies (pp. 115–117).

Sokhi Huda
STATE ISLAMIC UNIVERSITY (UIN) SUNAN AMPEL SURABAYA

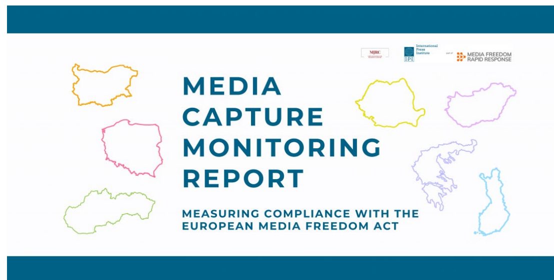
EVENT

MARIUS DRAGOMIR WINS THE MEDIA AND DEMOCRACY KAROL JAKUBOWICZ AWARD 2025

Marius Dragomir – the editor of the Media Capture Monitoring Report project – received the Media and Democracy Karol Jakubowicz Award 2025 for the Media Capture Monitoring Report, a research report by the Media and Journalism Research Center and the International Press Institute (co-authored with Zsuzsa Detreköi). A special Media and Democracy Karol Jakubowicz Award 2025 was awarded to Basil Kerski for his book *Europejczycy z kantonu Polska* (eng. *Europeans from the Canton Poland*; Kraków; Budapeszt; Syrakuzy: Wydawnictwo Austeria, 2024).

The Selection Committee awarded Marius Dragomir for the originality and thoroughness of his research, offering a systematic and pioneering perspective on media ownership, regulatory frameworks, and the complex relationships between media and politics. The *Media Capture Monitoring Report* is a joint initiative of the Media and Journalism Research Center and the International Press Institute, and the report was co-authored with Zsuzsa Detreköi.

Photo 1. Cover of *Media Capture Monitoring Report*. Vienna, Austria: International Press Institute



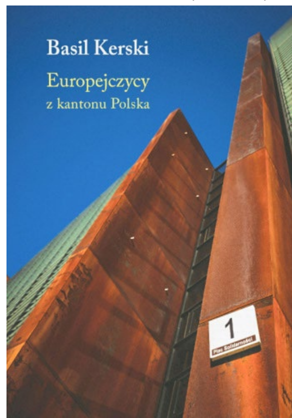
Dr. Marius Dragomir is the Director of the Media and Journalism Research Center, a researcher at the University of Santiago de Compostela and a visiting professor at Central European University (CEU) in Vienna, specializing in journalism, research design, advocacy, and media policy. Formerly Director of the Center for Media, Data and Society (CMDS), he has spent over two decades designing and managing global media research initiatives. He leads projects like the Media Influence Matrix, the State Media Monitor, and is the editor of the Media Capture Monitoring Report, a pioneering study of media ownership and

political influence. A former journalist in Romania, Dr. Dragomir also founded Media Power Monitor to explore the intersections of media, business, and politics worldwide.

Photo 2. Dr. Marius Dragomir. Source: https://mariusdragomir.com/

Moreover, the Selection Committee recognized *Europejczycy z kantonu Polska* by Basil Kerski with a special Media and Democracy Karol Jakubowicz Award 2025. The collection of essays reflects on the determinants of the crisis of democracy and the responsibility of the media and journalists for shaping perceptions of reality and public sentiment. Kerski highlights the cultural values of Europe as a foundation for redefining the European Community in the face of global geopolitical changes.

Photo 3. Book cover of Kerski B. (2024). *Europejczycy z kantonu Polska* (eng. E*uropeans from the Canton Poland*; Kraków; Budapeszt; Syrakuzy: Wydawnictwo Austeria, 2024)



The Media and Democracy Karol Jakubowicz Award continues to recognize outstanding scholarly and journalistic contributions to the study of media systems, media policies, media ethics, and public service media — with a strong emphasis

on their societal impact, in Central and Eastern Europe, and beyond. A full list of this year's nominees can be found on the Polish Communication Association's website: https://www.ptks.pl/en/awards/the-media-and-democracy-karol-jakubowicz-award.

The Award Ceremony took place during the 7th Congress of the Polish Communication Association, dedicated to the theme of the human being at the center of contemporary media transformation. The Congress was held in Katowice (Poland) on September 22–25, 2025, with the Award Ceremony scheduled for the first day of the event.

Photo 4. Dr. Marius Dragomir receives the Award from Prof. I. Hofman during the VII Congress of the PCA on September 22, 2025 at the Silesian Museum in Katowice. Photo by: Marcel Goldmann (University of Silesia in Katowice)

Photo 5. Basil Kerski receives the Special Award from Prof. Mariusz Kolczyński during the VII Congress of the PCA on September 22, 2025 at the Silesian Museum in Katowice (right). Photo by: Marcel Goldmann (University of Silesia in Katowice)

On behalf of the Committee and all the stakeholders, we warmly congratulate the Winners!

Dagmara Sidyk-Furman and Michał Głowacki
UNIVERSITY OF WARSAW, POLAND

AUTHORS

Márton Bene (PhD) is a research professor at the ELTE Centre for Social Sciences, and an assistant professor at the Faculty of Law, ELTE, Eotvos Lorand University. He is also the leader of MTA–HUN-REN TK Lendulet "Momentum" PRISMA Research Group. His research interests are in political communication, social media and politics, and political behavior. Email: bene.marton@tk.hu

Ianis Bucholtz, Dr. sc. comm. (Vidzeme University of Applied Sciences, Valmiera, Latvia) is a Professor and leading researcher at the Social, Economic, and Humanities Research Institute. His research centres on journalism, online publics, and media literacy. E-mail: ianis.bucholtz@va.lv

Lucie Čejková is a junior researcher and a PhD candidate at the Department of Media Studies and Journalism, Faculty of Social Studies, Masaryk University. She focuses on conspiracy thinking among news media audiences with a particular interest in media and political attitudes.

Dren Gërguri, Ph.D. (University of Prishtina, Prishtina, Kosovo) is a Lecturer in the Department of Journalism, Faculty of Philology, a Fulbright scholar and former TV journalist. His research agenda is at the intersection of disinformation, journalism, political communication, and AI. Email: dren.gerguri@uni-pr.edu

Marju Himma, Ph.D. (University of Tartu, Tartu, Estonia) is an Associate Professor of journalism studies at the Institute of Social Studies. Her research combines studies on news work, information disorder, and science communication. E-mail: marju.himma@ut.ee

Dmytro Iarovyi, Ph.D. (Vytautas Magnus University, Kaunas, Lithuania; Kyiv School of Economics, Ukraine). Dmytro is writing his second Ph.D. in Communication and Information at VMU, and serves as Academic Director of PhD in Public Policy at KSE. His research and teaching draws on resilience to information disorders, political communication and cognitive security. Email: dmytro.iarovyi@vdu.lt; diarovyi@kse.org.ua.

Kristina Juraitė, Ph.D. (Vytautas Magnus University, Kaunas, Lithuania) is a professor and head of the Department of Public Communications. She is also a cofounder of the UNESCO Chair on Media and Information Literacy for Inclusive Knowledge Societies. Her research focuses on digital media and social change, communication ethics and accountability, media and information literacy, civic agency and digital resilience. Email: kristina.juraite@vdu.lt

Maia Klaassen, MA (University of Tartu, Tartu, Estonia), is a Junior Research Fellow in Media Literacy at the Institute of Social Studies. Her doctoral research focuses on information resilience and interventions for countering information disorders in the context of ever-changing variables in the information ecosystem. E-mail: maia.klaassen@ut.ee

Jacek Kołodziej, professor at the Jagiellonian University, in the Department of Research on Social Change in Europe, Institute of European Studies. In his research, he explains processes of social and political change based on selected concepts from the fields of social communication, media, and politics. Email: jacek.kolodziej@uj.edu.pl

Ragne Kõuts-Klemm, Ph.D. (University of Tartu, Tartu, Estonia) is an associate professor of journalism sociology at the Institute of Social Studies. Her research focuses on media use and media systems in transforming societies. Email: ragne.kouts@ut.ee

Gergely Ferenc Lendvai, Ph.D. candidate in scientometrics and sociology of science at the Ludovika University of Public Service, Budapest, Hungary. His research examines sociological theory, computational social sciences, human rights, and disinformation. Email: lendvai.gergely.ferenc@uni-nke.hu

Krista Lepik-Verliin, Ph.D. (University of Tartu, Tartu, Estonia) is a lecturer in information studies at the Institute of Social Studies. Her scientific interests include information literacies, librarianship, and the role of memory institutions in social and digital transformations. E-mail: krista.lepik@ut.ee

Darren Lilleker P.h.D. (Bournemouth University, Poole, UK) is Professor of Political Communication in the Faculty of Media, Science and Technology. His research covers the production and impacts of political communication in all its forms. Email: dlilleker@bournemouth.ac.uk

Alena Pospíšil Macková is an assistant professor at the Department of Media Studies and Journalism, Faculty of Social Studies, Masaryk University. Her research focuses on political communication, social media, news consumption, and political attitudes.

Todd Nesbitt, Ph.D. (University of New York in Prague) works in the School of International Relations and Media Studies. His research focuses on political communication, propaganda, and media ownership. Email: tnesbitt@unyp.cz

Martina Novotná is an assistant professor at the Department of Media Studies and Journalism, Faculty of Social Studies at Masaryk University. Her research interests include online political communication and participation, with an emphasis on the roles of polarization, incivility, and intolerance. Email: mnovotna@fss.muni.cz

Veronika Patkós (PhD) is a Senior Research Fellow at the ELTE Centre for Social Sciences in Budapest and an Associate Professor at Eötvös Loránd University, Faculty of Law. Her research focuses on comparative politics and voting behavior, with particular emphasis on mass polarization and partisanship.

János Tamás Papp, Ph.D. is an Assistant Professor at the Faculty of Law and Political Sciences, Pázmány Péter Catholic University, Budapest. His research focuses on press and media law, with a particular focus on how online platforms shape democratic discourse. Email: papp.janos.tamas@jak.ppke.hu

Tess Slavickova, PhD (University of New York in Prague) works in the School of International Relations and Media Studies. Her research focuses on political discourse and American presidential rhetoric. Email: tslavickova@unyp.cz

Tamás Attila Szikora is a researcher at the Institute of the Information Society, Ludovika University of Public Service, Budapest. His research focuses on the impact of social media platforms on democratic discourse, particularly during election campaigns. Email: szikora.tamas@uni-nke.hu

Sten Torpan is a lecturer in crisis sociology at the University of Tartu. His work bridges classical sociology and communication research with computational social science in the domain of disaster research and communication studies. He's focused on developing training methods and scientific simulation methodologies to improve risk and crisis communication practice. Email: stentor@ut.ee

Krzysztof Wasilewski is an associate professor of political science and media studies at the Faculty of Humanities, Koszalin University of Technology, Poland. His research interests include digital and alternative media, media framing, political communication, and the American media system. Email: krzysztof.wasilewski@tu.koszalin.pl

Veronika Zavřelová, Ph.D. (University of New York in Prague) works in the School of International Relations and Media Studies. Her research focuses on applied linguistics and intercultural communication. Email: vzavrelova@unyp.cz

CENTRAL EUROPEAN JOURNAL OF COMMUNICATION (ISSN 1899-5101)
CEJC.PTKS.PL

Central European Journal of Communication is an Open Access journal which welcomes online submissions. The website cejc.ptks.pl offers information about Publication Ethics and Submission Guidelines for authors, editors and reviewers.

Central European Journal of Communication is published by the Polish Communication Association. It is financed by the Polish Communication Association and the following supporting institutions:

- Faculty of Management and Social Communication, Institute of Journalism, Media and Social Communication, Jagiellonian University, Cracow, Poland;
- Faculty of Philology, University of Wrocław, Poland;
- Faculty of Theology, Institute of Media Education and Journalism, Cardinal Stefan Wyszyński University, Warsaw, Poland;
- Faculty of Journalism, Information, and Book Studies, University of Warsaw, Poland;
- Institute of Media, Journalism, and Social Communication, University of Gdańsk, Poland;
- Institute of Social Communication and Media, Maria Curie-Skłodowska University, Lublin, Poland;
- Jan Kochanowski University, Kielce, Poland;
- University of Warmia and Mazury, Olsztyn, Poland.

Central European Journal of Communication is indexed in the following databases: ARIANTA, BazHum, CEEOL, CEJSH, ERIH PLUS, EBSCO, SCImago Journal & Country Rank, SCOPUS (CiteScore, SJR, SNIP) and Web of Science Core Collection (Emerging Sources Citation Index), as well as scored on the list of scientific journals by the Ministry of Science and Higher Education, Poland.

Central European Journal of Communication
ul. Uniwersytetu Poznańskiego 5, 61-614 Poznań, Poland
cejc.ptks.pl
journal@ptks.pl
The Polish Communication Association
residental address: ul. Koszarowa 3, 51-149 Wrocław, Poland
correspondence address: ul. Głęboka 45, 20-612 Lublin, Poland
www.ptks.pl
office@ptks.pl

Central European Journal of Communication

www.cejc.ptks.pl

Central European Journal of Communication is published twice a year (in Spring and Fall) by the Polish Communication Association. It engages in critical discussions on communications and media in Central and Eastern Europe and beyond. CEJC welcomes submissions of both theoretical and empirical research from a wide range of disciplinary approaches. We also publish papers on methods and concepts, book reviews, conference reports, interviews with scholars and media practitioners (policy-makers, media managers and journalists). The journal is indexed in several scientific databases, including SCOPUS, Web of Science Master Journal List (Clarivate), Central and Eastern European Online Library, Central European Journal of Social Sciences and Humanities and on the list of scientific journals of the Ministry of Education and Science in Poland.

ISSN 1899-5101