

János Tamás Papp – Levente Nyakas

Understanding online disinformation



National Media and Infocommunications Authority | Hungary
Media Council • Institute for Media Studies

Understanding online disinformation

Concepts, mechanisms, and policy responses

Sorozatszerkesztő:

Koltay András – Nyakas Levente: 2012–2019

Nyakas Levente – Szadai Károly: 2020–

Understanding online disinformation

Concepts, mechanisms, and policy responses

JÁNOS TAMÁS PAPP – LEVENTE NYAKAS

Institute for Media Studies

2026

The research was supported by:



All rights reserved.

© Mónika Andok, Péter Bányász, Gergely Ferenc Lendvai, Zsuzsanna Monori, Péter Nemere,
Levente Nyakas, Ronan Ó Fathaigh, János Tamás Papp, Andrej Škol kay, 2026
© Institute for Media Studies of the Media Council of the National Media and Infocommunications
Authority, 2026

Table of contents

Introduction	9
<i>Concepts, dynamics, and governance of online disinformation</i>	
JÁNOS TAMÁS PAPP	
Mapping the landscape of fake news definitions	13
JÁNOS TAMÁS PAPP	
1. Introduction	13
2. The renaissance of fake news	13
3. The concept of the term fake news	16
4. Divergent interpretations of fake news	17
5. Summary of definitions and taxonomy of fake news	19
6. Misinformation and disinformation	21
7. Conclusion	23
Disinformation from a communication and media studies perspective	27
MÓNIKA ANDOK	
1. Introduction	27
2. Fake news and the ritual model	29
3. A new media ecosystem	34
4. Process analyses	36
5. Media bias and trust in the news media	39
6. Research on the current situation in 2023	41
7. Artificial Intelligence and disinformation	46
8. Conclusions	49
Disinformation and its legal implications related to freedom of expression	55
<i>Four issues for consideration regarding regulation and conceptualization</i>	
GERGELY FERENC LENDVAI	
1. Introduction	55

2. Four issues to consider – from the pervasive nature of conceptualization to the Facebook Oversight Board	56
3. Unmasking disinformation: Striking at the heart of freedom of expression – Four inquiries to consider	57
3.1. Conceptual challenges	57
3.2. Liability and platform regulation	58
3.3. Alternative solutions: ‘platform courts’	61
3.4. The phenomenon of chilling effect	63
4. Conclusion	64

The relevance of psychological operations in social media – Conceptual overview **69**

Part 1

PÉTER BÁNYÁSZ

1. Introduction	69
2. Literature review	70
2.1. Social media	70
2.2. Fake news	75
2.3. Misinformation, disinformation and malinformation	82
2.4. Psychological operations	86
3. The future	89

The relevance of psychological operations in social media – Findings from empirical research **101**

Part 2

PÉTER BÁNYÁSZ

1. Introduction	101
2. Methodology	101
3. Results	102
3.1. Scientometrics	102
3.2. COVID-19	117
3.3. Russian-Ukrainian war	123
4. Conclusion	142

The European Union’s online disinformation regulation in practice **141**

LEVENTE NYAKAS – PÉTER NEMERE

1. Introduction	141
2. Characteristics of the European information space	141
2.1. The well-informed citizenry and democracy	142
2.2. The news media as the custodian of the European public sphere	144
2.3. Seeking the truth through ongoing debate	145

2.4. Platformization, social editors and the information crisis – challenges of the deliberative democracy model	146
3. The Practice of online disinformation regulation	150
3.1. The European Union’s approach – policy steps to tackle disinformation	150
3.2. The Code of Practice on Disinformation	152
3.3. The Digital Services Act	157
3.4. The European Media Freedom Act (EMFA)	159
3.5. Transparency and Targeting of Political Advertising (TTPA)	161
4. Conclusions	162

Regulation of online disinformation in the Netherlands **167**

RONAN Ó FATHAIGH

1. Introduction	167
2. Regulatory background	167
2.1. Government-wide strategy for tackling disinformation	168
2.2. Ministry of the Interior as a “trusted flagger” for reporting disinformation	169
2.3. Global Declaration on Information Integrity Online	171
2.4. Dutch Code of Conduct Transparency Online Political Advertisements	172
2.5. Regulation on state-owned media outlets engaged in disinformation and propaganda	173
2.6. Digital Services Act Implementation Bill	175
2.7. Dutch Advertising Code and Code for Social Media & Influencer Marketing	176
2.8. Guidelines of the Press Council	176
3. Practice	177
3.1. Facebook accounts shut down over COVID-19 disinformation	177
3.2. Parliamentary videos removed from YouTube	178
3.3. Politicians social media accounts shut down for disinformation	180
4. Conclusion	181

Case study: Slovakia **183**

ANDREJ ŠKOLKAY

1. Introduction	183
2. Regulatory overview	186
2.1. State regulation	188
2.2. Conceptual regulatory documents	200
2.3. Self and co-regulation of online disinformation	202
3. Institutions/stakeholders involved in tackling disinformation: issues, powers, obligations, tools and procedures	210
3.1. State authorities	210
3.2. Non-governmental and private entities	214
4. Practice	217
4.1. Case study: Constitutional challenges with the ban on ‘alternative’ news and current affairs portals in early 2022	217

4.2. Case study: Globsec and its methodological problem with the ‘vulnerability index’	214
4.3. Case study: Educational activism which turns into disseminating disinformation	215
4.4. State authorities and dissemination of fake news online	216
4.5. NGO Konšpirátori.sk and controversial assessment of online news webportal <i>Standard.sk</i>	216
5. Conclusions	217

French online disinformation regulation in practice **223**

PÉTER NEMERE

1. Introduction	223
1.1. How do the French become informed?	224
1.2. The regulation of fake news before the Infox Law	224
1.3. Disclosing false information in France	225
2. The changes introduced by the Infox Law	226
2.1. The Infox Law	226
2.2. The intersections of EU legislation and the Infox Law	230
2.3. Platform annual reports	232
3. Other measures under the Infox Law	235
3.1. Ban of Russian media outlets in France	235
3.2. Judicial proceeding under the Infox Law	236
3.3. Other initiatives to counter online disinformation	238
4. Conclusions	239

Scaremongering – Criminal law responses to fake news and the problem of restrictions on freedom of expression defending public peace in Hungary **243**

ZSUZSANNA MONORI

1. Basic principles of criminal law restrictions on freedom of expression	244
1.1. The constitutional nature of freedom of expression and its restrictions	244
1.2. The unique role of criminal law in the Hungarian legal system – the last resort	246
1.3. Criminal law restrictions on the grounds of public interests	248
2. Scaremongering – a controversial regulation against fake news	250
2.1. Defining the crime of fake news: the statutory elements of scaremongering	251
2.2. Critiques and the constitutionality	253
2.3. Examples from the legal practice	255
3. Concluding thoughts	259

Introduction

Concepts, dynamics, and governance of online disinformation

JÁNOS TAMÁS PAPP

In recent years, the integrity of public communication has become a focus of scholarly concern, not because societies have suddenly discovered the existence of false statements, but because the conditions under which false claims attain visibility, credibility, and political consequence have been structurally reconfigured by digital infrastructures which prioritize speed and engagement over deliberative verification. Within this transformed environment, the general function historically associated with journalism—namely the production of timely and accountable knowledge about matters of public interest—is increasingly contested by a dense ecology of actors, formats, and distribution channels, which are able to imitate journalistic appearance while remaining detached from journalistic responsibility; thereby complicating the practical distinction between information and manipulation.

The notion of ‘fake news’ has emerged as a commonly used label within this context, yet its analytical usefulness is constrained by the political and rhetorical trajectories which have given it prominence. This is because it is frequently used not only to designate demonstrably fabricated content, but also to delegitimize unfavourable reporting or to reframe normative disagreements. For this reason, the starting point of this book is a disciplined scepticism toward the term as a self-evident category, coupled with an insistence that productive inquiry requires explicit articulation of definitions, of the mechanisms presumed to connect communicative practices to social harms, and of the institutional contexts within which both the harms and the proposed remedies are constituted.

Contemporary debate is often framed as a conflict between truth and falsehood; however, the crux of the matter is more the fragility of social processes which enable diverse actors to converge, at least provisionally, on shared reference points for evaluating contested claims. These processes include not only fact-checking and editorial accountability, but also trust among institutions and the technical mediation of attention through recommendation systems and platform design. When these processes are disrupted, the consequence is not necessarily universal gullibility, but rather a fragmentation of authority, in which competing accounts persist without any mutually recognized adjudication procedures, and in which the strategic exploitation of uncertainty can become more consequential than the production of any single false statement.

Accordingly, this book approaches disinformation and related forms of information disorder as phenomena which are simultaneously social, technological, and legal, and which therefore require analytical frameworks capable of connecting micro-level features of messages to macro-level patterns of mobilization, institutional trust, and democratic resilience.

It is in this integrative spirit that the arguments set out examine the logic behind production of misleading content, the distribution dynamics that render certain narratives disproportionately visible, the audience practices through which such content is interpreted and recirculated, and the governance instruments that aim to mitigate harm while remaining compatible with freedom of expression and with democratic pluralism.

A central challenge addressed throughout is that ‘fake news’ has been used, in scholarly and public discourse alike, as a convenient umbrella for practices that differ in intention, structure, and effect, including deliberate fabrication designed to deceive, strategic propaganda, opportunistic clickbait, context-stripping and recombination of authentic materials, satire and parody, as well as negligent or careless dissemination of inaccuracies that may nevertheless have serious consequences. This heterogeneity has direct analytical implications: if the object of study is defined too broadly, it risks collapsing distinct mechanisms into a single undifferentiated category and thereby obscuring causal pathways; if it is defined too narrowly, it risks excluding forms of manipulation that are empirically salient and socially harmful, and may also misalign with the operational categories employed by policymakers, platforms, and civil society actors. For these reasons, the opening contributions provide a substrate for the definitional landscape and develop taxonomies to clarify which elements recur across influential definitions, most notably the relationship between falsity and news-like presentation, the role of intentional deception, and the strategic mimicry of journalistic conventions which functions as a substitute for credibility in digital environments.

The book’s interdisciplinary approach is motivated by the observation that no single discipline can exhaustively account for the phenomenon’s generation, diffusion, reception, and governance. Communication and media studies contribute theories of credibility, representation, and ritualized news consumption, which explain why superficially journalistic forms can function as persuasive signals even when institutional guarantees are absent, and why audiences may rely on inherited authenticity cues, which can easily be reproduced by malicious actors in networked spaces. Legal scholarship contributes frameworks for evaluating interventions under freedom of expression standards and for distinguishing between unlawful content and content which is harmful yet lawful – a distinction that becomes critical when governance instruments target systemic patterns of dissemination, as opposed to individual items of speech.

This book places disinformation within broader security and strategic contexts, including the study of psyops and coordinated influence practices, where misleading narratives may be deployed not merely to persuade on a single issue, but to erode trust in institutions, intensify polarization, or disrupt collective responses during crises. Case-oriented and empirical work presented here illustrates how major events of the 2020s, including the pandemic and war-related information environments, created conditions in which uncertainty, fear, and information overload have increased vulnerability to manipulation, while also complicating the task of distinguishing between erroneous claims, deliberate deception, and politically motivated framing.

A distinctive feature of the contemporary landscape, and a recurring theme in this book, is the acceleration of synthetic content production through generative artificial intelligence, which is reducing the marginal cost of fabricating plausible text, imagery, and audio, thereby changing the operational constraints under which disinformation campaigns can be

mounted. When persuasive linguistic fluency and convincing identity cues can be produced at scale, the credibility problem shifts from the detection of falsehoods to the governance of an ecosystem in which authenticity signals are increasingly contestable, and in which the standard of proof for reliable verification may be practically unattainable for ordinary users, who are confronted with information overload. Therefore, this book treats AI not only as a new instrument for manipulators, but also as a dual-use technology which is able to support detection, verification, and research, while simultaneously expanding the range of deceptive techniques and complicating attribution, intent assessment, and proportionality analysis in regulatory contexts.

Against this background, the book presents disinformation governance as a problem of institutional design under conditions of uncertainty, with regulators needing to balance competing goods and where the central risk is not only the persistence of false beliefs, but the erosion of shared epistemic reference points necessary for democratic decision-making. Contributions addressing regulatory practice emphasize that interventions can themselves become politically contested, particularly when measures are perceived as discriminating against particular views, when legal bases are ambiguous, or when enforcement relies on cooperation with private platforms, whose incentives do not always align with public values. The resulting governance landscape is characterized by layered instruments, including hard-law obligations, co-regulatory codes, transparency and reporting duties, and the development of institutional actors tasked with monitoring and coordination, each of which raises distinct questions of accountability, effectiveness, and compatibility with guarantees for freedom of expression.

The internal logic of this book follows from these premises. It begins by detailing the conceptual apparatus needed to distinguish various forms of information disorder and to clarify why the definitions chosen matter for research design and policy calibration. It then examines the socio-technical environment in which disinformation circulates, focusing on the construction of credibility, attention dynamics, and audience practices, which shape receptivity and retransmission. Subsequent chapters address methodological approaches for studying disinformation, including empirical mapping and case-oriented analyses, and develop accounts of coordinated influence practices which connect micro-level content features to macro-level strategic objectives such as polarization and distrust in institutions. Finally, the book turns to governance, analyzing legal frameworks and regulatory practice, with particular attention to developments in Europe and national-level implementation, which illustrate both the promise and the limitations of current approaches in addressing systemic risks without normalizing broad restrictions on speech.

This book's guiding assumption is that regulation and governance in this field have to be understood as a problem of institutional design under conditions of uncertainty, because both overreaction and underreaction carry structural costs: overly broad legal categories risk creating chilling effects and shifting decisions into opaque private enforcement, whereas an exclusive reliance on voluntary measures may fail to address systemic amplification mechanisms and repeated campaign dynamics. In this respect, the book treats freedom of expression not as a rhetorical counterweight to regulation, but as a constitutive constraint that has to be translated into concrete safeguards, procedural guarantees, and proportionality-sensitive tools, especially where interventions target lawful but harmful forms of communication.

In order to move beyond abstract principles, this book includes a dedicated set of case studies, which reconstruct how different European jurisdictions have confronted online disinformation through their own legal doctrines, administrative practices, and political contexts. The case-study section explicitly places comparative analysis front and centre, while demonstrating that the same European-level pressures, including the Digital Services Act, can lead to markedly different national choices in terms of competences, enforcement styles, and the balance between public authority and platform responsibility. The Slovak case study examines regulatory and practical responses from 2018–2023, and highlights how initiatives aimed at combating disinformation can become constitutionally controversial, particularly where website blocking and other restrictive measures appear without clear legal bases, transparent criteria, or robust procedural protections; this chapter also discusses how subsequent regulatory changes linked to the Digital Services Act expand the competence of the media regulator and reshape the institutional setting for future interventions. By contrast, the Dutch case study presents a framework in which policy explicitly places freedom of expression and freedom of the press at the centre, and where the state rejects the idea that qualifying content as disinformation and fact-checking should be primary government duties; it then shows how, within a largely self-regulatory environment, courts have nevertheless played a significant role in disputes involving platforms, including cases in which judicial decisions required platforms to restore accounts when due-process guarantees were inadequate. The Hungarian and French chapters add further comparative depth by illustrating different pathways of legal and regulatory engagement with false or misleading claims. The Hungarian analysis discusses, among other issues, the criminal-law concept of ‘scaremongering’ and the constitutional reasoning used to assess restrictions which are justified by reference to public peace, including the pandemic-related expansion of the offence of scaremongering, when special legal measures are in place. The French case study reconstructs the national approach associated with the Law ‘infox’ and the practice of Arcom, and it underscores both the limited number of judicial proceedings under the special election-related mechanism and the growing determining role of EU-level measures in shaping national options and enforcement patterns.

Mapping the landscape of fake news definitions

JÁNOS TAMÁS PAPP

1. Introduction

The concept of fake news has emerged as a significant and contentious issue in contemporary discourse, yet it eludes a concise and universally accepted definition. The complexity of defining fake news stems from its multidimensional nature, where it intersects with various fields such as media studies, political science, sociology and information technology. This paper seeks to unravel the intricacies involved in defining fake news, while analyzing the term from multiple perspectives and exploring the challenges that contribute to its elusive nature.

The concept of fake news can be ostensibly quite clear: it is factually false and can be considered news in terms of its form and appearance. However, this narrow perspective has expanded, encompassing a broad spectrum of misinformation and disinformation, ranging from satire and parody to propaganda and biased reporting. The digital era has further complicated the current affairs landscape, thanks to the rapid proliferation of online platforms which enable information to be disseminated swiftly and the boundaries between credible journalism and misleading content to become blurred. This convergence of factors creates a breeding ground for fake news, while making it increasingly complicated to identify and categorize such 'news'. Moreover, the subjective nature of truth and reality in news reporting adds a further layer of complexity. What one individual or group may perceive as biased or misleading may be interpreted as truthful and objective by another, making the task of establishing a clear-cut definition even more daunting. This subjectivity is often influenced by cultural, political, and social factors, highlighting the multifaceted nature of fake news and the diverse lenses through which it is viewed and understood.

This paper aims to dissect these complexities, exploring the various dimensions that contribute to the elusive nature of fake news. By examining the term through various disciplinary lenses, analyzing the impact of digital technology, and considering the subjective interpretation of information, this paper endeavors to provide a comprehensive understanding of why fake news remains widely discussed and deeply challenging to define conclusively.

2. The renaissance of fake news

In recent years, the world has become increasingly familiar with fake news, and digital platforms worldwide have not yet devoted sufficient resources to preventing its spread.¹

¹ GOSZTONYI (2023): 85.

This is problematic because fake news undermines public trust in institutions and in both online and conventional media and can undermine the foundations of democratic institutions by denying citizens access to credible information on which to base informed decisions. Moreover, misinformation is often used to promote radical and extremist ideas and activities and, by distorting public discourse, is also considered by many as seriously undermining freedom of expression itself. Mass online disinformation campaigns are often used by both domestic and foreign actors to create mistrust and social tension, or even to pose a threat to the integrity of democratic electoral processes. According to a 2018 survey, 83% of Europeans believe that fake news poses a problem for democracy in general.² According to the European Commission, disinformation undermines public confidence in the institutions and in online and conventional media, and it can undermine the foundations of democratic institutions by preventing citizens from accessing credible information to rely upon for making informed decisions.³ In addition, disinformation often serves radical and extremist ideas and activities, and in some cases even greatly affects the freedom of expression itself. Online disinformation campaigns aimed at the masses are widely used by both domestic and foreign actors to incite mistrust and social tensions, and, furthermore, such disinformation campaigns by foreign countries can jeopardize the integrity of democratic election processes⁴.

The absence of a clear and objective definition of fake news has been a consequential contributing factor to the challenges encountered in regulatory efforts aimed at addressing this issue. This lack of a precise definition creates ambiguity and opens the door to divergent interpretations, making it difficult for regulatory authorities to establish coherent and effective strategies for combating the dissemination of false or misleading information. Without a well-defined framework to delineate what constitutes fake news, there exists a risk of inadvertently stifling legitimate journalistic expression or impinging upon freedom of speech, further complicating regulatory endeavors. Consequently, the failure to establish an objective definition of fake news hampers the development of targeted and proportionate regulatory responses, which are essential in the context of an evolving media landscape characterized by information proliferation and digital platforms.⁵

The problem of fake news is not new, but it has become highly topical again in our times, as it plays a prominent role in public debates, in the press and in the decline or rise of socio-political thought and influence. Fake news is more of a threat today than in the past because its potentially harmful effects are amplified by modern technology. One of the potential benefits of the easy availability of information through the Internet and social networking sites is that it increases the possibility of participation in democratic processes and

² <https://ec.europa.eu/digital-single-market/en/news/final-results-eurobarometer-fake-news-and-online-disinformation>

³ European Commission: A European approach to tackling online misrepresentation. COM (2018) 236 final, Brussels of 26/04/2018. 1–2.

⁴ Ibid.

⁵ SUGOW (2019): 35.

makes them more inclusive. However, the downside of the ‘cheap talk’ enabled by the Internet is that the technology allows mass influence with fictitious information much more effectively than was previously the case.⁶

Most authors see two main reasons for the renaissance of fake news and it having greater prevalence than ever before. To begin with, there is the popularity and importance of traditional media and the opportunities offered by social networking sites.⁷ This is because by using their networking methods, social media are able to manipulate the content consumed by users in order to disseminate misleading or false information, either through simple sharing or promotion, or by influencing users.⁸ In many ways, social media sites provide a perfect environment for the spread of fake news. Firstly, they are easily accessible to anyone who wants to enter this market, as they can reach many people in return for a small financial investment – or even free of charge. Secondly, social media feeds actually favour fake news, as short, quick-to-read information in a couple of sentences or with a resonating title can help spread false content.⁹ All this is contributed to by the mixed nature of feeds, in which, after a while, the posts may converge in the user’s mind; as such, it is not so much the person (or the name of their page) or the sharer that matters, but the content, title or thumbnail. Thirdly, the filter bubble created by social networking sites, or the ‘daily self’, is also a breeding ground for spreading news that matches our ideology, even if it is not necessarily true.¹⁰ The novelty and sensational value of fake news (especially if astonishing or almost unbelievable) is much greater than that of simple (‘boring’) news, and the reaction it triggers is even more intense. The deep, emotional reactions evoked, which are usually some kind of disgust or outrage, make the user much more likely to pass the news on to others.

At the same time, the life cycle of fake news spreading on social media is quite short; such news quickly takes to the air and then quickly disappears (or is replaced by a newer, bigger sensation). However, the correction and clear refutations following the fake news are hardly shared; they reach very few people, so the distorting effect of the fake news can be highly significant.¹¹ Nabiha Syed names five factors that help spread fake news on social media: filtering, communities, amplification, speed, and profit-seeking. Filtering is necessary, because without it an unprocessable amount of information would flood the user. Hence, it is necessary to arrange the posts displayed in some form, and here the platform algorithm has a huge influence on what (pseudo) information a user gets. Social sites (as the name suggests) are built on communities and connect, in cyberspace, like-minded people who would most likely never have met in real life. In this way, the communities which form are communities where fake news corresponding to their common view can spread without criticism, and the resulting pseudo information is amplified by ever-spreading shares and potential distribution on other platforms. What is more, this pseudo information is able to proliferate at an incredible speed. Finally, according to Syed, the spread of fake news and

⁶ PAPP (2023): 144.

⁷ ALLCOTT–GENTZKOW (2017): 215.

⁸ European Commission: A European approach to tackling online misrepresentation. COM (2018) 236 final, Brussels of 26/04/2018. 5.

⁹ NICE (2019): 213.

¹⁰ DIFRANZO–GLORIA–GARCIA (2017): 32–35.

¹¹ BAPTISTA–GRADIM (2020): 6–8.

the traffic they generate results in large amounts of profit for social network operators, who therefore do not take effective action against it.¹²

3. The concept of the term fake news

While a colloquial understanding of fake news is pervasive, precisely delineating this term within the realm of factual reporting, deception, propaganda, opinion, satire, sarcasm, or purely entertaining content presents a formidable challenge. The academic discourse further underscores the intricate nature of this challenge, given that scholars often categorize false statements in divergent manners, thereby exemplifying the multifaceted and intricate nature of this issue. When considering the conceptual analysis of fake news, it should be noted that it is not merely a semantic, conceptual analysis. When we look at fake news, we are not talking about a detailed analysis of the meaning of the word, but about the social phenomenon of fake news as it exists today and is experienced by everyone.

There is a prevailing view which suggests that the term fake news has been subject to such a broad array of interpretations and applications by various researchers that it has undergone semantic dilution, leading to a loss of its original and precise signification. Consequently, proponents of this perspective contend that the term should no longer be used.¹³ Numerous observers concur that the designation fake news ought to be avoided due to its appropriation by certain politicians who employ it as a means to deflect criticism. However, Björnstjern Baade disagrees with this approach. He thinks the term is here to stay and will continue to be used by politicians and in everyday conversations.¹⁴ Ryan Walters argues that we should not discard the term fake news, as that would not make it disappear from public discourse. Instead, the term should be rehabilitated and given a scientifically comprehensible meaning.¹⁵

Fake news encompasses a spectrum of disinformation, misinformation, and malinformation, each with distinct nuances, making it elusive to delineate definitively. Moreover, the fluidity of information flows in the digital age, driven by rapidly evolving communication technologies and the democratization of content creation, complicates attempts to demarcate fake news from other forms of information disorder. Additionally, the subjectivity inherent in discerning the veracity of news reports, combined with the potential for biased interpretations and confirmation biases among audiences, further exacerbates the difficulty in constructing a universally applicable definition of fake news. This intricate interplay between technology, intent, and human cognition underlines the formidable task of devising a precise, universally acceptable definition of fake news within the realm of academic discourse.

In the following section, the concept of fake news will be broken down by considering what some important academics have to say about it. The goal is to distill the most significant ideas and categories to help us better grasp what fake news really means, and to syn-

¹² SYED (2017): 346–353.

¹³ VERSTRAETE–BAMBAUER–BAMBAUER (2017): 4.

¹⁴ BAADE (2020)

¹⁵ WALTERS (2018): 119.

thesize the conceptual foundations and salient categories that these academics have identified as crucial in the comprehension and delineation of fake news.

4. Divergent interpretations of fake news

In scholarly discourse concerning this subject matter, Alex Gelfert's definition is frequently cited by academics. According to Gelfert, the definition of fake news can be characterized as follows: "fake news is the deliberate presentation of (typically) false or misleading claims as news, where the claims are misleading by design".¹⁶ Somewhat similarly David Klein defines fake news simply as the online publication of intentionally or knowingly false statements of fact.¹⁷ However, detractors argue that these particular definitions are excessively reductionist in their conceptualization.¹⁸ According to Ryan Walter's proposal for a definition, we do not have to take into account the motive for publishing fake news; the term fake news has only three elements: (a) content presenting itself as a piece of news (b) that makes objectively false assertions that given events have occurred (c) in a materially false manner.¹⁹ According to Lili Levi, the category of fake news can mean click-hunting titles, politically motivated misinformation, satire and conspiracy theories all at the same time, but it can also mean press outlets that are disliked by the stigmatizer, or unintentional errors by the press. Hence, in her view, it would be more appropriate to use the term *fabricated false information*.²⁰

According to other authors, fake news can be defined as articles that exhibit both the semblance and substance of authentic news reports, deliberately incorporating at least one substantive factual assertion that can be empirically demonstrated as false, devoid of any protective justification under the auspices of fair information dissemination.²¹ Similarly, Neil Levy says fake news is the presentation of false claims that purport to be about the world in a format and with a content that resembles the format and content of legitimate media organizations.²² Regina Rini argues that a fake news story is one that purports to describe events in the real world, typically by mimicking the conventions of conventional media reporting, yet is known by its creators to be substantially false, and is transmitted with the two goals of being widely re-transmitted and of deceiving at least some of its audience.²³

Anderau delineates two principal criteria integral to the conceptualization of fake news: the *misleading condition* and the *news condition*, each comprising multiple nuanced sub-conditions. The discourse underscores that the deceptive intent inherent in fake news is paramount; this encompasses a spectrum of deceitful communicative acts, notably including lies. Anderau posits that the intentional fabric of deception is what demarcates fake news from mere inaccuracies or subpar journalism. He names four contentious dimensions

¹⁶ GELFERT (2018): 108.

¹⁷ KLEIN-WUELLER (2017): 6.

¹⁸ SMITH-PERRY (2021): 342.

¹⁹ WALTERS (2018): 120.

²⁰ LEVI (2017): 246–248.

²¹ CALVERT et al. (2018): 103–104.

²² LEVY (2017): 20.

²³ RINI (2017): 45.

within the fake news discourse: the relevance of truth value, the criticality of intentionality, the potential necessity for a minimum audience threshold, and the hypothesized dynamic nature of fake news.

Through illustrative exemplars, Anderau articulates a stance advocating that not all instances of misinformation should be automatically categorized under fake news and that factual content can morph into fake news contingent upon the presence of duplicitous intent. The author challenges prevailing notions stipulating that there be a minimum audience for the categorization of fake news or that it should be intrinsically dynamic in nature, instead advocating for a recognition that fake news is inherently grounded in the intent to deceive from the point of its inception, irrespective of the dynamics of any subsequent reception or interpretation.

The culmination of Anderau's scholarly endeavours is the proposition of a novel, nuanced definition of fake news: "fake news is misleading information that is intentionally published and presented as news, with the deliberate aim of deceiving its recipients regarding its authenticity as legitimate news". This definition accentuates the deliberate and intentional aspect of deception in fake news and underscores the critical importance of the creator's motive to misrepresent the authenticity of the content as genuine news.²⁴ Other authors also stress the importance of the fact that only the deliberate dissemination of false news constitutes fake news and any regulation should involve close attention to the protection of freedom of expression and the right to be wrong.²⁵

According to Ozbay and Alatas, even low-quality or incomplete news can be fake news.²⁶ Some also argue that it is not merely about what the author intended when defining fake news. They emphasize the importance of the audience's beliefs and how they perceive the news. Essentially, they say that whether something is considered fake news depends not only on the author's intent but also on what people think about it. In this view, if news is considered fake because people simply do not believe it, even though there is no deliberate deception involved, it is more like a work of fiction than malicious misinformation. However, dealing with these subjective and hidden factors, such as what the author meant and what the audience believes, is quite challenging.

Some propose using mechanism design to tackle this issue by addressing the information gap. The idea of "news you don't believe" is a murky concept which implies that if people disbelieve fake news, then it does not do much harm and may not need intervention. However, if disbelief is directed at genuine news, it presents a different problem: the news is still true but might cause harm, posing tricky challenges in terms of how we design and implement solutions.²⁷ Closely interlinked with this is the concept of the *post-truth era*, which, arguably, more aptly characterizes the contemporary phenomenon in which emotions and personal beliefs play a greater role in shaping public opinion than objective facts.²⁸

²⁴ ANDERAU (2021): 1.

²⁵ SUGOW (2019): 46.

²⁶ OZBAY-ALATAS (2020)

²⁷ VAN ALSTYNE (2023): 7.

²⁸ FARHOUDINIA-OZTURKAN-KASAP (2023): 7.

A study of fake news legislation in Asian countries, comparing legislation in a total of 10 countries,²⁹ identified common elements of legislation as (1) spreading false information likely to cause confusion, public harm, or panic, (2) publishing or giving false information causing concern about safety of a person or property, and (3) inciting to commit offenses or disturb social cohesion or damage to national security.³⁰ A similarly nuanced approach can be found in a study by Robert Smith and Mark Perry who identify fake news as “deliberate publication or distribution of material that contains disinformation or misinformation that was misleading by design and (1) the material is used to defame an individual; and/or (2) genuine sources are imitated; and/or (3) content is false and meant to deceive or harm; and/or (4) headlines, visuals, or captions do not support the content; and/or (5) genuine content is shared with false contextual information; and/or (6) genuine information or imagery is manipulated to deceive.”³¹

Another study encompassing an exhaustive analysis of over 60 distinct definitions of fake news within the academic corpus, identified three critical factors related to fake news: one, completely or partially false statements; two, intent to mislead; three, presented in news format³²). This lead to the formulation of the subsequent conceptual definition: “fake news is a type of online disinformation (1), with (2) misleading and/or false statements that may or may not be associated with real events, (3) intentionally created to mislead and/or manipulate a public (4) through the appearance of a news format with an opportunistic structure (title, image, content) to attract the reader’s attention, in order to obtain more clicks and shares and, therefore, greater advertising revenue and/or ideological gain”.³³

5. Summary of definitions and taxonomy of fake news

Considering the diverse definitions discussed earlier, it is clear that there are some key criteria that should be included in any definition of fake news. These common key elements are:

- Factual, objective falsity: fake news involves intentionally false information presented as fact.
- Intent to deceive: creators aim to deceive their audience deliberately. Some definitions include creator awareness that their content is false.
- Mimicry of authentic news: it often mimics the style and format of legitimate news, so adding a veneer of credibility.
- Diverse manifestations: fake news encompasses various forms, from clickbait to conspiracy theories.
- Transmission goals: fake news is designed to be widely shared and deceive some of its audience.

²⁹ Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam.

³⁰ SMITH–PERRY–SMITH (2021): 10.

³¹ SMITH–PERRY (2021): 346.

³² BAPTISTA–GRADIM (2022): 637.

³³ Ibid., 640.

Additionally, it is also very important to recognize and categorize different types or subcategories of fake news. Brittany Vojak classifies fake news into four categories. She distinguishes between intentional and accidentally disseminated fake news, and within each of these groups, she distinguishes between two more types. Within the intentional category: profit-oriented and influence-seeking fake news; and within the accidental category: viral news and news published due to negligent journalism.³⁴

Researchers at the University of Arizona have established several categories of fake news:³⁵ the first category includes satire, which is intentionally false, published for financial gain, and not intended to really mislead readers. In the second category are so-called hoaxes, which are also intentionally false and published for financial gain, but with the express purpose of genuinely misleading readers. The third category is propaganda, which is deliberately made up of false or misleading information, with the express purpose of genuinely misleading readers; however, the purpose of publication is not primarily financial gain but to gain political or similar influence.³⁶ Thus, according to the authors of the study, the two issues to be examined are whether the publisher actually wants to mislead readers and whether they do so for financial or other purposes.³⁷

Randall D. Nice also distinguishes between fake news in terms of motivation, arguing that fake news has four categories: ideologically motivated, financially motivated, state propaganda and satire.³⁸ Hunt Allcott and Matthew Gentzkow separate fake news – which is probably false and misleading news and newspaper articles – from journalistic mistakes, rumors, conspiracy theories, satire, and untrue statements by politicians.³⁹

Damian Tambini sets up six categories for what it is customarily called fake news: (1) intentional lies for political purposes, (2) intentional lies for economic purposes, (3) parody, (4) bad journalism (the dissemination of unconfirmed false information), (5) news that is ideologically opposed and (6) news questioning the traditional system of power.⁴⁰ Out of these, perhaps the latter deserves more explanation. Under this category, Tambini talks about when in any political community, there's a prevailing shared mindset or orthodoxy known as *zeitgeist*, which shapes how facts and events are perceived. Attempts to present information from a different perspective can be seen as fake, especially in areas such as business news, where groupthink can exaggerate market changes. It is crucial to differentiate between facts, their presentation (framing), and the editorial decisions behind what becomes newsworthy, as all these factors influence our understanding of the news.⁴¹

Claire Wardle also distinguishes between six types of fake news, based on the ways in which false information is disseminated: (1) use of an authentic source in an inappropriate context; (2) fake pages that mimic original news sites and exploit the credibility of their

³⁴ VOJAK (2017): 130.

³⁵ VERSTRAETE–BAMBAUER–BAMBAUER (2017): 5.

³⁶ LENDVAI (2023): 1236–1237.

³⁷ VERSTRAETE–BAMBAUER–BAMBAUER (2017): 7.

³⁸ NICE (2019): 211–212.

³⁹ ALLCOTT–GENTZKOW (2017): 214.

⁴⁰ TAMBINI (2017): 4–5.

⁴¹ Ibid., 4.

brand; (3) fake news sites; (4) visual elements (images, videos) and memes that spread false information; (5) manipulated content; (6) parody content (satire).⁴²

We should also mention, within the category of misleading information, the so-called hoax. A hoax makes real what seems impossible at first, making non-existent cases or conspiracies credible by using the appropriate frameworks. The term basically refers to chain letters or deceptive online content designed to generate more and more mail traffic by reaching as many users as possible, or to make fun of those who believe and share false information. It is important to note that the purpose of hoaxes is not to make money, so in such cases, one cannot really refer to them as fraud.⁴³ They can include urban legends, which have some basis in reality but are not verifiable and mostly fictional stories,⁴⁴ or propaganda that follows highly one-sided or single-channel communication, where one of its characteristic tactics is to trigger panic in order to direct its audience to where security is promised.⁴⁵ Clickbait is a form of fake news shared on social media. It is a sensational headline that encourages clicks to a separate webpage by providing some false information.⁴⁶

6. Misinformation and disinformation

By the onset of the 2020s, a notable shift had occurred within both academic discourse and legal literature regarding the terminology used to denote the deliberate dissemination of false or misleading information. This shift entailed a preference for the term ‘disinformation’ over its predecessor, fake news, underscoring the recognition of the need for a more precise and encompassing term to encapsulate the multifaceted nature of such information. The adoption of ‘disinformation’ within academic discourse reflected a desire for greater specificity, as well as a departure from the colloquial and somewhat subjective connotations associated with fake news.

Legal literature, in parallel, embraced the term as it resonated with the legal dimensions of misinformation campaigns, propaganda, and the intentional spread of falsehoods.⁴⁷ This terminological development reflected a growing consensus that the problem had extended beyond merely falsified news articles and was encompassing a broader spectrum of deceptive content disseminated across diverse media platforms. As a result, ‘disinformation’ emerged as a more apt and scientifically grounded term to address the intricate challenges posed by the manipulation of information in the evolving digital landscape.⁴⁸

This approach distinguishes between untrue information published accidentally and fake news published deliberately and in bad faith, calling the former misinformation and the latter disinformation. Misinformation refers to false or inaccurate information that is unintentionally or inadvertently spread, often without malicious intent. It can arise from genuine mistakes, mis-

⁴² https://www.cjr.org/tow_center/6_types_election_fake_news.php

⁴³ SZÜTS (2019): 428.

⁴⁴ VESZELSZKI (2017): 55.

⁴⁵ FEHÉR–KIRÁLY (2017): 40.

⁴⁶ CHUA–PAL–BANERJEE (2021): 291.

⁴⁷ KENYERES–SZÜTS (2024)

⁴⁸ PÉREZ–ESCOLAR–LILLEKER–TAPIA–FRADE (2023): 77–78.

understandings, misinterpretations, or errors in reporting. Misinformation can be disseminated through various channels, including social media, conventional media, word of mouth and on-line platforms.⁴⁹ Unlike disinformation, which involves the deliberate and intentional spread of false information, misinformation is typically the result of human error or misinformation campaigns that inadvertently convey inaccurate facts. Addressing misinformation often involves providing corrections, fact-checking, and promoting accurate information to mitigate its impact on public understanding and decision-making.⁵⁰

Within the field of European Union legislation and associated documentation, the terminology employed to address the issue of misleading or false information predominantly leans toward the usage of the term *disinformation*. However, it is pertinent to acknowledge that nuances exist within the definitions applied, even if these disparities are relatively minor. The High Level Expert Group on fake news and Online Disinformation set up by the European Union defines disinformation as “all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit”.⁵¹ However, online activities that are otherwise infringing (e.g. defamation, libel, hate speech) and currently regulated by Member States or the EU are not included in this definition. The purpose of disinformation is destructive; it raises doubts by distorting and falsifying facts, confusing people, and weakening their trust in institutions and accepted political processes.⁵² The report commissioned by the Committee on Civil Liberties, Justice and Home Affairs of the European Parliament states that disinformation is: information (i) designed to be false, manipulative or misleading, (ii) with the intention of generating insecurity, tearing cohesion or inciting hostility, or directly to disrupt democratic processes, (iii) it is a topic of public interest and (iv) often uses automated dissemination techniques to amplify the effect of the communication.⁵³ The Strengthened Code of Practice on Disinformation⁵⁴ refers to the European democracy action plan⁵⁵ for the definition of disinformation, which states that misinformation is “false or misleading content shared without harmful intent though the effects can still be harmful,” and “disinformation is false or misleading content that is spread with an intention to deceive or secure economic or political gain and which may cause public harm”.⁵⁶

⁴⁹ PERSILY–TUCKER (2020): 11.

⁵⁰ RENDA (2018)

⁵¹ Final report of the High Level Expert Group on Fake News and Online Disinformation. <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>

⁵² Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on the implementation of the Action Plan against Disinformation. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019JC0012&from=EN>

⁵³ Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States. Study Requested by the LIBE committee. [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU\(2019\)608864_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf)

⁵⁴ The Strengthened Code of Practice on Disinformation 2022. <https://ec.europa.eu/newsroom/dae/redirection/document/87585>

⁵⁵ 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. Brussels, 3.12.2020. COM(2020) 790.

⁵⁶ Ibid., Chapter 4. Countering Disinformation.

7. Conclusion

The concept of fake news, while widely recognized and discussed, remains an elusive and multifaceted issue in contemporary discourse. Its complexity arises from its intersection with various academic fields, including media studies, political science, sociology, and information technology. This paper has undertaken an exploration of the intricacies involved in defining and categorizing fake news, considering various perspectives and the divergent interpretations and categorizations within academic literature.

The fundamental characteristic of a fake news publication is in its lack of truthfulness. The primary factual assertions presented in fake news pieces are deliberately concocted and false. Moreover, fake news publications are deliberately or consciously deceptive. False news providers lack a reasonable belief in the veracity of the presented information. Most definitions of fake news fail to encompass negligent and reckless false publications of fact, which may be subject to legal action, including erroneous publishing by mainstream media sources.⁵⁷ Furthermore, the term fake news has transitioned to that of disinformation within academic and legal discourse, reflecting the recognition of a broader spectrum of deceptive content and manipulation of information in the digital landscape. The distinction between misinformation and disinformation highlights the importance of intent in defining the spread of false or misleading information, whether accidental or deliberate.

While the concept may seem straightforward at first glance, it encompasses a broad spectrum of misinformation and disinformation, from satire to propaganda, all masquerading as news. The digital era has exacerbated this complexity, as online platforms enable rapid information dissemination, blurring the lines between credible journalism and misleading content. Moreover, the subjectivity inherent in discerning the truth and reality in news reporting adds another layer of complexity. What one person may perceive as biased or misleading may be considered truthful by another, depending on the influence of cultural, political, and social factors. This subjectivity underscores the multifaceted nature of fake news and the diverse lenses through which it is viewed and understood.

Although disinformation and fake news may be disproved by tangible facts, a certain margin of bias might nevertheless remain.⁵⁸ This means that even when disinformation and fake news are debunked or proven false by specific and verifiable evidence, there may still be some degree of bias or subjective interpretation lingering in people's perceptions or beliefs. In other words, while the objective facts may refute the false information, individuals may continue to hold onto some degree of their pre-existing bias or perspectives, which can influence how they interpret or accept the information. This highlights the idea that fully eradicating the influence of disinformation and biases can be challenging, given that people's prior beliefs and inclinations can persist despite factual evidence to the contrary.

In conclusion, the study of fake news, disinformation and misinformation is essential in the contemporary information ecosystem. Developing precise definitions and categorizations is a complex but crucial endeavor, as it underpins regulatory efforts, safeguards freedom of expression, and informs strategies to combat the spread of false information. As technology and society continue to evolve, scholars, policymakers, and stakeholders must remain vigilant and adaptable in addressing the multifaceted challenges posed by fake news and misinformation in the digital age.

⁵⁷ KLEIN-WUELLER (2017): 6.

⁵⁸ QUINTEL-ULLRICH (2019): 14.

References

- ALLCOTT–GENTZKOW (2017) = Hunt ALLCOTT – Matthew GENTZKOW: Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, Vol. 31., No. 2. (2017) 211–236. <https://doi.org/10.1257/jep.31.2.211>
- ANDERAU (2021) = Glenn ANDERAU: Defining fake news. *KRITERION – Journal of Philosophy*, Vol. 35., No. 3. (2021) 197–215. <https://doi.org/10.1515/krt-2021-0019>
- BAADE (2020) = Björnstjern BAADE: Don't call a spade a shovel. *Verfassungsblog: On Matters Constitutional*, 14 April 2020. <https://doi.org/10.17176/20200415-032634-0>
- BAPTISTA–GRADIM (2020) = João Pedro BAPTISTA – Anabela GRADIM: Understanding fake news consumption: A review. *Social Sciences*, Vol. 9., No. 10. (2020), Art. no. 185. <https://doi.org/10.3390/socsci9100185>
- BAPTISTA–GRADIM (2022) = João Pedro BAPTISTA – Anabela GRADIM: A working definition of fake news. *Encyclopedia*, Vol. 2., No. 1. (2022) 632–645.
- CALVERT et al. (2018) = Clay CALVERT – Stephanie MCNEFF – Austin VINING – Sebastian ZARATE: Fake News and the First Amendment: Reconciling a Disconnect Between Theory and Doctrine. *University of Cincinnati Law Review*, Vol. 86., No. 1. (2018) 99–138.
- CHUA–PAL–BANERJEE (2021) = Alton YK CHUA – Anjan PAL – Snehasish BANERJEE: “This will blow your mind”. Examining the urge to click clickbaits. *Aslib Journal of Information Management*, Vol. 73., No. 2. (2021) 288–303. <https://doi.org/10.1108/AJIM-07-2020-0214>
- DIFRANZO–GLORIA-GARCIA (2017) = Dominic DIFRANZO – Kristine GLORIA-GARCIA: Filter bubbles and fake news. *XRDS: Crossroads, The ACM Magazine for Students*, Vol. 23., No. 3. (2017) 32–35.
- FARHOUDINIA–OZTURKCAN–KASAP (2023) = Bahareh FARHOUDINIA – Selcen OZTURKCAN – Nihat KASAP: Fake news in business and management literature: a systematic review of definitions, theories, methods and implications. *Aslib Journal of Information Management*, Vol. 77., No. 2. (2025) 306–329. <https://doi.org/10.1108/AJIM-09-2022-0418>
- FEHÉR–KIRÁLY (2017) = FEHÉR, Katalin – KIRÁLY, Olivia: Álhíresülés. A hamis hírek dinamikája a médiában (Fake news – The dynamics of fake news in the media). *Századvég*, 2017/2. 39–48.
- GELFERT (2018) = Axel GELFERT: Fake news. A definition. *Informal Logic*, Vol. 38., No. 1. (2018) 84–117. <https://doi.org/10.22329/il.v38i1.5068>
- GOSZTONYI (2023) = Gergely GOSZTONYI: *Censorship from Plato to social media. The complexity of social media's content regulation and moderation practices*. Cham, Springer, 2023. <https://doi.org/10.1007/978-3-031-46529-1>

- KENYERES–SZÜTS (2024) = KENYERES Attila Zoltán – SZÜTS Zoltán: Az álhírek (fake news) definíciós kísérletei és nyelvi jelei (Attempted definitions and linguistic signs of fake news). *Magyar Nyelvőr*, 2024/2. 203–235. <https://doi.org/10.38143/nyr.2024.2.203>
- KLEIN–WUELLER (2017) = David KLEIN – Joshua WUELLER: Fake news. A legal perspective. *Journal of Internet Law*, Vol. 20., No. 10. (2017) 5–13. <https://ssrn.com/abstract=2958790>
- LENDVAI (2023) = Gergely Ferenc LENDVAI: Media in war. An overview of the European restrictions on Russian media. *European Papers*, Vol. 8., No. 3. (2023) 1235–1245. <https://doi.org/10.15166/2499-8249/715>
- LEVI (2017) = Lili LEVI: Real “Fake News” and fake “Fake News”. *First Amendment Law Review*, Vol. 16. (Symposium Issue) (2017) 232–327. <https://journals.law.unc.edu/firstamendmentlawreview/wp-content/uploads/sites/8/2018/03/falr-volume-16-symposium-issue3.pdf>
- LEVY (2017) = Neil LEVY: The bad news about fake news. *Social Epistemology Review and Reply Collective*, Vol. 6., No. 8. (2017) 20–36. <http://wp.me/p1Bfg0-3GV>
- NICE (2019) = Randall D. NICE: Reviving the lost tort of defamation. A proposal to stem the flow of fake news. *Arizona Law Review*, Vol. 61., No. 1. (2019) 205–230.
- OZBAY–ALATAS (2020) = Feyza Altunbey OZBAY – Bilal ALATAS: Fake news detection within online social media using supervised artificial intelligence algorithms. (February 15, 2020) *Physica A: Statistical Mechanics and its Applications*, Vol. 540., Art. no. 123174. (2020). <https://doi.org/10.1016/j.physa.2019.123174>
- PAPP (2023) = János Tamás PAPP: Recontextualizing the role of social media in the formation of filter bubbles. *Hungarian Yearbook of International and European Law*, Vol. 11., No. 1. (2023) 136–150. <https://doi.org/10.5553/hyiel/266627012023011001012>
- PÉREZ-ESCOLAR–LILLEKER–TAPIA-FRADE (2023) = Marta PÉREZ-ESCOLAR – Darren LILLEKER – Alejandro TAPIA-FRADE: A systematic literature review of the phenomenon of disinformation and misinformation. *Media and Communication*, Vol. 11., No. 2. (2023) 76–87. <https://doi.org/10.17645/mac.v11i2.6453>
- PERSILY–TUCKER (2020) = Nathaniel PERSILY – Joshua A. TUCKER: *Social media and democracy. The state of the field, prospects for reform*. Cambridge, Cambridge University Press, 2020. <https://doi.org/10.1017/9781108890960>
- QUINTEL–ULLRICH (2019) = Teresa QUINTEL – Carsten ULLRICH: Self-regulation of fundamental rights? The EU code of conduct on hate speech, related initiatives and beyond. In: Bilyana PETKOVA – Tuomas OJANEN (Eds.): *Fundamental rights protection online. The future regulation of intermediaries*. Cheltenham, Edward Elgar, 2020.
- REDA (2018) = Andrea REDA: *The legal framework to address fake news. Possible policy actions at the EU level*. Brussels, Policy Department for Economic, Scientific and Quality of Life Policies, 2018. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/619013/IPOL_IDA\(2018\)619013_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/619013/IPOL_IDA(2018)619013_EN.pdf)

- RINI (2017) = Regina RINI: Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal*, Vol. 27., No. 2. Supplement (2017) 43–64. <https://doi.org/10.1353/ken.2017.0025>
- SMITH–PERRY–SMITH (2021) = Robert Brian SMITH – Mark PERRY – Nucharee Nuchkoom SMITH: ‘Fake News’ in Asean. Legislative responses. *Journal of ASEAN Studies*, Vol. 9., No. 2. (2021) 117–137. <https://doi.org/10.21512/jas.v9i2.7506>
- SMITH–PERRY (2021) = Robert SMITH – Mark PERRY: Fake news and the convention on cybercrime. *Athens Journal of Law*, Vol. 7., No. 3. (2021) 335–358. <https://doi.org/10.30958/ajl.7-3-4>
- SUGOW (2019) = Abdulmalik SUGOW: The right to be wrong. Examining the (im)possibilities of regulating fake news while preserving the freedom of expression in Kenya. *Strathmore Law Review*, Vol. 4., No. 1. (2019) 19–46. <https://doi.org/10.52907/slr.v4i1.108>
- SYED (2017) = Nabihah SYED: Real talk about fake news. Towards a better theory for platform governance. *Yale Law Journal Forum*, Vol. 127. (2017) 337–357. <http://www.yalelawjournal.org/forum/real-talk-about-fake-news>
- SZÜTS (2019) = SZÜTS, Zoltán: Online. *Az internetes kommunikáció és média története, elmélete és jelenségei* (Online. History, theory and phenomena of Internet communication and media). Budapest, Wolters Kluwer, 2019.
- TAMBINI (2017) = Damian TAMBINI: *Fake news. Public policy responses*. (Media Policy Brief, 20.) London, Media Policy Project, London School of Economics and Political Science, 2017. <https://researchonline.lse.ac.uk/id/eprint/73015/>
- VAN ALSTYNE (2023) = Marshall W. VAN ALSTYNE: *Free speech and the fake news problem*. (April 10, 2023). <http://dx.doi.org/10.2139/ssrn.4414261>
- VERSTRAETE–BAMBAUER–BAMBAUER (2017) = Mark VERSTRAETE – Derek E. BAMBAUER – Jane R. BAMBAUER: Identifying and countering fake news. *Arizona Legal Studies Discussion Paper*, No. 17-15., 2017.
- VESZELSZKI (2017) = VESZELSZKI, Ágnes: A hamis hírek extra- és intralingvális jellemzői (Extra- and intralingual features of fake news). *Századvég*, 2017/2. 51–82.
- VOJAK (2017) = Brittany VOJAK: Fake news. The commoditization of internet speech. *California Western International Law Journal*, Vol. 48., No. 1., Art. no. 5. (2017) 123–158. <https://scholarlycommons.law.cwsl.edu/cwilj/vol48/iss1/5>
- WALTERS (2018) = Ryan M. WALTERS: How to tell a fake. Fighting back against fake news on the front lines of social media. *Texas Review of Law & Politics*, Vol. 23., No. 1. (2018) 111–179.

Disinformation from a communication and media studies perspective

MÓNIKA ANDOK

1. Introduction

Over the last ten years, both the media and public organizations – for example, the European Council or the World Health Organization during the coronavirus epidemic – have paid increasing attention to the phenomenon known in the press as the spread of fake news; in fact in 2017, fake news was chosen as the Collins English Dictionary word of the year. Three prominent political events – the election of Donald Trump as president in 2016, the result of the Brexit vote and the French presidential election campaign – placed fake news under intense academic focus.¹ Several media research journals published special issues and the social media platform Facebook, which was held responsible for the spread of fake news, also reacted to what happened.² Therefore, social media has become nowadays a powerful source for fake news dissemination.³

Many problems have become interwoven with the dysfunctional operation of the news media, which is becoming increasingly apparent. The impact of fake news has reached various areas and disciplines beyond online social networks and society, such as economics, psychology, political science, health science and environmental science (e.g., climate change). To explore this impact, the technological elements of the changing media ecosystem, and the radically changed cultural practices of content production and reception have to be taken into account.

The definition of fake news according to the Oxford English Dictionary is as follows: “false reports of events, written and read on websites”.⁴ Furthermore, disinformation is construed as: “false information that is given deliberately”.⁵ Meanwhile, the European Commission defines the disinformation as follows: “false, inaccurate, or misleading information

¹ WARDE–DERAKHSHAN (2017)

² WEEDON–NULAND–STAMOS (2017); NEWMAN et al. (2017).

³ AÏMEUR–AMRI–BRASSARD (2023): 2.; SHARMA et al. (2019); SHU et al. (2017).

⁴ ‘Fake news’ [dictionary entry]. *Oxford Learner’s Dictionaries*, <https://www.oxfordlearnersdictionaries.com/definition/english/fake-news>.

⁵ ‘Disinformation’ [dictionary entry]. *Oxford Learner’s Dictionaries*, <https://www.oxfordlearnersdictionaries.com/definition/english/disinformation>.

designed, presented and promoted to intentionally cause public harm or for profit”⁶ For a comparison of the various definitions, see the paper by Akhtar et al. in detail.⁷

The spread of fake news is often classified as one of the defining phenomena of a broader category: the post-truth era. This article does not seek an answer to the question of whether we really live in a post-truth era, as this question is not related to media theory, given that truth is a concept that is mostly dealt with in philosophy or theology. Media studies can adequately deal with the issue of real, realistic and fictional display, or representation as it is called. Representation includes both the process of media presentation and its result. It was already well known with regard to conventional mass communication that classical metaphors of representation based on transparency, such as ‘window to the world’ and ‘mirror to reality’ cannot be applied, as they do not merely provide an impression or a reflection. Since each medium has its own unique agency that will shape the representation.⁸ At the same time, in the case of news content appearing in conventional mass media, (media) representation presupposes the existence of reality.⁹

The issue of network communication is also a thorny one because in addition to physical reality, we also have to deal with forms of representation such as virtual worlds or augmented reality. Therefore, it is expedient to discuss the phenomenon not from the point of view of truth or reality, but of authenticity, to which Gelencsér and Szűts refer.¹⁰ The creation and construction of credibility differs according to medium and historical-cultural traditions, although the effects of globalization brought about significant standardization in the case of news media. In this regard, we can identify an institutional strategy that can be linked to news organizations – news agencies, editorial offices, publishers – as it is in their interest to preserve their credibility and reliability. The media theory literature refers to this phenomenon by calling these signals credibility indicators.¹¹

I would like to cite two examples from Hungarian press history, one related to Lajos Kossuth, the other to the opening of the Hungarian Telegraph Office. In 1841 Jenő Landerer signed a contract with Lajos Kossuth, making Kossuth editor of the *Pesti Hírlap* newspaper. In the contract “the only thing the publisher stipulated was that the editor should accept articles that are worthy of publication due to their intrinsic merits and written in a decent tone, even if they might be in conflict with the editor’s opinion, so that accusation of one-sidedness that excludes other opinions should not serve as an obstacle to the life of the newspaper,” says Ferenczy.¹² Additionally, when Géza Egyesy and Hugó Maszák endeavored to start the Hungarian Telegraph Office in 1880, to justify receiving support, they wrote in their application: “they propose the editing and distribution of objective telegrams corre-

⁶ A multi-dimensional approach to disinformation. Report of the independent High Level Expert Group on Fake News and Online Disinformation. European Commission, 2018. <https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation>.

⁷ AKHTAR et al. (2023)

⁸ ALLAN (2010); FISKE (2011).

⁹ BELTING (2009): 21.

¹⁰ GELENCSÉR–SZÜTS (2020): 44.

¹¹ NYGREN–BROUNÉUS–SVENSSON (2019); LANIUS (2021).

¹² FERENCZY (1887): 78.

sponding to reality".¹³ In other words, news institutions as far back as the 19th century took care to ensure that their credibility was not open to question.

In the construction of credibility, we can identify not only elements linked to institutions, but also textual strategies, i.e. the way in which the text refers to itself as an objective, credible, factual account, and was written according to the textual characteristics typical of such accounts – for example, news is written in third person singular, according to the rules of 5W+H¹⁴ However, while the practices and forms of these credibility constructions have stabilized during the past century on conventional mass media platforms, a new strategy adapted to the (technological and content) characteristics of the medium of the Internet has not yet crystallized in network communication. At the same time, the audience still decides on the reliability of the content based on the old authenticity signals and credibility indicators, but these signals are already included in the content on the social network sites in many cases, with intent to mislead.

Not only is the term *post-truth era* inadequate in a media theoretical framework, but the name fake news is also misleading. This is why many researchers recommend that instead of objectivity, the issue of credibility should be the focus when describing the misleading operation of news. From this, it also follows for media research that it is advisable to choose the ritual theory of communication and not the transmission theory as the exploratory analysis framework of the fake news phenomenon.¹⁵ This thesis analyses the topic within this theoretical framework, with a focus on social media.

2. Fake news and the ritual model

Almost without exception, all analyses in the literature begin with the statement that the term fake news is inappropriate and even misleading in describing the phenomenon, as it suggests that fake news in the media consists of texts published in the form of news genre, but with false content. The term *fake news* is defined in the Collins English Dictionary as false and often sensational information disseminated under the guise of news reporting. Over time, this term has evolved and become synonymous with the spread of false information. Allcott and Gentzkow provided the first formal definition of *fake news*, characterizing it as news articles that are intentionally and verifiably false and designed to mislead readers. Subsequent definitions in the literature concur that the authenticity of fake news is fundamentally false, meaning it is non-factual.¹⁶ However, other types of media content are also classified here.¹⁷ What kind of contents are these and how can they be typologized? The political sphere tries to appropriate the term such that it uses the term for media and media content that do not support their beliefs and political orientation.

In this study, I will present two typologies, both of which are among the most frequently cited in the media theory literature. One is the division of Warde et al., developed

¹³ PIRITYI (1996): 34.

¹⁴ ANDOK (2013): 181–194.

¹⁵ ACZÉL (2017); WARDE–DERAKHSHAN (2017).

¹⁶ ALLCOTT–GENTZOW (2017)

¹⁷ WEEDON et al. (2017)

in 2017, at an early stage of the emergence of fake news.¹⁸ The other is the division proposed by Aïmeur et al., which was published in 2023 and includes the results of the previous studies.¹⁹ First, I present the previous model, which is that of Warde et al. Warde and their fellow researchers introduce a triple terminology instead of the concept of fake news: they propose the terms misinformation, disinformation and malinformation, precisely defining what they mean by one or the other. Two sets are projected onto each other, one representing false information (misinformation) and the other harmful information (malinformation), and the intersection of the two will contain both false and harmful (disinformation) information.²⁰ Those in the false set are called misinformation, those in the harmful set are called malinformation, and those in the intersection of the two sets are called disinformation. In the case of false information, the content is misleading or the text connects information elements in a misleading way; the bad and harmful information can be identified as leaks with harmful intent, confusion or hate speech.²¹

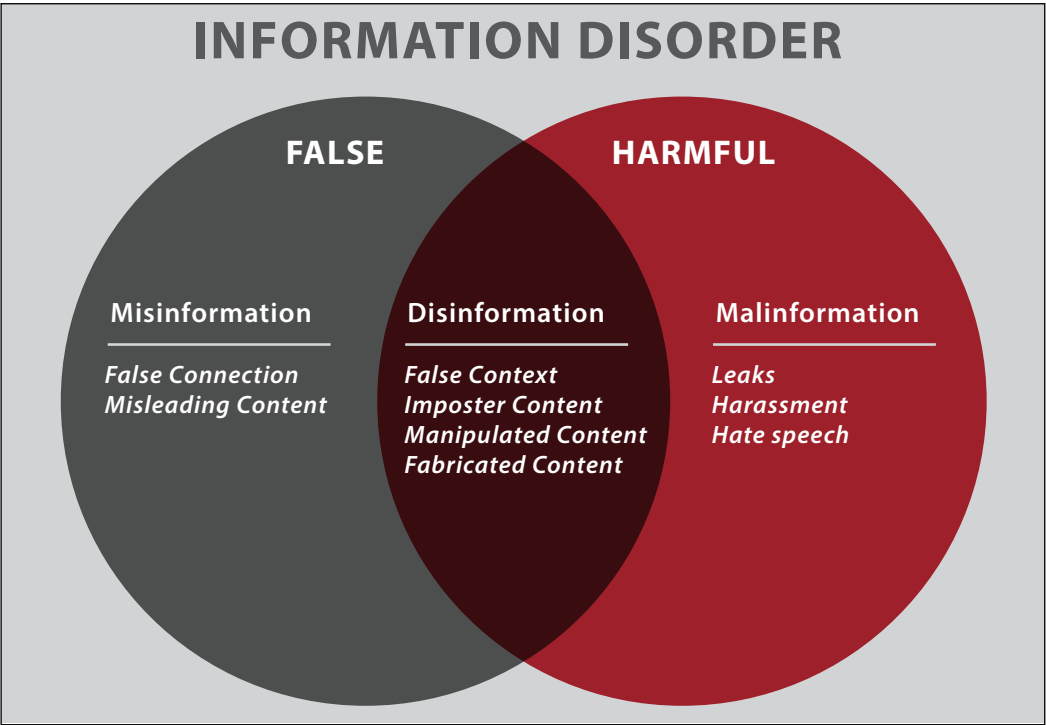


Figure 1. Typology of fake news based on WARDE–DERAKHSHAN (2017)

The above typology is analyzed by Warde et al. both in the system of media communication and in its dynamics. Among the elements of communication, the agent, the message and the interpreter are examined separately. While in terms of process, the mechanisms of crea-

¹⁸ WARDE–DERAKHSHAN (2017)
¹⁹ AÏMEUR–AMRI–BRASSARD (2023)
²⁰ WARDE–DERAKHSHAN (2017)
²¹ WARDE–DERAKHSHAN (2017): 5.

tion, production and distribution are explored for the three different types. They do this in such a way that their entire analysis is carried out within the ritual model of communication, while consciously avoiding the transmission approach. The analytical framework of the ritual theory is highlighted and this is considered more appropriate, because the ritual model is able to explain precisely how we create, maintain and shape our social reality through communication.²² According to the cultural or ritual view of communication, it is not only the spatial extension of messages that is important for a society or a culture: the maintenance of the community and the culture over time has even greater significance. This maintenance is achieved not only through the transfer of information, but also by sustaining beliefs and convictions related to convictions.

The ritual model of communication was developed by American researcher James Carey. "The driving force behind the ritual approach to communication is not the transmission of messages in space, but the cohesion of society in time."²³ When Carey differentiates between the transmission and the ritual models, he does not draw the line between good and bad, correct and incorrect, scientific and non-scientific, because both models give a valid description of communication. The transmission model focuses on the flow of information, while the ritual perspective sheds light upon the aspects of culturally sustaining communication, and the creation and maintenance of a more meaningful cultural world.

The study of communication – based on Carey's theory – means the examination of the real social processes that create the symbolic cultural forms important to a community, fill them with meaning and put them to social use.²⁴ In other words, in this approach, the main problem is not the one-off appearance of fake news, but the fact that the dominance of fake news can undermine the centuries-old practices of authentic news communication and those forms of social cooperation based on it. James Carey's work also influenced the practice of journalism and the public journalism movement that developed in the mid-1990s. He was not primarily interested in the internal, narrative systems and genre issues of journalism, but rather in the close relationship between democracy and journalism. In other words, journalism, including the culture of news communication, is closely related to the quality of democracy both at the global and national level, which is why the spread of fake news poses a serious threat.

Carey's life's work clearly reflects his deep commitment to democracy: in his thinking on the ritual model, he openly rejects the instrumentalist understanding of communication and media. The media of communication are not tools – in terms of will or goals – but social forms belonging to human life that reproduce our thoughts, actions and social relationships. This is precisely why it becomes a relevant theoretical framework for the study of fake news. It is particularly important to mention that aspect of the ritual view according to which the rite (or media communication perceived as a rite) is particularly suitable for framing attention and thereby delineating boundaries and values in the life of the community.²⁵ The ability to direct and frame attention is extremely important due to the dominant spread

²² CAREY (2009); ANDOK (2013); ANDOK (2017).

²³ CAREY (2009): 15.

²⁴ CAREY (2009): 24.

²⁵ ANDOK (2017): 103–106.

of fake news on social media. After all, one of the effects/goals of social media is the control of attention.²⁶

This mechanism can be considered in detail, in keeping with the analysis of English media researcher Nick Couldry. In the case of media content which can be interpreted as ritual communication, the broader framework – the framework of understanding and interpretation – is particularly important. Couldry connects the issue of ritual and power at this point by focusing on legitimation, given that ritual action is highly suitable for strengthening and communicating widely shared values, and thereby creating their legitimacy. In the words of Mary Douglas: the rite focuses attention through framing; it refreshes the memory, connects the present with the relevant past, and thereby also supports perception.²⁷ The rite does not involve random associations, but rather focuses attention on fundamental values, categories and hierarchies that organize our world, and that is why it is essential that news should not be fake, but real. At this point, Couldry also describes his definition of media rites: media rites are formal activities organized around basic media-related categories and boundaries, the performance of which frames or suggests broader media-related values.²⁸ According to Couldry, the most important thing from the point of view of media rites is not what is explained in the media text, but the formal relationship between various types of actions. According to Bell, the most important feature of ritualization is how it organizes our movements in space, by helping us to experience our constructed environment as real. Ritualization thereby reproduces the symbolic power inherent in categorization, which is drawn by the rite.²⁹

Media rites also function in the same way as Bell described regarding religious rites, but in this case the space is the ritual media space described by Couldry. In this ritual media space, we experience the constructed elements as real. In this way, media provide new forms of ritual experience, not only in the case of real news but also in that of fake news. The misleading and deceptive mechanisms of fake news attract and frame attention in a false way and offer misguided experiences and patterns of incorrect boundary setting based on disinformation, in terms of both the internal and external boundaries of the social environment.

Another typology for fake news was published by Aïmeur et al. in 2023.³⁰ The literature published until 2022 was processed and the terms related to fake news were listed and grouped. “We present the general context from which the fake news problem emerged (i.e. online deception). We review existing definitions of fake news, identify the terms and features most commonly used to define fake news, and categorize related works accordingly. We propose a fake news typology”.³¹ According to their research, the existing terms can be separated into two groups. The first group represents the general terms, which are information disorder, false information and fake news, each of which includes a subset of terms from the second group. The second group represents the elementary terms, which are misinformation, disinformation and malinformation. The literature agrees on the definitions for

²⁶ SZÜTS (2018): 428.

²⁷ DOUGLAS (1984): 64., cited by COULDRY (2003): 25.

²⁸ COULDRY (2003): 29.

²⁹ BELL (2009)

³⁰ AÏMEUR–AMRI–BRASSARD (2023)

³¹ AÏMEUR–AMRI–BRASSARD (2023): 4.

the latter group, but there is still no agreed-upon definition for the first group. *Figure 2* presents a model of the relationship between the most used terms in the literature.³²

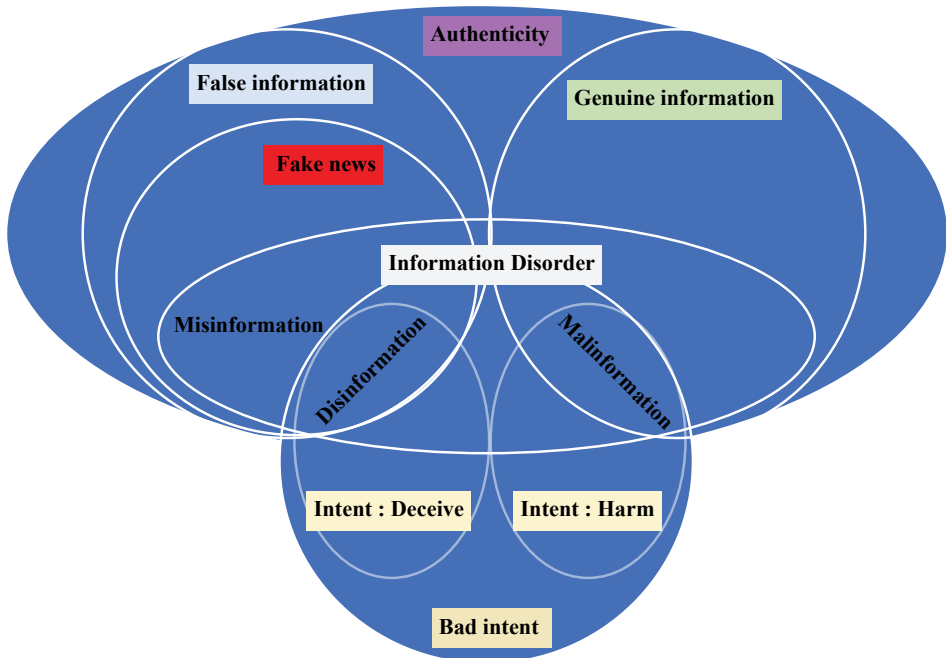


Figure 2. Modeling of the relationship between terms related to fake news
(source: AÏMEUR-AMRI-BRASSARD [2023]: 7.)

The two major approaches in previous typologies are taken in order: one is typology according to intention, while the other is according to content.

“Most researchers classify fake news based on the intent (...). However, other researchers (...) focus on the content to categorize types of fake news through distinguishing the different formats and content types of data in the news (e.g., text and/or multimedia). Recently, another classification was proposed by Zhang and Ghorbani (2020). It is based on the combination of content and intent to categorize fake news. They distinguish physical news content and non-physical news content from fake news. Physical content consists of the carriers and format of the news, and non-physical content consists of the opinions, emotions, attitudes and sentiments that the news creators want to express.”³³

³² AÏMEUR-AMRI-BRASSARD (2023): 7.

³³ AÏMEUR-AMRI-BRASSARD (2023): 8. For the intended typology, it is mentioned this sentence fragment is unclear: Collins et al. 2020; Bondielli and Marcelloni 2019; Zannettou et al. 2019; Kumar et al. 2016; Wardle 2017; Shu et al. 2017; Kumar and Shah 2018. Mentioned for content-based typologies: Parikh and Atrey 2018; Fraga-Lamas and Fernández-Caramés 2020; Hasan and Salah 2019; Masciari et al. 2020; Bakdash et al. 2018; Elhadad et al. 2019; Yang et al. 2019b. See *ibid*.

Zhang and Ghorbani’s next step was to build their own typology based upon this, so as to also take into account the peculiarities of multimedia interfaces, as shown in *Figure 3*.

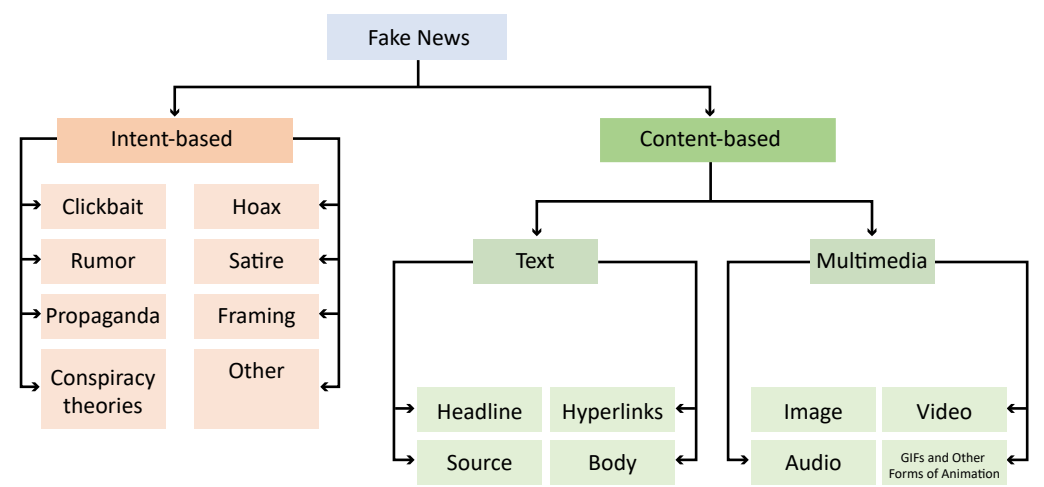


Figure 3. Fake news typology according to Aïmeur et al.
(source: AÏMEUR–AMRI–BRASSARD [2023]: 8).

3. A new media ecosystem

The increase in the proportion of *fake news* within media content was facilitated by the transformation of the mass communication system, the emergence of network communication and social media. Moreover, it is no longer just about appearance, but about the fact that online media have become the primary source of news consumption.³⁴ New media technologies, particularly social media, have democratized content creation and this content can be misleading or simply false. At the same time, these new media have also democratized the ability to manipulate and made it possible to spread that manipulation widely through the network, while bypassing professional journalistic control in its role as gatekeeper.

That said, misleading communication itself is not a new phenomenon: in fact it is perhaps one of the oldest forms of face-to-face communication, if we think of deception, lies, rumors, slander and urban legends. It is not new in terms of media communication either, as propaganda and hoaxes have been under discussion for a long time. However, with the use of social media technology – as part of a radically changed media ecosystem – misleading communication has entered a new era. What are the hallmarks of this new era of deceptive communication?

With new technology (smartphones and social media) it is extremely easy to create and share content.

³⁴ Digital News Report 2017: 10.

Information consumption has turned from a private activity into a publicly observable activity (which can be archived and analyzed). User attention pattern data is accessible and can be analyzed.

The spread of information in network communication takes place in real time; not necessarily with high speed, but certainly simultaneously with the event itself.

The way content is shared through social media is predominantly characterized by the fact that the focus is not on the source and format of the information (text, image, video), but almost exclusively on its content.

The use of network communication – at the search engines level – reinforces the process.³⁵

The same phenomenon is described by Weedon and his fellow researchers – who are development engineers at Facebook – as civil initiatives are increasingly present in the constantly changing information ecosystem, and the growth of individual access and operation has become important in political dialogue. The speed and scale of information consumption, the diversity and impact of influence mean both new opportunities and new challenges.³⁶ In addition to all these, it is important to highlight the echo chamber and filter bubble effect in social media.³⁷ Social media is not inherently a news source, its primary function is not the transfer of information. Within the communities created with the filtering settings, their own views are echoed by the users' acquaintances, and people feel comfortable in these echo chambers, which in turn reinforce their attitudes.

El-Bermawy, an Egyptian media scholar believes that the metaphor of the global village is misleading in relation to social media, because rather digital islands of isolation have been created on this interface. The more the network culture becomes personalized, the more the users become isolated from each other. This phenomenon cannot be described and cannot be fully grasped with the transmission theory of communication, but the ritual model can succeed in doing so. After all, it is not the truth or falsity of a single piece of news or information that is important, but the bias of the system itself, and this can primarily be examined on a macro level. Moreover, in some cases, with certain types of news referred to by Warde as malinformation, the basic information itself is true, and it is actually the framing which makes it ambiguous and culturally or morally distorted.

Starting in 2017, Lucy Küng made ongoing analysis of the economic impact of the new media ecosystem on the news media. Her research shows that twenty-five years after the rise of the Internet, a new media ecosystem was created. In terms of methodology, Küng conducted sixty in-depth interviews with decision-makers related to the business part of the news media. The general trend that can be revealed from these is that the conventional strategic sovereignty of the media industry is decreasing, content production is increasingly vulnerable both to technological change and, according to the 2020 data of the Digital News Report, to the news consumers' expectation.³⁸ The dominance of Google and Facebook is particularly strong. What can a media company do if it is financed by venture capital and loses the possibility of 'saying no' in terms of content production?

³⁵ WARDE–DERAKHSHAN (2017)

³⁶ WEEDON et al. (2017)

³⁷ SUNSTEIN (2013); PARRISER (2011).

³⁸ Digital News Report 2020: 41.

Küng revealed other three significant changes. The first is the strengthening of dialogue between news editors and news consumers; the second is that this dialogue is increasingly a dialogue between equal parties; and the third is that the editors want to shape the news in such a way that encourages the reader to engage with it: not merely to consume it but also to form an opinion about it, report it and share it on their social media profile. In the emerging new business models, the dual structure of the media market continues to expand, and it is possible to sell products in three types of markets. In addition to the previous two markets – the sale of content to readers and reader availability sold to advertisers – data analysis companies are now entering as a third market, whose market profile involves the analysis and further sale of download data. Since the majority of media market revenues already come from the digital market, the importance of online distribution becomes central to creating a marketing strategy or plan. It is possible to witness the development of new forms of online digital storytelling and the further expansion of graphic data visualization.³⁹

4. Process analyses

As briefly mentioned above, Warde et al. have analyzed the three types of fake news they identified (misinformation, malinformation and disinformation) within both structural and dynamic media systems. Among the structural elements of communication, the agent, the message and the interpreter are examined separately.

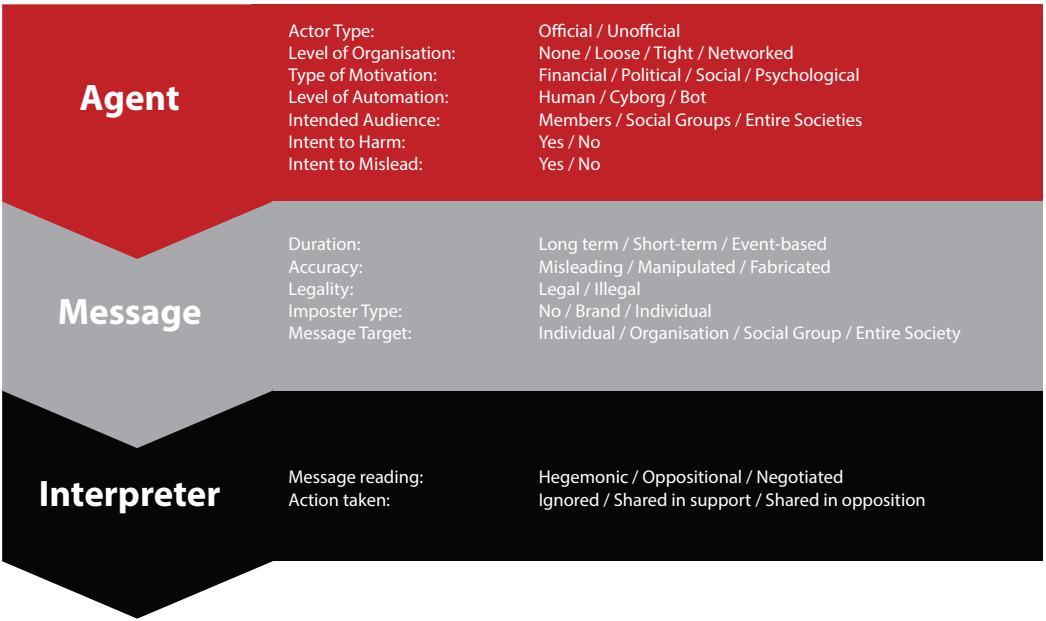


Figure 4. Process analysis of the operation of fake news based on WARDE–DERAKHSHAN (2017)

³⁹ KÜNG (2017): 27–28.

Firstly, it is necessary to consider the structural elements, the question of the agent, the message and the interpreter-receiver. In the case of the agent, Warde et al. investigated who created the message (official or unofficial source, automated technology), who disseminated it (what kind of community it is intended to reach) and what the presumed motivation behind it is (economic, political, social, and psychological). Is the message likely to have a short or long-term effect? How accurate and authentic is the branded content from an official source? From empirical analyses, it was determined that the following characteristics are the most important for a given message:

1. Is it expected to cause an emotional reaction, and if so, how much?
2. Does it contain a strong visual element?
3. How strong is the narrative it uses?
4. Is it a repeated, circulated message?

In the case of structural components, they mention the receiver, quoting Hall's three types of coding (hegemonic, negotiated and oppositional), but emphasize the primacy of emotional processing when using social media. The primacy of emotions and sentiments on digital surfaces is not new. As early as the mid-90s, it was noticed how many negative emotions and frustration surround the use of the Internet.⁴⁰ Since the emergence of social media in 2000, the range of investigations has expanded into two major areas: one related to information processing which has been influenced by emotions, and the other related to the issue of sentimental public sphere. The reception of information in social media basically works with emotional framing.⁴¹ This is also true on a macro level, since social media is an emotion-based public sphere.⁴² This is maintained by the technological design itself and the data analysis (datafication and sentiment analysis) based upon it.⁴³ In 2017, Wagner and Boczkowski also conducted inclusive research in which they interviewed 71 subjects and revealed why and to what extent the level of trust in the news media in the United States is low.⁴⁴

In the second part of their analysis, Warde et al. have explored the mechanisms of creation, production and distribution in the case of the three different types of false information.

⁴⁰ WALLACE (2016)

⁴¹ NABI (2003); COMAN-COMAN (2017); DAFONTE-GÓMEZ (2018).

⁴² ANDOK (2020); GELENCSEŔ-SZŰTS (2020): 45–49.; KISS (2013); PAPACHARISSI (2014).

⁴³ VAN DIJCK (2014): 198.; SZŰTS (2018): 462.

⁴⁴ WAGNER-BOCZKOWSKI (2018): 874.

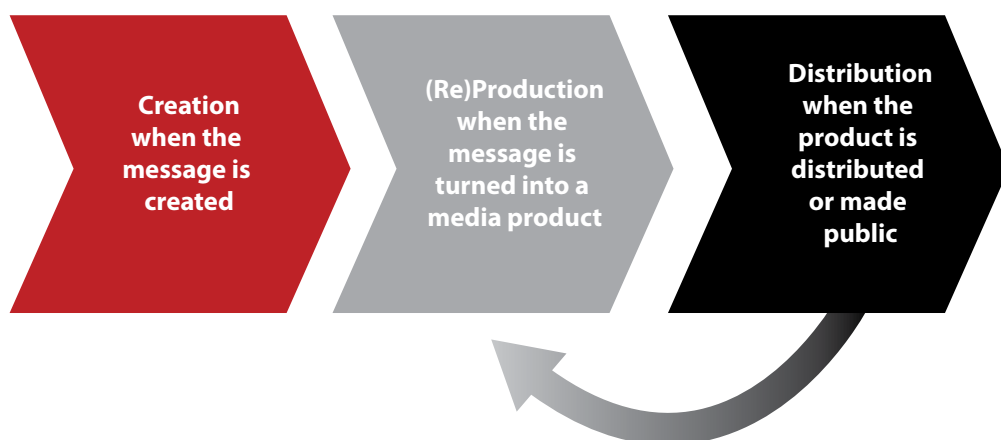


Figure 5. *Mechanisms of operation of fake news based on WARDE–DERAKHSHAN (2017)*

Facebook engineers described the process from the social media provider's side and identified it as information operations.⁴⁵ Information operations are carried out by state, economic or civil organizations, domestic or foreign, to achieve their political, economic or strategic goals. The methods of these information operations are various: they may take the form of fake news, misleading news or disinformation, but their goal is the same – to manipulate public opinion. They divided the process of spreading fake news into three phases: the first is targeted data collection, with respect to the way that van Dijck discusses datafication and Couldry and Hepp discuss deep mediatization.⁴⁶ IT analysis and evaluation of user data enables distributors of misleading news to define and reach micro-target groups (micro-targeting). The second is the creation of content and the wide distribution of this content through false amplifiers, i.e. with profiles of persons or fake organizations whose purpose is to deceive. These false amplifiers can have an impact not only with their shared content (posts), but also with their comments. In terms of impact assessments related to sharing, it is worth mentioning the 2017 research by Vargo, Guo and Amazeen, in which they considered whether fake news published between 2014 and 2016 had the power to determine the agenda on a wider scale. They carried out extensive data analysis which showed that the agenda setting effect of the fake news flow on the real news media was detectable on several occasions, although not in all cases.⁴⁷

Since 2016, public attitudes have remained largely unchanged, although there is some evidence of increasing ambivalence over time. This shift may be attributed to a decline in concerns about echo chambers and filter bubbles, even as high levels of anxiety about misinformation persist. Alternatively, this ambivalence could result from the growing use of multiple social networks, each employing algorithmic news selection in distinct ways, complicating the formation of consistently positive or negative views among users.⁴⁸

⁴⁵ WEEDON–NULAND–STAMOS (2017)

⁴⁶ VAN DIJCK (2014); COULDRY–HEPP (2017); ANDOK (2022).

⁴⁷ VARGO–GUO–AMAZEEN (2017)

⁴⁸ FLETCHER (2023)

5. Media bias and trust in the news media

Extensive research has been conducted to determine whether algorithmically driven platforms indeed overexpose users to like-minded views, while filtering out information that might be contrary to their beliefs, and thereby creating feedback loops that reshape their worldviews. However, recent studies on news exposure, particularly those by Ross Arguedas et al. (2022), suggest that this is not the case. Current evidence indicates that platform use tends to increase the diversity of individuals' news repertoires. Nevertheless, as platforms evolve, the debate persists, and the overall impact on users' attitudes and beliefs remains inadequately understood.⁴⁹

Additionally, there is limited knowledge regarding individuals' attitudes and beliefs about algorithmic news selection itself. This understanding is crucial, as many of the concerns about echo chambers and filter bubbles are based on the assumption that audiences are passive, credulous, and unreflective consumers of information. Media audiences, news consumers themselves have perceived the dysfunctions of mass communication, especially social media, in relation to the dissemination of news and information, which has caused a significant loss of trust in the system.

The Reuters Institute within the University of Oxford measures and analyses user data related to the news media every year. It is worthwhile to consider their most relevant data from the *Digital News Project* between 2017 and 2023. The Reuters Institute examined news consumers' distrust in news media across seven European countries, Australia and the United States. Participants named bias and polarity as the main reasons for distrust. About 67% of them expressed the opinion that those in power influence the news media in accordance with their own political or economic interests. Young people (under 35) and people with low socio-economic status in particular hold this view. News consumers trust the big brands of news television the most, and social media the least. Only 24% think the latter platform is reliable, with the main criticism being that it is unable to separate facts from opinion. Some 40% of all respondents believe that journalists do their job well. In the spring of 2020, during the first wave of the pandemic, the data showed that news consumers demonstrated a renewed appreciation for conventional news sources.⁵⁰

In general, trust in journalism has fallen by more than 20% in the United States over the past twenty years. According to Gallup's 1997 research, this figure was 53%, while in 2016 it was only 32%.⁵¹ At the same time, the gap between people with high or low incomes has grown: those with higher incomes are more satisfied with the news service and trust it more.

A further significant observation is that multi-point of view has become the norm for young people, which is an important characteristic of social media.⁵² The more detailed report clearly states that news consumers perceive the bias of the media, primarily in a political sense, and as a result they are quite distrustful of the news media.⁵³ In countries where political polarization is very strong, there is even less trust in the news media. The 2017 re-

⁴⁹ FLETCHER (2023)

⁵⁰ Digital News Report 2020: 11.

⁵¹ SWIFT (2016)

⁵² NEWMAN–FLETCHER (2017): 36.

⁵³ Digital News Report 2017

port mentions Hungary as third, after the United States and South Korea. However, it has also been shown that the filter bubble and the echo chamber effects are stronger for those who use social media than for those who do not. It is also appears that social media users more easily share news when they know that their friends' opinions about the news will be the same as their own.

In Hungary, between 2016 and 2020, the changes in media which were used as a news source were as follows: the use of the online medium as a news source fell from 88% to 86%, television from 72% to 67%, social media remained at 64%, while in the case of the printed press we see the largest drop from 27% to 15%.⁵⁴ The smartphone has become the primary tool of news consumption, and smartphones are now just as important in the home as they are outside. Regarding Hungary, this means that between 2016 and 2020, the number of people who primarily use smartphones for news consumption rose from 43% to 67%.⁵⁵ Although the role of social media as a news source has continued to grow over the past years, the rate of growth has slowed. Thus, the report treated as distinct categories the social network sites (Facebook, Instagram) and text messaging applications (Facebook Messenger, WhatsApp), the role of which is increasingly important in online news consumption. However, both forms are dominated by Facebook and Facebook Messenger. Two-thirds of users and news consumers remember the interface on which they found the news (Google, Facebook) rather than the brand of the news media itself.

The reports revealed several strategies for accessing news between 2016 and 2020, some of which are also typical of Hungarian news consumers:

1. Direct news consumption from news sites of digital media;
2. Social media as a source of news;
3. Actively searches for news using search engines;
4. News aggregator sites are mostly used in Japan (40%) and South Korea (37%);
5. They are informed about them via e-mail in Belgium (34%) and Portugal (34%);
6. Mobile notifications are relied upon in Taiwan (32%) and Hong Kong (27%);
7. Get information through the Messenger application.

In relation to Hungary, in 2020 they also measured which option involves a greater risk of encountering misleading information. According to the direct consumption of news from the news pages of digital media, 20% of Hungarian news consumers have a chance of encountering fake news. Referring to social media as a source of news, 40% of Hungarian news consumers believe that there is a chance of reading fake news. According to 10% of news consumers, if someone actively searches for news using search engines, there is a chance of finding misleading news. According to 14% of Hungarian news consumers, there is a chance of obtaining false information through the Messenger application.⁵⁶ On the whole, the recipients prefer text news (71%), 14% prefer news consisting of text and video, while only 7% prefer video news. The research examined the difference that can be measured between the general level of trust in the news media and the trust level of the sources used by the user. This means that on a global level, 38% of news consumers generally trust

⁵⁴ Digital News Report 2020: 72–73.

⁵⁵ Digital News Report 2020: 72.

⁵⁶ Digital News Report 2017; Digital News Report 2020: 18–19.

the news media and 46% trust the medium, which they read, listen to or watch themselves.⁵⁷ According to this aspect, Hungary ranks first in terms of the size of the difference. Trust in the news media is 27%, while in the case of the sources used by the person under analysis, it is 53%.⁵⁸ The study found a similarly large difference in the United States, where the two figures were 38 and 53%.

A clear distinction and separation of facts and opinions is a basic requirement conventional news services. In this regard, the report discusses the fact that new consumers consider the news media to be twice as credible as the social media in this aspect. The willingness to pay for news available online is low in Hungary, but it increased from 4% to 10% between 2016 and 2020.⁵⁹ This figure can be said to be similar across Europe, but in the United States, far more people – about 17% – are willing to do so. In the 2017 survey, the following answers were given to the question why they would rather not pay for online news:

1. Because I can access these contents for free (54%);
2. My favorite sites do not use pay-wall (29%);
3. Online news is not worth paying for (22%);
4. I cannot afford this expense (18%);
5. I don't want to use my bank card or credit card (11%);
6. I am not interested in the news (9%).

The researchers supplement this data series with the cases in which the respondents paid for online content in the past few years, and according to the results, 23% paid for video content (Netflix), 16% for audio content (Spotify), 15% for mobile applications and 11% for an online game. The research assessed the percentage of respondents who consciously avoid the news. At one end of the scale we find Greece with a figure of 57%, at the other end Japan with 6%, while in the case of Hungary this figure was 31%.

6. Research on the current situation in 2023

Regarding the data on the current situation, this is presented below on the basis of the 2023 Digital News Report of Reuters-Oxford university international research. Among the Hungarian research, the 2022 survey was carried out by the Association of Hungarian Content Providers, and the Media Market Report was published by the National Media and Information Communication Authority in 2023.

International data show the following 2023 trends: the most notable findings in this year's report pertain to the evolving nature of social media, evidenced by declining engagement with established platforms such as Facebook and the rising prominence of TikTok and other video-centric networks. Despite an increasing fragmentation of channels and the heightened public concern regarding misinformation and algorithms, dependence on these intermediaries continues to grow. Their data illustrate that this shift is significantly influenced by the habits of younger generations, who have grown up with social media and now

⁵⁷ Digital News Report 2020: 15.

⁵⁸ Digital News Report 2020: 72.

⁵⁹ Digital News Report 2020: 73.

frequently prioritize influencers or celebrities over journalists, even when choosing sources for their news consumption.⁶⁰

Publishers have exhibited uncertainty regarding the adaptation of their storytelling to new platforms. According to a Digital News Report 2023, approximately half of top publishers are now creating content for TikTok, while others remain hesitant due to concerns about Chinese government influence and the lack of monetization. Nonetheless, concerns about the unchecked dissemination of misinformation on the platform and the potential for engaging hard-to-reach younger audiences have persuaded some news organizations to establish a presence there, despite the associated risks.⁶¹ “Although the averages for TikTok are relatively low, usage is much higher with younger groups and in some Asia-Pacific, Latin American, and African countries. It has played a role in spreading both information and misinformation in recent elections in Kenya and Brazil and has grown strongly in parts of Eastern Europe, where the Ukrainian conflict has increased its profile.”⁶²

It is extremely interesting that the report uses the term *fake news* less in 2023, and is instead using the terms misinformation and disinformation. The Digital News Report 2023 summarizes as follows: the shifts in platform usage described, including the transition of attention to new networks, do not appear to alleviate public concerns about misinformation and disinformation. Across various markets, over half of the population (56%) express apprehension regarding their ability to discern real from fake news online, a 2 percentage point increase from the previous year. Notably, those who primarily rely on social media for news exhibit greater concern (64%) compared to those who do not use social media at all (50%). Countries with high levels of concern about misinformation also tend to have high levels of social media news usage.

While social media use is not necessarily the cause of misinformation, documented issues on these platforms and greater exposure to a wider range of sources appear to influence confidence in the accuracy of information encountered. Analysis of the types of misinformation reveals that dubious health claims related to COVID-19, including from anti-vaccination groups, remain common, along with false or misleading political information. In Slovakia, nearly half of the respondents reported encountering misinformation about the Ukraine conflict in the past week, a proportion roughly double that in the UK, United States, or Japan. Perceived misinformation about climate change is approximately three times higher in the United States (35%) compared to Japan (12%). Some prominent politicians, media opinion writers, and groups aligned with fossil fuel interests continue to dismiss the scientific consensus and undermine green policies. Environmental groups report that climate change denialism has seen a ‘stark comeback’ on social media, with misleading advertisements on Facebook and other networks and viral hashtags such as #climatescam on Twitter, which critics claim has failed to properly moderate harmful content since Elon Musk’s takeover of the platform, which is now called X.

In summary, the evolving platform environment reveals that our reliance on social media continues to grow, but a diverse array of platforms now competes to fulfill different roles in our lives, with news often becoming less central to their functions. Concurrently, there is increasing

⁶⁰ NEWMAN et al. (2023): 10.

⁶¹ NEWMAN et al. (2023): 13.

⁶² NEWMAN et al. (2023): 13–14.

dissatisfaction with how news content is selected (via algorithms), the accuracy of the content (misinformation), and the quality of debate (participation). Newer platforms like Instagram and TikTok, optimized for younger users with more visual content, frequently necessitate more tailored investment from publishers with minimal return in terms of traffic or revenue. With further innovations driven by Artificial Intelligence (AI) and automation on the horizon, publishers must increasingly focus on how these intermediaries can attract new users, foster deeper connections, and support their core business objectives.⁶³

Turning to the Hungarian situation, it is first useful to present the data for 2022.⁶⁴ The research conducted on a sample of 2,000 people in 2022 revealed the following regarding news consumption. Nearly two-thirds of the respondents read online news several times a day, and another 17% read it once a day. Among those reading several times a day, men, residents of Budapest and people with higher education are present in a significantly higher proportion than the average. Every second respondent reads online news on the web pages of national news portals, but the proportion of those viewing social media shares or online applications of news sites, watching TV news and listening to radio news is also high.

Regarding trust in news sources, respondents have the greatest trust in news from television, radio and national news portals. Looking at the entire sample of respondents, the proportion of those who trust news shared by a personal acquaintance is also particularly high. Where trust in each potential news source is analyzed only among the respondents who consume the given news source, television, radio, and online portals are still considered the most reliable sources. The position of the news shared by a personal acquaintance changes compared to the result measured by the entire sample – it becomes the least reliable news source. Most of the respondents spend more than 1 to 2 hours a day watching TV. Some 34% of respondents spend 1 to 2 hours per day browsing social media, 31% less than 1 hour, and 16% spend more than 2 hours per day on social networks. About two-thirds of respondents use social media for information purposes (reading, listening to videos and for viewing). Slightly more than a quarter of them are accustomed to reacting to shared content, 18% share content they find themselves, while only 10% actually reshare content which has been shared with them. The highest proportion of respondents open and share scientific articles on social media, followed by content related to health, politics, economy and recipes.

In terms of checking the authenticity of news, 32% of respondents check the news they read regularly for authenticity, while another 54% do so occasionally. Such checks typically consist of a search for the publication of the news on another portal, or checking the authenticity of a site itself; only about one out of four respondents uses fact-checking sites. The majority of people feel considerably more protected against fake news than the level of protection that they perceive their acquaintances or the entire population of the country have.

A majority of respondents, 53%, believe that they are confronted with fake news almost every day. Another 23% feel that they have come across fake news several times a week, and 9% at least once a week. Media consumption habits strongly influence opinions

⁶³ NEWMAN et al. (2023): 17.

⁶⁴ Fake news questionnaire survey for the Association of Hungarian Content Providers (MTE): Fake news – Kérdőíves kutatás a Magyar Tartalomszolgáltatók Egyesülete részére. Gemius, February 2022. https://mte.hu/wp-content/uploads/2022/03/gemiusAdhoc_MTE_fakenews_2022_02_FIN.pdf

about fake news. Those who read contents in foreign languages and read the news several times a day are significantly more likely than the average to think that they encounter fake news almost every day, while among ‘hardcore’ TV viewers, the proportion of those claiming the same falls considerably below the average. Based on the responses, the responsibility of standing up against fake news falls (or should fall) primarily on the media – including the management, newspapers, journalists, professional organizations. Compared to the average, a significantly larger percentage of people holding diplomas of higher education consider that the task of fighting against fake news belongs to the media. However, only 14% of the respondents would trust that a public organization against fake news could operate successfully.

When considering science vs. conspiracy theory, international research has clearly shown that there is a connection between conspiracy theorizing and consumption of fake news. During a cluster analysis, respondents were divided into two large groups based on nine questions about attitudes. Those open to science ‘believe’ in the existence of fake news, so they check what appears in the media assertions and they are generally less likely to automatically accept what others assert. According to *conteo* believers, there is no significant problem with fake news and they have a tendency to accept news published without verification. Based on the responses, it appears that people susceptible to conspiracy theories are significantly more active on social media, than those groups which are open to science.

Hungarian-related data can also be found in the 2023 Media Market Report of the NMHH (National Media and Infocommunication Authority). The diversity of news consumption is indicated by the fact that 4% of the population aged 15 to 75 source their information from all four types of media – television, radio, print and online media – 24% from three, and 44% from two types of media. Only 23% get their information from a single type, while 3% of the population sample does not receive information about world affairs from any type of media. The Internet is an increasingly important source of information: the proportion of combinations types of media where online media is included increased to 94% from 92%, as measured in 2021. At the same time, it is undeniable that the constantly expanding range of media on offer also carries risks. Some 74% of Internet users over the age of 16 who use social networking sites have encountered harmful content and expressions, with the largest proportion being confronted with fake news (54%) and fraudulent ads (46%). While 37% of users reported the incident to the operator of the site, and the same proportion blocked the publisher or unsubscribed, a third of users did not take any action. The spread of fake news can also be facilitated by cognitive, social and algorithmic biases. This is complemented by the phenomenon that the spread of information among social media users takes place largely in homogeneous clusters (*filter bubbles*), where users mostly trust their close friends and acquaintances.⁶⁵

Although the problem of fake news is not new, the burden it creates has been growing significantly in this age of the Internet and social media. It is no coincidence that this challenge now plays a prominent role in public debates, in the press and in socio-political public thinking. At the same time, it is difficult to measure the effects and social significance of fake news using empirical methods for technical and research methodological reasons. Ac-

⁶⁵ Media Market Report 2023. NMHH, 2023. 15. https://english.nmhh.hu/document/242930/NMHH_Media_Market_Report_2023.pdf

According to the research, among those who reported harmful content and expressions to some authority, most of them turned to the police (48%) or the NMHH and its various institutions (26%). The material compiled on behalf of the NMHH summarizing the results of various studies dealing with the topic concluded that potentially impactful *fake news* affects only a relatively small number of users, but the extent to which different age groups and social groups are impacted differs greatly –depending on the level of media competence and understanding.⁶⁶ Research has shown that when it comes to news shared on Facebook, the person sharing the news matters more than the original source of the news: users are more likely to believe that the news is accurate and balanced when it is shared by someone they know and trust. Social media algorithms reinforce this phenomenon by favoring the interactions of close acquaintances, so in the end it is the popularity of a particular news item which is important, not its truthfulness and reliability.

The effect of fake news on voter behavior is somewhat limited, precisely because of the closed filter bubbles. However, the indirect consequences of this phenomenon are far more worrying: it increases the level of cynicism, apathy and mistrust, confuses and overwhelms media consumers, causes mental fatigue, and ultimately contributes to the growth of extremism and social polarization. The spread of fake information and fake news keeps the press and traditional media in a double bind: on the one hand, they have to constantly refute fake news spreading on the Internet (which takes resources away from the distribution of real news), and on the other hand, they have to fight the general loss of trust caused by fake news.

Some 24% of the respondents of the aforementioned research conducted by Scores Group experience a that a social media site deleted or blocked a content they had uploaded. However, more than a third of those sanctioned (37%) did not receive any justification related to the decision. 7% of the Internet users included in the research found themselves in a situation where a social networking site deleted or disabled their profile, and in 43% of such cases, the social networking site did not even justify the decision.

Commissioned by NMHH, Scale Research researched moderation practices on global online platforms in 2022, using the trial purchase method. Despite the fact that the basic moderation rules for individual social media interfaces are set out in user regulations and in documents related to community guidelines, the research shows that implemented practice does not necessarily comply with the guidelines. For example, according to the experience of the test buyers, most of the platforms also blocked posts whose content did not violate the guidelines set by the given site, but there were also examples where the social media site did not block posts that they should have removed according to their own rules. In their rules, social media sites give considerable space to subjective decision-making, meaning that whether posts are compliant is contingent upon this subjectivity. Such an approach does not result in consistent moderation practice.

Those without a high school diploma have more than average trust in the print media and TV news, while graduates trust national news sites. People who watch a great deal of TV have more trust than average in TV news and printed newspapers, while among those who do not watch TV at all, trust in all news sources is lower than average. 85% of the respondents believe that they come across fake news at least weekly, most of them ‘almost every day’ or ‘several times a week’. A slightly higher proportion

⁶⁶ Media Market Report 2023: 115.

of men (57%) think that they encounter fake news almost every day than women (51%). Among the age groups, the youngest (39%) believe that they discover fake news among news content the least. The level of trust in individual news sources depends on education and media consumption habits.⁶⁷

7. Artificial Intelligence and disinformation

In his study, Andreas Kaplan examines the impact on democracy of technological developments in the field of Artificial Intelligence. He mentions three serious sources of danger – supervision, manipulation and frustration – which can be defended against in three main ways. The options for this defence are related to the technology itself, regulation and education.

“Artificial Intelligence represents a threat to democracy which needs to be taken seriously. There are at least three areas in which AI might be a danger to democratic life and mechanisms: supervision, manipulation, and frustration. Firstly, states now have very advanced means of controlling and supervising their citizens’ daily behavior, which could be abused by governments to limit freedoms. Secondly, citizens can increasingly be manipulated in their voting behavior by ample use of Artificial Intelligence and social media. Thirdly and finally, such AI-driven supervision and manipulation can, in short, lead to citizens’ frustration and their deciding to no longer take part in democracy.”⁶⁸

Democracy can be protected from the negative effects of AI by developing technology in the right direction, carefully regulating it, and educating users about media literacy.

“Protecting Democracy: Technology, Regulation, Education – the previous section clearly shows that Artificial Intelligence definitively represents a threat to democracy. Next, three areas will be discussed that could help to avoid misuse of AI and to limit its danger to democracy: technology, regulation, and education. Technology and AI itself can be applied to detect unethical and illegal behaviour. Regulation will be necessary and has the potential to define voter manipulation and the dissemination of fake news. Finally, the question also arises of how to educate (future) citizens and make them more conscious of various manipulation techniques.”⁶⁹

According to other researchers, among the four most harmful effects on the public and democracy are fake news, disinformation, and deep fakes.⁷⁰

The appearance of Artificial Intelligence tools in journalism and media content production is not new. Many AI tools have been used in the media world for decades, from automatic cameras to translation programs or content recommendation systems. Artificial In-

⁶⁷ Media Market Report 2023: 178.

⁶⁸ KAPLAN (2020): 153.

⁶⁹ KAPLAN (2020): 156.

⁷⁰ KERTYSOVA (2018): 64.

telligence as a tool is related to the issue of disinformation in two ways: for one, it helps to identify and detect fake news and sources of disinformation, and from this point of view, it has a positive effect in the social sense; for another, AI tools are also suitable for supporting, helping and making the sharing of fake news and disinformation more effective.⁷¹ This duality is dealt with in depth in media theory literature and research.⁷²

In the context of information operations, AI solutions have been particularly effective in detecting and removing illegal, dubious, and undesirable content online. AI techniques have also been successful in screening for and identifying fake bot accounts – techniques known as bot-spotting and bot-labeling.⁷³ In their examination of disinformation detection methods, Bontridder et al. evaluate four distinct approaches.⁷⁴ Firstly, they discuss the use of end-to-end machine learning models trained on labeled data, which consist of articles identified as either false or accurate. These models can autonomously differentiate between true and false articles. However, there are significant limitations to this method: for one, it relies on large volumes of labeled data, which can be challenging to obtain; and for another, the model outputs lack transparency and are susceptible to biases inherent in the datasets. The second of the four approaches considered is the detection of factual inaccuracies through content comparison with external evidence. This method is more effectively executed by human fact-checkers due to the complexities and subtleties of natural language, which are difficult to codify. Nevertheless, automated approaches can also be employed, such as verifying claims against provided sources or utilizing unstructured external web information. The third approach is the identification of misleading stylistic elements – inferring the article's intent through stylistic analysis – can be performed by trained human experts or machine learning algorithms. However, it is important to note that the presence of a misleading style does not necessarily correlate with the accuracy or truthfulness of the information presented. Fourthly and finally, the analysis of metadata, such as the profile and attributes of the individual sharing the content, is explored as an alternative to content analysis.⁷⁵

Montoro and his colleagues list seven ways in which AI can contribute to fact-checking.⁷⁶ These seven are as follows: (1) monitoring, recognition, and prioritization of content susceptible to verification; (2) evaluating whether claims are verifiable or not and topic prioritization; (3) searching for previous successful verification that applies to the same case; (4) retrieval of evidence for further investigation; (5) semi-automated classification in categories (hoax, misleading content, false context, etc.); (6) dissemination of successful verification; and (7) speeding up writing and documenting the fact-checks.⁷⁷

Santos and Carrilho suggested in their 2023 paper that fact checking should be supplemented with language and sentiment analysis.⁷⁸ “In addition to automated fact-checking, which may also incorporate language analysis and sentiment analysis, the fight against disinformation requires a multifaceted approach that involves not only

⁷¹ AL-ASADI–TASDEMIR (2022)

⁷² ACZÉL–VESZELSZKI (2023)

⁷³ KERTYSOVA (2018): 59.

⁷⁴ BONTRIDDER–PUOLLET (2021)

⁷⁵ BONTRIDDER–PUOLLET (2021): 32.

⁷⁶ MONTORO et al. (2023)

⁷⁷ MONTORO et al. (2023): 123.

⁷⁸ SANTOS–CARRILHO (2023)

the use of AI and other technological tools but also human verification.”⁷⁹ As they delve into the subsequent section on the intersection of AI and blockchain technologies, automated fact-checking, especially when combined with these two technologies, is shown to be able to encompass the three key stages of conventional fact-checking: identification, verification, and distribution.⁸⁰

There are several limitations to the application of automated techniques to detect and counter disinformation. The first significant shortcoming is the risk of over-blocking lawful and accurate content – the ‘over-inclusiveness’ feature of AI.⁸¹ Observers warn that if this is left unchecked, bias in algorithms may lead to decisions which can have a collective, disparate impact on certain groups of people. For instance, algorithms trained on historical datasets have shown to replicate social biases, notably those against women, which then influence decisions made by computers ranging from recruitment for jobs to mortgages.⁸²

When AI techniques are used to create fake content, the product is called a *deepfake*.⁸³ Deepfakes (a portmanteau of deep learning and fake) are the product of two AI algorithms working together in a so-called Generative Adversarial Network (GAN). GANs are best described as a way to algorithmically generate new types of data from existing datasets. For example, a GAN could analyse thousands of pictures of Donald Trump and then generate a new picture that is similar to the images analyzed but not an exact copy of any of them. This technology can be applied to various types of content – images, moving images, sound, and text, but the term deepfake is primarily used for audio and video content. Deep Fakes Application of AI to audio and video content production presents an even greater challenge. The so-called ‘deep fakes’ – digitally manipulated audio or visual material that is highly realistic and virtually indistinguishable from real material – were initially used in the movie industry. However by 2030, deep fakes could become indistinguishable from genuine information and easier to produce. Telling the difference between the original and manipulated content may become close to impossible for news consumers, and progressively difficult for machines.⁸⁴ Deep fakes make it possible for malign actors to deny the truth in two ways: not only may fake videos be passed off as real in order to create doubt, but it can be claimed that authentic information is fake. As the public becomes more educated about the threats posed by deep fakes, the latter technique is likely to become more plausible.⁸⁵ In any event, all researchers without exception agree that media literacy helps people to identify fake news and it is therefore imperative to improve people’s news literacy.⁸⁶

⁷⁹ SANTOS–CARRILHO (2023): 681.

⁸⁰ SANTOS–CARRILHO (2023): 684.

⁸¹ KERTYSOVA (2018): 60.

⁸² KERTYSOVA (2018): 61.

⁸³ BONTRIDDER–PUOLLET (2021): 32.

⁸⁴ KERTYSOVA (2018): 66.

⁸⁵ KERTYSOVA (2018): 68.

⁸⁶ AÏMEUR–AMRI–BRASSARD (2023); HAMELEERS et al (2022).

8. Conclusions

A number of stakeholders are attempting to address the biased nature of news consumption and the explosive spread of fake news. These endeavors include initiatives from the news media themselves or ones implemented through cooperation between civilians and media professionals. Among the solutions proposed, there are those which emphasize technological, socio-societal, media-centric, regulatory or educational approaches.⁸⁷

In addition, there are many examples of fact-checking sites, for example the US-based <https://www.factcheck.org>, some of which have been set up specifically to expose the lies of political opponents. We also find an example where news aggregator site collects news from media with a markedly different orientation in relation to each news item, and provides the reader with a comparative evaluation, such as the All Sides initiative (<https://www.allsides.com/unbiased-balanced-news>). Overall, it appears that a repair-corrective mechanism has begun to take shape from within the information system, through cooperation of industry players and users, political players and organizations responsible for education.

References

- ACZÉL (2017) = ACZÉL Petra: Az álhír. Kommentár a jelenség értelmezéséhez. *Századvég*, 2017/2. 5–25.
- ACZÉL–VESZELSZKI (2023) = ACZÉL Petra – VESZELSZKI Ágnes: *Deepfake. A valótlan valóság*. Gondolat, Budapest, 2023.
- AÏMEUR–AMRI–BRASSARD (2023) = Esma AÏMEUR – Sabrine AMRI – Gilles BRASSARD: Fake news, disinformation and misinformation in social media. A review. *Social Network Analysis and Mining*, Vol. 13. (2023) Art. no. 30. <https://doi.org/10.1007/s13278-023-01028-5>
- AKHTAR et al. (2023) = Pervaiz AKHTAR – Arsalan Mujahid GHOURI – Haseeb Ur Rhman KHAN – Mirza AMIN UL HAQ – Usama AWAN – Nadia ZAHOOB – Zaheer KHAN – Aniqua ASHRAF: Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions. *Annals of Operations Research*, Vol. 327. (2023) 633–657. <https://doi.org/10.1007/s10479-022-05015-5>
- AL-ASADI–TASDEMIR (2022) = Mustafa A. AL-ASADI – Sakir TASDEMIR: Using artificial intelligence against the phenomenon of fake news. A systematic literature review. In: Mohamed LAHBY – Al-Sakib Khan PATHAN – Yassine MALEH – Wael Mohamed Shaheer YAFOOZ (Eds.): *Combating fake news with computational intelligence techniques*. Cham, Springer, 2022. 39–54. https://doi.org/10.1007/978-3-030-90087-8_2
- ALLAN (2010) = Stuart ALLAN: *News Culture*. Third edition. Berkshire, McGraw-Hill, 2010.
- ALLCOTT–GENTZKOW (2017) = Hunt ALLCOTT – Matthew GENTZKOW: Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, Vol. 31., No. 2. (2017) 211–236. <https://doi.org/10.1257/jep.31.2.211>
- ANDOK (2013) = ANDOK Mónika: *A hírek története*. Budapest, L'Harmattan, 2013.
- ANDOK (2017) = ANDOK Mónika: *A kommunikáció rituális elmélete*. Budapest, Gondolat, 2017.

⁸⁷ ACZÉL (2017): 13.

- ANDOK (2020) = ANDOK Mónika: A társadalmi nyilvánosság kutatásának affektív fordulata. *Jel-Kép*, 2020/2. 23–38. <https://doi.org/10.20520/JEL-KEP.2020.2.23>
- BELL (2009) = Catherine BELL: *Ritual theory and ritual practice*. (Foreword by Diane JONTE-PACE). Oxford – New York, Oxford University Press, 2009 [1992].
- BELTING (2009) = Hans BELTING: *A hiteles kép. Képviták mint hitviták*. Budapest, Atlantisz, 2009.
- BONTRIDDER–POULLET (2021) = Noemi BONTRIDDER – Yves POULLET: The role of artificial intelligence in disinformation. *Data & Policy*, Vol. 3. (2021) Art. no. 32. <https://doi.org/10.1017/dap.2021.20>
- CAREY (2009) = James W. CAREY: *Communication as Culture. Essays on Media and Society*. Revised edition. New York – London, Routledge, 2009. <https://doi.org/10.4324/9780203928912>
- COMAN–COMAN (2017) = Ioana A. COMAN – Mihai COMAN: Religion, popular culture and social media. The construction of a religious leader image on Facebook. *ESSACHESS Journal for Communication Studies*, Vol. 10., No. 2. (2017) 129–143. <https://essachess.com/3/index.php/jcs/article/view/208>
- COULDRY (2003) = Nick COULDRY: *Media rituals. A critical approach*. London – New York, Routledge, 2003.
- COULDRY–HEPP (2017) = Nick COULDRY – Andreas HEPP: *The mediated construction of reality*. Cambridge, Polity Press, 2017.
- DAFONTE–GÓMEZ (2018) = Alberto DAFONTE–GÓMEZ: Audience as medium. Motivations and emotions in news sharing. *International Journal of Communication*, Vol. 12. (2018) 2133–2152. <https://ijoc.org/index.php/ijoc/article/view/6790/2351>
- FEHÉR–KIRÁLY (2017) = FEHÉR Katalin – KIRÁLY Olívia: Álhíresülés. A hamis hírek dinamikája a médiában. *Századvég*, 2017/2. 39–48.
- FERENCZY (1887) = FERENCZY József: *A magyar hírlapirodalom története 1780–1867*. Budapest, Lauffer V., 1887.
- FISKE (2011) = John FISKE: *Television culture*. Second edition. London – New York, Routledge, 2011.
- FLETCHER (2023) = Richard FLETCHER: Attitudes to algorithms and their impact on news. In: Nic NEWMAN – Richard FLETCHER – Kirsten EDDY – Craig T. ROBERTSON – Rasmus KLEIS NIELSEN: *Reuters Institute Digital News Report, 2023*. Oxford, Reuters Institute, 2023. 32–35. <https://doi.org/10.60625/risj-p6es-hb13>
- GERENCSÉR–SZŰTS (2020) = GERENCSÉR Péter – SZŰTS Zoltán: A társadalmi nyilvánosság webkettes szerkezetváltozása. *Jel-Kép*, 2020/2. 39–51. <https://doi.org/10.20520/JEL-KEP.2020.2.39>
- HALL (2007) = Stuart HALL: Kódolás – dekódolás. In: ANGELUSZ Róbert – TARDOS Róbert – TERESTYÉNI Tamás (szerk.): *Média, nyilvánosság, közvélemény. Szöveggyűjtemény*. Gondolat, Budapest, 2007. 131–142.

- HAMELEERS (2022) = Michael HAMELEERS: 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*, Vol. 25., No. 1. (2022) 110–126. <https://doi.org/10.1080/1369118X.2020.1764603>
- KAPLAN (2020) = Andreas KAPLAN: Chapter 9. Artificial intelligence, social media, and fake news. Is this the end of democracy? In: Ayşen GÜL – Yıldız ERTÜRK – Paul ELMER: *Digital transformation in media and society*. Istanbul, Istanbul University Press, 2020. 149–161. <https://doi.org/10.26650/b/ss07.2020.013.09>
- KERTYSOVA (2018) = Katarina KERTYSOVA: Artificial intelligence and disinformation. How AI changes the way disinformation is produced, disseminated, and can be countered. *Security and Human Rights*, Vol. 29., No. 1–4. (2018) 55–81. <https://doi.org/10.1163/18750230-02901005>
- KISS (2013) = Kiss Balázs: Érzelmek és politikatudomány. *Politikatudományi Szemle*, 2019/3. 7–28. <https://doi.org/10.30718/POLTUD.HU.2019.3.7>
- KÜNG (2017) = Lucy KÜNG: *Going digital. A roadmap for organisational transformation*. Oxford, Reuters Institute, 2017. <https://doi.org/10.60625/risj-xce5-7e64>
- MONTORO-MONTARROSO et al. (2023) = Andres MONTORO-MONTARROSO – Javier CANTÓN-CORREA – Paolo ROSSO – Berta CHULVI – Ángel PANIZO-LLEDOT – Javier HUERTAS-TATO – Blanca CALVO-FIGUERAS – M. José REMENTERIA – Juan GÓMEZ-ROMERO: Fighting disinformation with artificial intelligence. Fundamentals, advances and challenges. *Profesional de la información*, Vol. 32., No. 2. (2023) <https://doi.org/10.3145/epi.2023.may.22>
- NABI (2003) = Robin L. NABI: Exploring the framing effects of emotion. Do discrete emotions differentially influence information accessibility, information seeking, and policy preference? *Communication Research*, Vol. 30., No. 2. (2003) 224–247. <https://doi.org/10.1177/0093650202250881>
- NAGEL (2022) = Tyler W. NAGEL: Measuring fake news acumen using a news media literacy instrument. *Journal of Media Literacy and Education*, Vol. 14., No. 1. (2022) 29–42. <https://doi.org/10.23860/JMLE-2022-14-1-3>
- NEWMAN et al. (2017) = Nic NEWMAN – Richard FLETCHER – Antonis KALOGEROPOULOS – David A. L. LEVY – Rosmus KLEIS NIELSEN: *Reuters Institute Digital News Report 2017*. Oxford, Reuters Institute, 2017. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf
- NEWMAN et al. (2020) = Nic NEWMAN – Richard FLETCHER – Anne SCHULZ – SIMGE Andi – Rosmus KLEIS NIELSEN: *Reuters Institute Digital News Report 2020*. Oxford, Reuters Institute, 2020. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf
- NEWMAN et al. (2023) = Nic NEWMAN – Richard FLETCHER – Kirsten EDDY – Craig T. ROBERTSON – Rasmus KLEIS NIELSEN: *Reuters Institute Digital News Report 2023*. Oxford, Reuters Institute, 2023. <https://doi.org/10.60625/risj-p6es-hb13>

- NEWMAN–FLETCHER (2017) = Nic NEWMAN – Richard FLETCHER: *Bias, bullshit and lies. Audience perspectives on low trust in the media. (Digital news project.)* Oxford, Reuters Institute, 2017. <https://doi.org/10.2139/ssrn.3173579>
- NYGREN–BROUNÉUS–SVENSSON (2019) = Tomas NYGREN – Fredrik BROUNÉUS – Göran SVENSSON: Diversity and credibility in young people's news feeds. A foundation for teaching and learning citizenship in a digital era. *Journal of Social Science Education*, Vol. 18., No. 2. (2019) 87–109. <https://doi.org/10.4119/jsse-917>
- PAPACHARISSI (2014) = Zizi PAPACHARISSI: *Affective publics. Sentiment, technology, and politics.* New York, Oxford University Press, 2014.
<https://doi.org/10.1093/acprof:oso/9780199999736.001.0001>
- PARISER (2011) = Eli PARISER: *The filter bubble. What the internet is hiding from you.* New York, Penguin Press, 2011.
- PIRITYI (1996) = PIRITYI Sándor: *A nemzeti hírügynökség története 1880–1996.* Budapest, MTI, 1996.
- SANTOS (2023) = Fátima C. Carrilho SANTOS: Artificial intelligence in automated detection of disinformation. A thematic analysis. *Journalism and media*, Vol. 4., No. 2. (2023) 679–687. <https://doi.org/10.3390/journalmedia4020043>
- SHARMA et al. (2020) = Karishma SHARMA – Sungyong SEO – Chuizeng MENG – Shirisha RAMBHATLA – Yan LIU: Covid-19 on social media. Analyzing misinformation in Twitter conversations. *arXiv*, (2023) 2293–2309.
<https://doi.org/10.48550/arXiv.2003.12309>
- SHU et al. (2017) = Kai SHU – Amy SLIVA – Suhang WANG – Jilian TANG – Huan LIU: Fake news detection on social media. A data mining perspective. *ACM SIGKDD Explorer Newsletter*, Vol. 19., No. 1. (2017) 22–36. <https://doi.org/10.1145/3137597.3137600>
- SUNSTEIN (2013) = Cass R. SUNSTEIN: *Republic.com 2.0.* Budapest, CompLex, 2013.
- SWIFT (2016) = Art SWIFT: Americans' trust in mass media sinks to new low. *Gallup*, 14 September, 2016.
<http://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx>
- SZÜTS (2018) = SZÜTS Zoltán: *Online. Az internetes kommunikáció és média története, elmélete és jelenségei.* Budapest, Wolters Kluwer, 2018.
- TÓFALVY (2017) = TÓFALVY Tamás: *A digitális jó és rossz születése. Technológia, kultúra és az újságírás 21. századi átalakulása.* Budapest, L'Harmattan, 2017.
- VAN DIJCK (2014) = Jose VAN DIJCK: Datafication, dataism and dataveillance. Big Data between scientific paradigm and ideology. *Surveillance & Society*, Vol. 12., No. 2. (2014) 197–208. <https://doi.org/10.24908/ss.v12i2.4776>
- VARGO–GUO–AMAZEEN (2017) = Chris J. VARGO – Lei GUO – Michelle A. AMAZEEN: The agenda-setting power of fake news. A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, Vol. 20., No. 5. (2017) 2028–2049.
<https://doi.org/10.1177/1461444817712086>

- VESZELSZKI (2017) = VESZELSZKI Ágnes: Az álhírek extra- és intralingvális jellemzői. *Századvég*, 2017/2. 51–82.
- WAGNER–BOCZKOWSKI (2019) = María Celeste WAGNER – Pablo J. BOCZKOWSKI: The reception of fake news: The interpretations and practices that shape the consumption of perceived misinformation. *Digital Journalism*, Vol. 7., No. 7. (2019) 870–885.
<https://doi.org/10.1080/21670811.2019.1653208>
- WALLACE (2016) = Patricia WALLACE: *The psychology of the internet*. Second edition. New York, Cambridge University Press, 2016.
- WARDE–DERAKHSHAN (2017) = Claire WARDE – Hossein DERAKHSHAN: *Information disorder. Toward an interdisciplinary framework research and policy making*. (Council of Europe report DGI 09). Strasbourg, Council of Europe, 2017.
- WEEDON–NULAND–STAMOS (2017) = Jen WEEDON – William NULAND – Alex STAMOS: *Information Operations and Facebook*. Facebook Inc., 2017.

Disinformation and its legal implications related to freedom of expression

Four issues for consideration regarding regulation and conceptualization

GERGELY FERENC LENDVAI

1. Introduction

Noting the fact that in the digital age, the proliferation of disinformation has emerged as a pressing legal concern is no scholarly novum. However, the profound interpretation of the effects of disinformation, such as the intricate interplay between misrepresentation of facts and the cherished principle of freedom of expression, proposes more questions than answers to researchers, policymakers, and billions of users around the world. This paper aims to examine the exceedingly complex landscape where these two domains intersect, aiming to elucidate the multifaceted legal implications of disinformation for the fundamental right of free of expression. Disinformation – though not a new phenomenon¹ – spreads with unprecedented speed and even more unparalleled gravity,² leading to consequences such as the disruption and manipulation of public discourse, the erosion of trust in news and media, the harmful often systemic dissemination of false narratives in politics manifesting in political fake news and disinformation propaganda campaigns, and the fabrication of communication perspectives.³ The question rightly arises – if fake news is such a notable and proven harm regarding democratic values and social cohesion, how can the law and regulation specifically answer the aforementioned challenges? The importance of this question lies in the comprehension of the legal dimensions of disinformation within the context of freedom of expression and the challenges related to its limitations.

As enshrined in the relevant international human rights frameworks from the IC-CPR,⁴ through Europe⁵ to the Americas⁶ or Africa⁷ and national constitutions, freedom of

¹ FALLIS (2015): 402.

² BALL (2018)

³ KAPANTAI et al. (2020): 1303–1304.

⁴ International Covenant on Civil and Political Rights, 16 December 1966. <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>

⁵ European Convention on Human Rights (1950). https://www.echr.coe.int/documents/d/echr/convention_ENG

⁶ American Convention on Human Rights “Pact of San Jose, Costa Rica” (B-32), (1969). https://www.oas.org/dil/treaties_b-32_american_convention_on_human_rights.pdf

⁷ African Charter on Human and Peoples’ Rights (1986). https://au.int/sites/default/files/treaties/36390-treaty-0011_-_african_charter_on_human_and_peoples_rights_e.pdf

expression serves as the cornerstone of democratic societies.⁸ Though not an absolute right,⁹ the scope of interference with someone's freedom of expression is narrow.¹⁰ This narrow possibility of limitation lays the ground for further polemics. Though disinformation poses a significant threat to the integrity of information ecosystems, it also raises intricate legal questions regarding the boundaries of speech regulation, with respect to the striking of a delicate balance between combating disinformation and upholding free expression. This is a critical challenge that requires careful consideration of legal frameworks, regulatory measures, and ethical considerations.¹¹ This study explores the legal frameworks that govern disinformation and freedom of expression. It will encompass international human rights law, national legislation, and evolving regulations, as well as the role of private entities like Meta and Twitter in formulating content policies. Additionally, the study will delve into complex issues such as defining disinformation, assigning liability for its dissemination, and the roles of regulatory bodies, media, and technology companies in mitigating its effects.

2. Four issues to consider – from the pervasive nature of conceptualization to the Facebook Oversight Board

The selection of these four specific issues stems from their profound implications at the confluence of disinformation and freedom of expression, reflecting the complex challenges that necessitate a thorough examination. The conceptual challenges highlight the foundational necessity of a clear understanding of disinformation's (proposed) boundaries to ensure regulation remains effective without infringing on free speech rights.

Firstly, the conceptual challenges arising from the ever-evolving nature of disinformation demand a precise delineation of its scope, encompassing misinformation, satire or comical usage of disinformation, and legitimate expression. This issue calls for a perspicacious balance to prevent overreach in regulation that might inadvertently suppress freedom of expression. Secondly, the liability quandary emerges when attempting to hold individuals and platforms accountable for disseminating disinformation. This poses crucial questions regarding the boundaries of responsibility and the role of intermediaries. The third issue explores regulatory ventures, scrutinizing the effectiveness and potential pitfalls of governmental and non-governmental measures in combating disinformation without curtailing freedom of expression. Furthermore, the concept of 'platform courts' emerges as a novel mechanism for dispute resolution, engendering the fourth issue.¹² These digital tribunals, operating beyond traditional legal realms, raise questions of jurisdiction, fairness, and transparency while addressing disinformation-related disputes.¹³ Lastly, the chilling effect underscores how disinformation's pervasiveness may inadvertently suppress legitimate discourse as individuals fear legal consequences or social backlash. The discourse on chilling effect ac-

⁸ CASSESSE (2022): 10.

⁹ GUNATILLEKE (2021)

¹⁰ MENDEL (2010): 7–9.

¹¹ KOLTAY (2009)

¹² MAZUR–GRAMBLIČKOVÁ (2023)

¹³ E.g. SCHULTZ (2021).

centuates the delicate equilibrium required between curbing disinformation and preserving an environment conducive to open dialogue.

3. Unmasking disinformation: Striking at the heart of freedom of expression – Four inquiries to consider

As mentioned above, the core of the paper is the understanding of four critical issues that regulators and users face on a daily basis concerning disinformation and the regulation related thereto. These four issues collectively form the bedrock upon which the research is built, delving into their intricate connections, legal ramifications, and potential resolutions. By addressing these concerns, this study aims to advance the discourse surrounding the intricate interplay of disinformation and freedom of expression, navigating a complex landscape where the preservation of democratic values and an informed public remain paramount.

3.1. Conceptual challenges

There are conceptual challenges inherent in defining nearly all new media phenomena – disinformation is no exception. These challenges can be traced back to the very essence of all digital phenomena, however, in the context of disinformation specifically, defining the issue has emerged as a multifaceted phenomenon that defies easy categorization due to its intricate interplay with truth, falsehoods, and subjective interpretations of information.¹⁴ The very nature of disinformation involves the intentional distortion or manipulation of facts with the aim of deceiving or misleading the public,¹⁵ making its demarcation a daunting task. When making an attempt to define disinformation, it is of central importance to understand and consider the complex landscape in which disinformation operates and the subsequent impact it has on freedom of expression.¹⁶ It is also essential to identify which features distinguish disinformation from misinformation, as well as from misrepresentation of fact – terms which are often interconnected, but not absolutely interchangeable.¹⁷

As a premise it is crucial to state that at the heart of these conceptual challenges lies the dynamic and fluid nature of information dissemination in the digital era.¹⁸ Unlike traditional media, digital platforms empower a diverse range of actors, from individuals to state-sponsored entities, to disseminate information on a global scale without gatekeeping mechanisms.¹⁹ This democratization of information-sharing, while fostering a richer public discourse, simultaneously enables the rapid spread of unverified or deliberately fabricated content, and, consequently, the task of distinguishing between legitimate free speech and

¹⁴ HAMELEERS (2023)

¹⁵ LENDVAI (2023a)

¹⁶ JACOBS (2022): 280–286.

¹⁷ SHU et al. (2020)

¹⁸ AÏMEUR–AMRI–BRASSARD (2023)

¹⁹ See: WILDING et al. (2018).

harmful disinformation becomes intricate, as the line between expressing opinions and propagating falsehoods is blurred. Defining disinformation also intersects with broader questions of intent and the impact of the information in question.²⁰ Intent plays a pivotal role in distinguishing between genuine errors, satire, and deliberate manipulation of information.²¹ Nonetheless, ascertaining intent in the digital realm can be challenging, given the anonymity, pseudonymity, and decentralized nature of many online platforms.²²

One of the key legal instruments – though it is not binding – concerning disinformation is the United Nation Human Rights Council’s (hereinafter referred to as UNHRC) report on disinformation and freedom of opinion and expression.²³ The report states that there is no universally accepted definition of disinformation,²⁴ despite the fact that a number of state laws and acts aimed to establish a comprehensive concept of disinformation.²⁵ As per the UNHRC’s report “disinformation is understood as false information that is disseminated intentionally to cause serious social harm” in contrast to misinformation which is “the dissemination of false information unknowingly”.²⁶ However, it is open to debate how far the definition proposed in the report applies to new media phenomena in an effective sense. For instance, according to the two short definitions, the sole distinctive feature is the definitive element of ‘harm’. The binary solution evokes a crucial conceptual problem; how can disinformation be categorized for situations where false information is disseminated with mixed motives or where intentionality is ambiguous? This polemic leads to the objective understanding of ‘intent’, which has been discussed in the Rabat Plan of Action as well.²⁷

3.2. Liability and platform regulation

It is important to mention another legal instrument issued by signatories in the framework of the Guidance of the European Commission, namely ‘The 2022 Code of Practice on Disinformation’ (hereinafter referred to as COPD2022).²⁸ The COPD2022 was signed and presented on 16 June 2022, by 34 participating entities, represents a comprehensive effort to counter online disinformation. Building upon the original Code published in 2018, which described disinformation as an information which is verifiably false or misleading, which, cumulatively, is created, presented and disseminated for economic gain or to intentionally

²⁰ Cf. APPELMAN et al. (2022).

²¹ AÎMEUR–AMRI–BRASSARD (2023): 28–30.

²² KENNEDY (2006)

²³ United Nations Human Rights, Office of the High Commissioner: A/77/288: Disinformation and freedom of opinion and expression during armed conflicts – Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 12 August 2022. <https://www.ohchr.org/en/documents/thematic-reports/a77288-disinformation-and-freedom-opinion-and-expression-during-armed>

²⁴ Ibid. 5.

²⁵ See: Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG).

²⁶ Ibid. 15. §.

²⁷ Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred – Annual report of the United Nations High Commissioner for Human Rights, 2012. 11. https://www.ohchr.org/sites/default/files/Rabat_draft_outcome.pdf

²⁸ 2022 Strengthened Code of Practice on Disinformation. <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>

deceive the public and may cause public harm intended as threats to democratic political and policy-making processes as well as public goods such as the protection of EU citizens' health, the environment or security.²⁹ This new iteration incorporates a broader spectrum of commitments and measures aimed at addressing the dissemination of false information.³⁰ The true novum of the COPD2022, however, lies in the profound examination of platform issues when fighting disinformation.

The objective of the COPD2022 aligns with the Commission's Guidance from May 2021,³¹ striving to combat online disinformation through enhanced transparency, user empowerment, and industry cooperation. This new version comprises 44 commitments and 128 specific measures in total across multiple domains. One significant facet pertains to the concept of demonetization, an ambitious commitment, aiming to curtail financial incentives for those disseminating disinformation. Demonetization in the context of the enforcement of the COPD2022 therefore means that signatories commit to preventing advertisements from being placed next to disinformation and to avoid distributing advertising containing false information. To achieve this rather grandiose obligation, greater collaboration within the advertising sector is fostered, especially in the online advertising segment on platforms. The COPD2022 also emphasizes the importance of transparency in political advertising, which is in line with current European regulatory perspectives, namely the European Media Freedom Act.³² With regard to political disinformation propaganda, signatories also undertake to implement robust transparency measures, enabling users to identify political ads with ease; these measures include improved labelling, disclosing sponsors, ad spending, and display duration and potentially establishing efficient and searchable ad libraries for political advertising.

To safeguard the integrity of online services, the COPD2022 strengthens measures against manipulative behaviors which propagate disinformation, such as fake accounts, bot-driven amplification, and deep fakes.³³ While practice will tell how these measures manifest into enforcement, nonetheless, it can be underlined that enhanced cooperation among signatories will be pivotal in addressing these challenges, with a cross-service understanding of manipulative techniques and practices.

The COPD2022 also underlines that the fight against disinformation will be fought and won by the users. Therefore, it aims to empower users with tools for recognizing, understanding, and reporting disinformation. In this regard, safe design practices will be implemented to minimize the spread of false information, and transparency in recommendation systems will be increased to mitigate disinformation propagation. Furthermore, the focal objective of the COPD2022 is to enable researchers and the factchecking community with more effective tools and options for fighting disinformation on platforms. As for such

²⁹ 2018 Code of Practice on Disinformation: 1. <https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation>

³⁰ BORZ et al. (2024)

³¹ EU Commission (2021)

³² Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing a common framework for media services in the internal market (European Media Freedom Act) and amending Directive 2010/13/EU COM/2022/457 final. Brussels, 16.9.2022. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0457>

³³ Cf. GOSZTONYI–LENDVAI (2023).

tools, online platforms are to facilitate research on disinformation by providing access to data, fostering a better understanding of disinformation trends and future perspectives. The reach of factchecking will be extended across all EU Member States and languages, ensuring more consistent usage.

In the framework of effective and productive measures, the COPD2022 introduces two fundamental measures: (1) the establishment of the Transparency Center and (2) the monitoring provisions on efficacy of the COPD2022. To bolster accountability, the Transparency Center, which will be accessible to the public, offers insights into the implementation of measures. A Permanent Task-force within the EU framework will also maintain the COPD2022's relevance, adapting commitments to evolving technological, societal, and legislative changes. The COPD2022's efficacy will be monitored through qualitative and quantitative reporting, with periodic assessments carried out by the European Regulators Group for Audiovisual Media Services (ERGA) and the European Digital Media Observatory (EDMO). Signatories will have six months to implement their commitments, and progress will be continually reviewed and adjusted as necessary.

Though a progressive step towards more transparency and democratic values, the COPD2022 is not free of regulatory challenges. A central hurdle lies in accurately understanding the intent behind the dissemination of information. The COPD2022 hinges on distinguishing between disinformation and other content based on the motive behind it, yet discerning this intent proves intricate due to the multifaceted nature of motivation and the difficulty in establishing clear intent. Enforcing the measures in cooperation with online platforms and users may also prove harder in practice than in theory. This is due to the fact that the efficacy of the COPD2022 heavily relies on signatories diligently adhering to their obligations and implementing the outlined measures, so without robust enforcement mechanisms, the impact of the Code could – and evidently will – be limited, as entities might fail to take substantial action against the spread of disinformation. Lastly, the technological prerequisites of the proposal also pose difficulties. The COPD2023 assumes that platforms have the technical capability to accurately identify and remove disinformation, yet the task of differentiating between disinformation and nuanced content requires advanced content analysis algorithms – or even technologies which are not even developed yet to a point where they can be effectively used.

As the COPD2023 itself mentions the Digital Services Act (hereinafter referred to as the DSA), it is essential to briefly highlight the key provisions in the EU's landmark regulation of platforms. The DSA – which aims to mitigate the harm of caused by disinformation on online platforms³⁴ – proposes “diligence requirements for providers of intermediary services as regards the way they should tackle illegal content, online disinformation or other societal risks”.³⁵ This objective is stipulated in a number of ways in the Regulation – from horizontal and content-neutral general obligations for providers to additional obligations for very large online platforms such as Meta or X (the Twitter).³⁶ The DSA provides a firmer le-

³⁴ LEISER (2023): 4.

³⁵ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [hereinafter referred to as DSA] (2) pr.

³⁶ LEISER (2023): 4.

gal basis behind the self-regulatory efforts (such as the COPD2023),³⁷ by imposing stricter and more transparent advertising rules and codes of conduct for online advertising. However, users will also be capable of reporting disinformation due to the new notice and action mechanism.³⁸ One particular provision is to be referenced to with regard to very large online platforms; Article 34 of the DSA stipulates that “providers of very large online platforms and of very large online search engines shall diligently identify, analyze and assess any systemic risks in the Union stemming from the design or functioning of their service and its related systems, including algorithmic systems, or from the use made of their services”. This means that platforms with over 45 million monthly active users³⁹ are to assess systemic risks – such as risks related to disinformation – deriving from the operation of such platforms. This particular provision is highly significant as mega companies such as Meta, Amazon and Google are to investigate their own platforms and assess potential risks⁴⁰ and harms that billions of users may face. This will be of very great relevance to academics and parties with an interest in user experience, considering the unparalleled parts these platforms played thus far in disseminating disinformation⁴¹ (COVID-19 disinformation advertisements on Facebook is a recent instance of the above⁴²).

3.3. Alternative solutions: ‘platform courts’

Amid the complex landscape of combating disinformation on a national and international regulatory level concerning safeguarding freedom of expression, the exploration of alternative solutions has emerged as a compelling avenue for addressing the challenges presented by the proliferation of false information on online platforms, especially in the field of social media platforms.⁴³ As elucidated above, the rapid and pervasive spread of disinformation poses critical societal and regulatory concerns, impacting diverse realms such as public discourse, political trust, and social cohesion.⁴⁴ The lack of globally harmonized regulation on disinformation leads to institutional novums in the forms of innovative approaches that transcend conventional regulatory paradigms. The conceptual proposition of ‘platform courts’ is emblematic of such innovation, reflecting a transformative endeavor to reconcile these seemingly competing imperatives.

In the present paper, the term ‘platform courts’ is proposed to describe and represent the novel paradigm wherein dedicated and specialized mechanisms are envisioned by privately owned companies, such as Meta, in order to mediate disputes arising from disinformation propagated on digital platforms. These quasi courts⁴⁵ are conceived as independent arbiters equipped with the authority by the establishing company to assess and adjudicate

³⁷ DSA (106) pr.

³⁸ DSA Article 16.

³⁹ DSA Article 33.

⁴⁰ HUSOVEC (2024)

⁴¹ MARSDEN et al. (2019); cf. DSA 84 pr.

⁴² BRUNS–HARRINGTON–HURCOMBE (2020)

⁴³ NIKLEWICZ (2017)

⁴⁴ COLOMINA – SÁNCHEZ MARGALEF – YOUNGS (2021)

⁴⁵ KLEIN (2018)

cases specific to the given platform, wherein the spread of disinformation can and may be contested and judged. The underlying rationale is to establish a dedicated forum for addressing issues surrounding misinformation and disinformation, thereby offering a quasi-judicial space where the countervailing interests of combating falsehoods and preserving freedom of expression can be balanced.⁴⁶ One example of a platform court – which also happens to be the one with the widest scholarly interest – is the Facebook Oversight Board (hereinafter referred to as the FOB), which became fully operational in 2021.⁴⁷ It is essential to briefly address the establishment and main goals of the FOB. Often referred to as an “independent and transparent quasi Supreme court”,⁴⁸ the Fob was established by Facebook (now Meta) to combat systemic content moderation issues prevalent on the platform.⁴⁹ Intended to be objective, professional and independent from Facebook,⁵⁰ the FOB – consisting of experts around the world⁵¹ – judges individual content moderation cases that are of utmost importance with respect to the community guidelines of Facebook.⁵² The latter means that the FOB is careful to ensure that the ‘hard’ or ‘complex’ cases are brought before the FOB; cases that have an impact in the offline world or are of high relevance to public discourse; in short, cases which are complex, significant, globally relevant and can be used to inform the development of future Facebook rules.⁵³

The distinctive features of the FOB in contrast to a conventional human rights court lie in three primary categories: (1) rapidity of decisions (members have 90 days to make a decision in the given case as per the guiding Bylaws⁵⁴ and the Charter), (2) high emphasis on trans-disciplinary expertise and (3) specificity of the scope of authority – meaning that FOB’s decisions are solely binding on Facebook and nowhere else.⁵⁵ Drawing upon expertise from legal, political, media, and technological domains, the FOB holds the promise of informed and impartial adjudication. This is particularly pertinent given the intricate interplay of legal nuances, content moderation intricacies, and technological complexities intrinsic to the contemporary disinformation landscape. The FOB intends to establish a Facebook case law, though which content moderators of the platform have a clearer and more transparent, judgementbased direction to understand and examine cases involving hate speech, xenophobia, promotion of illegal substances and dissemination of sexual content. As for human rights comparisons, the FOB often makes reference to relevant human rights law in its decisions and supports its decisions with legally binding international human rights provisions, which represents an exceedingly severe, constitutional commitment by its members to the enforcement of human rights in the contractual relationship

⁴⁶ LENDVAI (2023b)

⁴⁷ O’KANE (2022): 167–170.

⁴⁸ KLONICK (2021)

⁴⁹ CHANDER (2012)

⁵⁰ KLEIN (2018)

⁵¹ ARUN (2022)

⁵² Facebook Oversight Board Charter [hereinafter referred to as FOB Charter (2023)]. https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf

⁵³ Ibid.

⁵⁴ Facebook Oversight Board Bylaws. <https://www.oversightboard.com/wp-content/uploads/2024/03/Oversight-Board-Bylaws.pdf>

⁵⁵ FOB Charter (2023)

between the user and Facebook.⁵⁶ With regard to the scope of authority, the Board holds the power to overrule Facebook's moderation decision, concluding in the fact that Facebook is obliged to reinstate or make the content available again.⁵⁷

Even though the FOB is often criticized for its lack of transparency and diversity inclusivity,⁵⁸ as well as the incredibly low number of cases decided,⁵⁹ this exemplar proposed by Facebook provides a tangible illustration of the potential advantages and challenges inherent in employing alternative mechanisms for content moderation and dispute resolution within the digital realm.

In summary, the pursuit of alternative solutions, notably the conceptualization of 'platform courts', stands as a pivotal trajectory within the discourse on disinformation and freedom of expression, because the endeavor to counteract the detrimental impact of false information while upholding democratic values necessitates innovative approaches which transcend conventional regulatory frameworks.

3.4. The phenomenon of chilling effect

The phenomenon of chilling effect is closely intertwined with freedom of expression. The concept of chilling effect stems from the nuanced understanding and interpretation of self-censorship and dampening of speech, arising from the fact that individual(s) or entity(es) perceive potentially harmful or detrimental consequences (which can be legal, economic or societal) for expressing their speech.⁶⁰ To understand how chilling effect is related closely to the regulation of disinformation, it is crucial to first explore the significance of the phenomenon concerning its destructive nature on vibrant exchange of ideas – no matter what their nature is – which is fundamental to a democratic society and public discourse.

In the context of freedom of expression chilling effect serves as a conundrum,⁶¹ which can also be linked with disinformation. The symbiotic relationship between disinformation and chilling effect constitutes a complex and multifaceted juncture which lies at the crux of contemporary debates concerning freedom of expression and the proliferation of false information. The rhetorics at the intersection between disinformation and chilling effect are based on the need for an adequate, balancing regulation. As a key statement, in relation to disinformation, the chilling effect is usually not the judicial intent, but rather a deleterious consequence.⁶² This means that chilling effect manifests as a latent consequence of vigorous efforts to counteract disinformation. As regulatory measures and content moderation strategies are deployed to quell the spread of falsehoods, the line between legitimate and lawful expression and potentially harmful information becomes increasingly blurred leading to greater scrutiny in speech

⁵⁶ SCHULTZ (2021)

⁵⁷ FOB Charter (2023)

⁵⁸ WONG-FLORIDI (2022): 261–264.

⁵⁹ LENDVAI (2023b)

⁶⁰ PECH (2021): 2.

⁶¹ Id.

⁶² Cf. FAJDIGA-ZAGORC (2023).

regulation – which essentially results in the emergence of chilling effect.⁶³ The risk of inadvertently stifling lawful discourse is amplified, prompting self-censorship among content creators, journalists,⁶⁴ and citizens who fear punitive actions or reputational damage. This can lead to a narrowing of public discourse, depriving democratic societies of the diverse perspectives vital for robust deliberation.⁶⁵

Therefore, the pivotal question is the following: how can regulators and policy-makers create a legal framework for disinformation where both avoidance of the negative consequences of chilling effect and adequate regulation on the dissemination of disinformation can be achieved? The question is rooted in the ambiguous nature of the problem; while curbing disinformation is crucial to fostering a wellinformed citizenry, the chilling effect underscores the potential cost of aggressive measures which restrict the free exchange of ideas. Consequently, the search for effective solutions demands a delicate equilibrium that mitigates the adverse effects of both disinformation and self-censorship. To approach the question from a theoretical perspective, a comprehensive framework which acknowledges the nuances of each context while acknowledging the potential consequences of overreach may be sufficient, meaning that national and international regulatory bodies, as well as online platform providers (especially providers of very large online platforms and very large online search engines) must be straightforward, yet cautious in their regulatory approaches, employing transparent and context-sensitive content moderation mechanisms which prioritize accuracy without undermining open discourse.

4. Conclusion

In conclusion, it can be stated – in view of the above – that the relationship between disinformation and freedom of expression is one of the more complex and multifaceted challenges that define the information landscape in the digital age. The four pivotal issues highlighted – conceptual challenges, liability quandaries, alternative regulatory, and the chilling effect – offer an in-depth exploration of the multifaceted dynamics involved in the studying of this intriguing legal and ethical issue.

In the face of these challenges, a nuanced and balanced approach is imperative to address and assess disinformation. Striking the, rather delicate and compound, balance between curbing disinformation and preserving the vibrancy of open dialogue requires collaboration between regulators, technology companies, civil society, and users. The study's comprehensive exploration contributes to an informed understanding of the intricate dynamics at play and underscores the necessity for evolving strategies that uphold the tenets of democratic societies, while navigating the complexities of the digital age. Ultimately, addressing the confluence of disinformation and freedom of expression necessitates adaptive and ethically informed approaches which safeguard democratic values without stifling the diverse voices essential to a thriving public sphere.

⁶³ PECH (2021)

⁶⁴ MUNORIYARWA–CHIUMBU (2019)

⁶⁵ BAUMBACH (2018): 93–96.

References

- AÏMEUR–AMRI–BRASSARD (2023) = Esma AÏMEUR – Sabine AMRI – Gilles BRASSARD: Fake news, disinformation and misinformation in social media. A review. *Social Network Analysis and Mining*, Vol. 13. (2023) Art. no. 30. <https://doi.org/10.1007/s13278-023-01028-5>
- APPELMAN et al. (2022) = Naomi APPELMAN – Stephan DREYER – Pranav Manjesh BIDARE – Keno C. POTTHAST: Truth, intention and harm. Conceptual challenges for disinformation-targeted governance. *Internet Policy Review*, 16 May 2022. <https://policyreview.info/articles/news/truth-intention-and-harm-conceptual-challenges-disinformation-targeted-governance/1668>
- ARUN (2022) = Chinmayi ARUN: Facebook's faces. *Harvard Law Review Forum*, Vol. 135. (2022) 236–264. <https://harvardlawreview.org/forum/no-volume/facebooks-faces/>
- BALL (2018) = Philip BALL: 'News' spreads faster and more widely when it's false. A study of information sharing on Twitter shows people more likely to retweet falsehoods. *Nature Online*, 18 March 2018. <https://doi.org/10.1038/d41586-018-02934-x>
- BAUMBACH (2018) = Trine BAUMBACH: Chilling effect as a European Court of Human Rights' concept in media law cases. *Bergen Journal of Criminal Law and Criminal Justice*, Vol. 6., No. 1. (2018) 92–114. <https://doi.org/10.15845/bjclcj.v6i1.1555>
- BORZ et al. (2024) = Gabriela BORZ – Fabrizio DE FRANCESCO – Thomas L. MONTGOMERIE – Michael Peter BELLIS: The EU soft regulation of digital campaigning. Regulatory effectiveness through platform compliance to the code of practice on disinformation. *Policy Studies*, Vol. 45., No. 5. (2024) 709–729. <https://doi.org/10.1080/01442872.2024.2302448>
- BRUNS–HARRINGTON–HURCOMBE (2020) = Axel BRUNS – Stephen HARRINGTON – Edward HURCOMBE: 'Corona? 5G? or both?' The dynamics of COVID-19/5G conspiracy theories on Facebook. *Media International Australia*, Vol. 177., No. 1. (2020) 12–29. <https://doi.org/10.1177/1329878X20946113>
- CASSESE (2022) = Sabino CASSESE: *Freedom of Expression as a Common Constitutional Tradition in Europe. Report of the European Law Institute*. Vienna, European Law Institute Secretariat, 2022. 1–44. https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Report_on_Freedom_of_Expression.pdf
- CHANDER (2012) = Anupam CHANDER: Facebookistan. *North Carolina Law Review*, Vol. 90., No. 5. (2012) 1807–1844.
- COLOMINA – SÁNCHEZ MARGALEF – YOUNGS (2021) = Carme COLOMINA – Héctor SÁNCHEZ MARGALEF – Richard YOUNGS: The impact of disinformation on democratic processes and human rights in the world. *European Parliament Working Paper*, 2021/April. <https://doi.org/10.2861/677679>
- FAJDIGA–ZAGORC (2023) = Mohor FAJDIGA – Saša ZAGORC: Freedom or fear of expression of judges? Exploring the 'Chilling Effect' on judicial speech. *European Constitutional Law Review*, Vol. 19., No. 2. (2023) 249–270. <https://doi.org/10.1017/S1574019623000093>

- FALLIS (2015) = Don FALLIS: What is Disinformation? *Library Trends*, Vol. 63., No. 3. (2015) 401–426. <https://doi.org/10.1353/lib.2015.0014>
- GOSZTONYI–LENDVAI (2023) = Gergely GOSZTONYI – Gergely Ferenc LENDVAI: Deepfake: A multifaceted dilemma in ethics and law. *Journal of Information Ethics*, Vol. 32., No. 2. (2023) 109–121.
- GUNATILLEKE (2021) = Gehan GUNATILLEKE: Justifying limitations on the freedom of expression. *Human Rights Review*, Vol. 22. (2021) 91–108. <https://doi.org/10.1007/s12142-020-00608-8>
- HAMELEERS (2023) = Michael HAMELEERS: Disinformation as a context-bound phenomenon. Toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory*, Vol. 33., No. 1. (2023) 1–10. <https://doi.org/10.1093/ct/qtac021>
- HUSOVEC (2024) = Martin HUSOVEC: The Digital Service Act's red line. What the Commission can and cannot do about disinformation. *Journal of Media Law*, Vol. 16., No. 1. (2024) 47–56. <https://doi.org/10.1080/17577632.2024.2362483>
- JACOBS (2022) = Leslie Gielow JACOBS: Freedom of speech and regulation of fake news. *The American Journal of Comparative Law*, Vol. 70., Suppl. 1. (2022) 278–311. <https://doi.org/10.1093/ajcl/avac010>
- KAPANTAI et al. (2020) = Eleni KAPANTAI – Androniki CHRISTOPOULOU – Christos BERBERIDIS – Vassilios PERISTERAS: A systematic literature review on disinformation. Toward a unified taxonomical framework. *New Media & Society*, Vol. 23., No. 3. (2020) 1301–1326. <https://doi.org/10.1177/1461444820959296>
- KENNEDY (2006) = Helen KENNEDY: Beyond anonymity, or future directions for internet identity research. *New Media Society*, Vol. 8., No. 6. (2006) 859–876. <https://doi.org/10.1177/1461444806069641>
- KLEIN (2018) = Ezra KLEIN: Mark Zuckerberg on Facebook's hardest year, and what comes next. (Apr 2, 2018) *Vox*, <https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-in-terview-fake-news-bots-cambridge>.
- KLONICK (2021) = Kate KLONICK: Inside the making of Facebook's supreme court. (February 12, 2021) *The New Yorker*, <https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court>.
- KOLTAY (2009) = KOLTAY András: *A szólásszabadság alapvonalai: magyar, angol, amerikai és európai összehasonlításban*. Budapest, Századvég, 2009.
- LEISER (2023) = Mark LEISER: Analysing the European Union's Digital Services Act provisions for the curtailment of fake news, disinformation, & Online Manipulation. *SocArXiv*, 2023. 1–13. <https://doi.org/10.31235/osf.io/rkxh4>
- LENDVAI (2023a) = Gergely Ferenc LENDVAI: Media in war. An overview of the European restrictions on Russian media. *European Papers*, Vol. 8. No. 3. (2023) 1235–1245.
- LENDVAI (2023b) = LENDVAI Gergely Ferenc: A Facebook "elitbírósa" kritikai megközelítésben. *Iustum Aequum Salutare*, 2023/4. 227–243.

- MARSDEN et al. (2019) = Cristopher MARSDEN – Trisha MEYER – Ian BROWN: Platform values and democratic elections: How can the law regulate digital disinformation? *Computer Law & Security Review*, Vol. 36. (2020) 1–18.
<https://doi.org/10.1016/j.clsr.2019.105373>
- MAZUR–GRAMBLIČKOVÁ (2023) = Jan MAZUR – Barbora GRAMBLIČKOVÁ: New regulatory force of cyberspace. The case of meta's oversight board. *Masaryk University Journal of Law and Technology*, Vol. 17., No. 1. (2023) 3–32.
<https://doi.org/10.5817/MUJLT2023-1-1>
- MENDEL (2010) = Toby MENDEL: *Restricting freedom of expression. Standards and principles*. Centre for Law and Democracy, Background Paper for Meetings Hosted by the UN Special Rapporteur on Freedom of Opinion and Expression, 2010. 1–23.
- MUNORIYARWA–CHIUMBU (2019) = Allen MUNORIYARWA – Sarah H. CHIUMBU: Big Brother is Watching: Surveillance Regulation and its Effects on Journalistic Practices in Zimbabwe. *African Journalism Studies*, Vol. 40., No. 3. (2019) 26–41.
<https://doi.org/10.1080/23743670.2020.1729831>
- NIKLEWICZ (2017) = Konrad NIKLEWICZ: Weeding Out out Fake fake Newsnews.: An Approach approach to Social social Media media Regulationregulation. *European View*, Vol. 2017/16., No. (2). (2017). <https://doi.org/10.1007/s12290-017-0468-0>
- O'KANE (2022) = Ruby O'KANE: Meta's private speech governance and the role of the oversight board. Les-sons from the board's first decisions. *Stanford Technology Law Review*, Vol. 25., No. 2. (2022) 167–209. <https://law.stanford.edu/publications/metasp-private-speech-governance-and-the-role-of-the-oversight-board-lessons-from-the-boards-first-decisions/>
- PECH (2021) = Laurent PECH: *The concept of chilling effect*. Open Society Foundations online publishing, 2021.
- SCHULTZ (2021) = Mårten SCHULTZ: Six problems with Facebook's Oversight Board. Not enough contract law, too much human rights. In: Judit BAYER et al. (eds.): *Perspectives on Platform Regulation: Concepts and Models of Social Media Governance Across the Globe*. Baden-Baden, Nomos, 2021. 145–163. <https://doi.org/10.5771/9783748929789-145>
- SHU et al. (2020) = Kai SHU – Suhang WANG – Dongwon LEE – Huan LIU: Mining disinformation and fake news. Concepts, methods, and recent advancement. *Lecture Notes in Social Networks*. Cham, Springer, 2020. 1–23.
https://doi.org/10.1007/978-3-030-42699-6_1
- WILDING et al. (2018) = Derek WILDING – Peter FRAY – Sacha MOLITORISZ – Elaine McKEWON: *The impact of digital platforms on news and journalistic content*. Sydney, Centre for Media Transition, 2018.
- WONG–FLORIDI (2022) = David WONG – Luciano FLORIDI: Meta's Oversight Board. A Review and Critical Assessment. *Minds & Machines*, Vol. 33. (2022) 261–284.
<https://doi.org/10.1007/s11023-022-09613-x>

The relevance of psychological operations in social media – Conceptual overview

Part 1

PÉTER BÁNYÁSZ

1. Introduction

The emergence of social media has transformed how people communicate and interact with each other. Billions of individuals worldwide use numerous platforms to connect with friends, family, and even strangers. The impact of these platforms on human behavior and decision-making has grown in importance, as they have become crucial tools for shaping public opinion and behavior.

The widespread dissemination of fake news on social media has become a significant concern in recent years, due to its potential impact on individuals and society. With social media becoming a primary news source for many people, the spread of misinformation has become increasingly rapid and widespread. The prevalence of fake news on social media has resulted in serious consequences, such as the propagation of conspiracy theories, the distortion of public opinion, and the erosion of trust in conventional media sources and democratic institutions.

As cybersecurity experts, we constantly search for ways to combat the spread of fake news. While technological advancements have provided numerous solutions to various problems, no infallible algorithm currently exists for detecting fake news. However, some tools can assist in detecting and preventing the spread of fake news to a certain extent.

One of the most powerful tools currently in use is an algorithm – that is, a type of computer program – designed to scan and analyze different types of online content, such as news articles, social media posts, or comments. Its job is to look for anything unusual or suspicious, like sudden changes in a text or strange word choices that don't seem natural.

These algorithms are trained to recognize certain patterns that are commonly found in misleading or false information. For example, if an article uses emotionally charged language or repeats certain keywords that often appear in fake news, the algorithm can flag it as potentially deceptive, so it can be checked more closely by human experts.

In addition to analyzing the content itself, there are also algorithms that focus on *how* information spreads. They examine the sharing behavior – for instance, if a piece of content is being shared very quickly and always by the same type of user accounts. If many of those accounts turn out to be fake profiles – often part of a so-called 'botnet', which is a group of automated accounts controlled by software – the system can detect this as suspicious activity and raise an alert. This helps prevent fake or harmful content from spreading too widely before it can be verified.

Despite the benefits of these tools, they are far from infallible, and they cannot guarantee the complete elimination of fake news. There are numerous ways that fake news can

be spread, and it can often be difficult to differentiate between genuine and fake news articles. Therefore, it is essential to remain vigilant and use these tools alongside other methods, in order to prevent the spread of fake news.

The technology designed to detect and prevent the spread of fake news itself faces a major hurdle due to the complexity of defining what exactly constitutes fake news. This issue becomes particularly evident when we ask readers to define what they mean by fake news. The definition of fake news is a subjective matter, much like the well-known humorous quote from Dylan Thomas about alcoholism: “An alcoholic is someone you don’t like, who drinks as much as you do”. Similarly, people will sometimes call legitimate information fake news, simply because it is not something they want to hear.

The difficulty of defining fake news also makes it challenging to develop a universal algorithm for detecting it. While various algorithms exist for scanning content and flagging suspicious activity, differentiating between genuine and fake news articles remains an unsolved problem. Fake news is often distributed through a variety of channels, making it even more difficult to track and prevent its spread. Therefore, combating the spread of fake news requires a multifaceted approach, including leveraging specialized technology and educating readers on identifying and avoiding fake news. By working together and remaining vigilant, we can minimize the impact of fake news and promote the dissemination of accurate information.

Understanding the spread of fake news on social media is crucial to developing effective strategies to combat it and promoting the dissemination of accurate information. By shedding light on the extent and factors contributing to the spread of fake news, this study aims to provide policymakers, journalists, and social media platforms with valuable insights on how to best address this issue.

2. Literature review

2.1. Social media

Social media has enabled people from diverse backgrounds to connect and form relationships, resulting in a robust social fabric. This has led to an abundance of information, providing users with a great deal of value. However, the downside of this information overload is it can sometimes become challenging for users to find credible and relevant information when needed. Social media has become an integral part of our daily lives: providing news updates, connecting with loved ones, seeking product reviews and recommendations, managing work, and staying up-to-date with the latest trends, among many others uses.¹

Many efforts have been made to explain social media, a majority of which have primarily engaged with the topic through marketing-related concepts. The rationale behind this is that in the early years of the 21st century, online marketing experts predominantly utilized social media platforms for professional purposes.

Oxford Dictionary describes social media as a set of websites and applications where users can create and share content on social networks.² Linked to this definition are those

¹ KAPOOR et al. (2018)

² Oxford English Dictionary: Social media. In: *Oxford English Dictionary*. Retrieved from <https://www.oed.com/search/dictionary/?scope=Entries&q=social%20media>.

of Andreas Kaplan and Michael Haenlein, who describe social media as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content”.³

This indicates that social media is primarily the result of user interaction, which may involve the partial or complete creation of new content by sharing and supplementing content by other users. Theoretically, this content can undergo constant evolution and expansion as it is updated with new information. While there appears to be little difference between the definition by Kaplan and Haenlein and the one proposed here, using the latter definition can significantly broaden the scope of what is meant by social media. On this basis, social media can also encompass those smartphone applications which facilitate interaction between users and serve as an integrative tool between various social media platforms. A notable example of such integration is Google, which initially operated as a search engine provider but now offers a wide range of social media tools (e.g., blogging, photo and video sharing, social networking, smartphone platforms, etc.) within a single entity.

In light of the definition proposed, social media can be characterized as platforms that fulfill the criteria of content production and sharing. Various theories have been developed to classify social media sites based on user types and applications. Ngai et al., for instance, present a conceptual framework positing three levels of interdependence. The authors contend that adopting their model can ultimately promote an inter and multidisciplinary approach for research into social media.

The first level refers to theories and models related to social media. The authors identify the following:

1. behavioral theories (e.g., personality traits,⁴ TAM⁵);
2. theories of social learning (for example, social capital,⁶ social identity⁷);
3. theories of mass communication⁸ (for example, parasocial relations⁹).

³ KAPLAN–HAENLEIN (2010)

⁴ The purpose of categorising each personality trait is to differentiate between individuals, thereby understanding the motivations behind their behavior and their interpersonal interactions. See also: PLÉH–BOROSS (2010).

⁵ The Technology Acceptance Model postulates that the perceived usefulness and ease of use of a technology are key determinants of its adoption by users. This model was originally formulated by Davis, but has since undergone further refinement. The TAM2 model incorporates a number of additional factors that can play a role in determining the utility of a technology. See also: VENKATESH (2000).

⁶ Social capital is a term that finds its origin in the world of economics, where it is closely related to the concept of capital. However, unlike financial capital, social capital is a resource rooted in the quality, quantity, and structure of interpersonal relationships. It is the value that individuals derive from their social networks, and it has the potential to positively impact people’s well-being. By fostering collective action, social capital can contribute to the prosperity of society as a whole. In essence, social capital is the glue that binds people together, enabling them to achieve more than they could on their own. See also: PINXTEN–LIEVENS (2014).

⁷ The concept of social identity encompasses an individual’s desire to cultivate a positive self-image by aligning themselves with a specific group. This inclination is based on the idea that belonging to a particular group is preferable to being associated with other groups. It revolves around a constructive sense of affiliation with one group over others. See also: TAJFEL (1974).

⁸ BLUMLER–KATZ (1974)

⁹ According to Veronika Munk (MUNK [2009]), parasocial relationships are a type of connection promote a sense of belonging, fosters group participation, and generates a feeling of emotional closeness to particular individuals, like celebrities. However, these relationships can supplant genuine, intimate human connections, resulting in adverse outcomes.

The second level includes the platforms. The authors listed the following:

1. content sharing sites (e.g., YouTube, TikTok, Picasa);
2. social bookmarking sites (e.g., Pinterest);
3. blogs, microblogs (e.g., Blogger.com, X));
4. virtual/online communities (e.g., Lonely Planet, Yahoo Answers);
5. social network sites (e.g., Facebook and LinkedIn);
6. virtual worlds (e.g., Second Life).

The third level covers the individual application areas that result from the integration of the previous levels. These areas include:

1. marketing;
2. knowledge sharing;
3. customer relationship management;
4. cooperation activities;
5. organizational communication;
6. education and training;
7. and others.

A number of studies have examined the academic literature on social media using the same network theory approach outlined above, e.g., Grover et al.¹⁰ or the previously cited Kapoor et al.

Social media networks have been frequently accused of inadequately addressing the issue of fake news. This criticism stems from their inability to effectively counteract the spread of misleading information, suggesting that they should implement precise algorithms to detect and block such content. However, effective algorithms capable of identifying fake news would have already been developed if it were possible for clear-cut parameters to be used to differentiate between fake news and opinions which are merely biased. This challenge seems all the more insurmountable during election campaigns, as the threshold for calling certain opinions 'biased' varies widely based on the tolerance levels in individual countries. Consequently, social media platforms must continue investing in research into and the development of algorithms to effectively combat the dissemination of fake news and biased opinions.

The scope of this paper is not aimed at presenting a thorough discourse on the hazards of social networking sites; instead, it seeks to investigate their role in shaping policy decisions. However, in some instances, this role is in fact interrelated with the risks associated with social networking sites, as the attributes derived from using these platforms may vary depending on the users' motives. A pertinent example is the data collection practices employed by social networking sites such as Facebook and Google, which collect various types of user data. As evidenced by the disclosures made by Edward Snowden, this kind of data gathering facilitates extensive state surveillance.¹¹ Moreover, these social networking sites also sell this data to third-party entities for targeted advertising purposes. Often, this data is not only relevant to consumption habits, but can also be used as a significant factor in influencing political attitudes. In order to sway users on political matters, market-oriented companies sometimes utilize the data they collect and/or purchase from social networking sites.

¹⁰ GROVER-KAR-DWIVEDI (2022)

¹¹ GREENWALD (2014)

Implementing big data technology for collecting, processing, and analyzing the vast amounts of data available within the new media system has brought about significant transformations in global politics and economics, particularly with regard to the business models of internet-based social networks.¹² Notably, recent advancements in social network analytics have facilitated microtargeting, a technique that enables the transmission of tailored messages promoting specific content or services to receptive audiences. Regrettably, this capability has been exploited in a number of unethical ways, with the most infamous example being the scandal involving Facebook and Cambridge Analytica, a British consulting and analytical company. Established in 2012, Cambridge Analytica promoted its services as a consultancy specializing in ‘behavioral communications’ campaigns based on the research of Aleksandr Kogan, a scholar affiliated with the University of Cambridge.¹³ Kogan developed a personality test application on Facebook called “This is your digital life.” At the time, Facebook’s privacy practices permitted the creation of such applications, but specified that only certain parties would be permitted to access the collected data. The application duly alerted users that their data would be collected, and financially incentivized them to complete the test. Ultimately, approximately 270,000 individuals completed the personality test. However, the application not only gathered data from the individuals who took the test, but also accessed additional data from their social connections. It is important to note that a wide range of other applications also took advantage of these unethical methodologies.¹⁴

A company obtaining user data does not, in itself, imply that it will use it in unlawful ways, nor does it guarantee its ability to analyze it effectively. A case in point is the 2011 “What is your Indian name?” test on Facebook, which garnered responses from almost 800,000 Hungarian users in just a few days. Subsequently, it was revealed that the app was sponsored by an online jewelry store that had used the test as a guerrilla marketing campaign to increase the number of followers on its Facebook page. Notably, Facebook’s advertising system was already equipped to sell targeted ads at the time, which the sponsor had utilized to select specific demographic groups (e.g. young people aged between 19 and 25 years, located in Budapest, enrolled in higher education, and following certain pages, among other criteria) for its advertisements. However, the real issue arises when data can effectively categorize user personalities and predict their future behavior patterns. The ‘Big Five’ personality model, which uses factor analysis to classify individuals into five different factor groups – Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism – based on specific characteristics¹⁵ served as the basis for Aleksandr Kogan’s personality test. Numerous studies have investigated the correlation between an individual’s Facebook activity and their personality, revealing that it is feasible to establish a link between Facebook likes and personality traits using the ‘Big Five’ personality model.¹⁶

By constructing an individual profile using data analytics, marketers can utilize targeting to deliver an advertisement which is more likely to generate clicks and purchases. This technique has proven particularly effective in election campaigns, with certain algorithms capable of identifying a user’s potential party affiliation and displaying adverts tailored to their personality. For instance, if a user is identified as having a neurotic personality

¹² LEVAK (2021)

¹³ HINDS–WILLIAMS–JOINSON (2020)

¹⁴ SYMEONIDIS et al. (2018)

¹⁵ NORMAN (1963)

¹⁶ HE et al. (2014); EŞKİSU–HOŞOĞLU–RASMUSSEN (2017); AZUCAR–MARENGO–SETTANNI (2018).

type, the algorithm will display content associated with violent acts, while a family-oriented user will be presented with ads that emphasize family traditions and values.

In order to build accurate user profiles, it is crucial to have a substantial amount of data. Social networking sites such as Facebook, Google, and Cambridge Analytica collect massive amounts of user data based on various factors. This data is not only collected by the companies themselves, but also purchased from third-party data brokers, and it can include everything from user demographics to their interests, online behavior, and even location data.

User data is not only sourced from data brokers. Facebook also collects data through technical means, including scripts such as the 'like' button, which is used for instant sharing. Additionally, Facebook collects data on people who upload pictures, including individuals not registered on the site. This data is then used to create profiles of these individuals, even if they have never used Facebook. One of the most intriguing aspects of Facebook's data collection is the existence of these 'shadow profiles': profiles on non-members, i.e. individuals who have never signed up for a Facebook account. In addition to uploading photos, synchronizing phone books or email accounts can also result in Facebook accessing this data and using it to create shadow profiles. Facebook has long denied the existence of these shadow profiles, but in the aftermath of the Cambridge Analytica scandal, CEO Mark Zuckerberg admitted their existence during a congressional hearing in April 2018.¹⁷ Zuckerberg claimed that these profiles were necessary for enable future account registrations, which was why the site had them.

Technical data collection is another way that companies such as Facebook can support profiling. This involves monitoring device hardware and software characteristics, such as the operating system, browser, mouse location, and battery charge. Companies legally collect this information, using it to improve user experience and to target advertising more effectively. The data used by Kogan was also lawfully obtained; their only unlawful conduct was selling it to Cambridge Analytica. It is important to note that the situation is different with applications used for phishing, which intentionally deceive users into providing personal information.

The correlation between the perceived relevance of news and its dissemination can be observed through the extent of information sharing. In February 2011, a journal exposed a US Air Force contract for developing "online identity management software" designed to create a botnet capable of influencing political decision-making by manipulating social media profiles.¹⁸ The software was required to meet specific criteria, such as bypassing geo-location restrictions to avoid compromising operations in the Middle East, instead using fabricated profiles that indicated a US location. These fabricated profiles were also required to have personalized backstories tailored to the target area, and involved the use of a virtual private network. It has been suggested that several states have developed botnets similar to this software in recent years, with X being a popular platform for such operations.¹⁹

¹⁷ Russel BRANDOM: Shadow profiles are the biggest flaw in Facebook's privacy defense. *The Verge*, 11 April 2018. <https://www.theverge.com/2018/4/11/17225482/facebook-shadow-profiles-zuckerberg-congress-data-privacy>

¹⁸ Stephen C. WEBSTER: Social media. Air Force ordered software to manage army of Fake Virtual People. *Global Research*, 30 May 2011. <https://www.globalresearch.ca/social-media-air-force-ordered-software-to-manage-army-of-fake-virtual-people/25060>

¹⁹ ABOKHODAIR-YOO-McDONALD (2015)

The significance of political movements organized through social media platforms is exemplified by the events that transpired during the ‘Arab Spring’ in late 2010.²⁰ The movement was initiated in Tunisia, where young people – predominantly from urban areas – organized peaceful protests on social networking sites. The movement quickly spread to other countries in the region, some governments responding with violence. However, the dissemination of images and videos of these protests on social media helped the movement gain international attention and support. The potential for the use of software tools similar to the aforementioned online identity management software during such events is significant. These tools could amplify existing discontent by creating a network of botnets consisting of fake profiles as well as actual individuals in the field. This would increase the sharing of related content, making it newsworthy enough to achieve international reach, possibly with some help from mainstream media.

A case in point is the 2014 incident when Russia accused Hungary of violating an EU arms export ban by selling T-72 tanks to Ukraine. The Russian Foreign Ministry relied on a report published on an obscure website called Hídfő.net which was linked to the far-right group Magyar Nemeth Arcvonal. The news quickly spread across international mainstream media, but it was later discovered that Hídfő.net was controlled by Russian national security services, and the news article was likely written by people connected to them.²¹ Disseminating fake news is one of the most effective methods of influencing political decision-making, and social media is a significant facilitator.

Without pluralistic consumption habits, the automated selection processes of social networking site algorithms, adapting to the user’s behavior, may lead to a ‘filter bubble’²², wherein users only encounter content that aligns with their existing beliefs and opinions and are rarely confronted with opposing viewpoints. This can create an impression that their narrow perspective represents reality as it is. This is closely related to the post-truth phenomenon,²³ whereby public opinion is driven by emotions and personal convictions rather than facts. As a result, objective facts lose their importance in the perception of reality, and are replaced by many parallel subjective realities. This contributes to the absorption of fake news and the confusion generated by disseminating alternative points of view questioning the validity of mainstream news releases; this is known as ‘noise-making’. This activity is widely employed by national security agencies of illiberal, authoritarian states, particularly against the European Union, to fragment it and impede its ability to act as a unified global political actor, leaving more room for the political ambitions of noise-making states.

2.2. Fake news

Fake news has a long history,²⁴ but its most recent and most concerning form is politically motivated fake news, which has raised several unanswered scientific questions.

²⁰ SPIER (2017); KHONDKER (2015).

²¹ For further analysis on the relationship between European far-right parties and Russia, see Political Capital’s analysis: JUHÁSZ et al. (2015).

²² SPOHR (2017)

²³ LEWANDOWSKY–ECKER–COOK (2017)

²⁴ In ancient Rome during the time of Julius Caesar, spreading fake news was a common tactic used to undermine political rivals.

According to Tandoc et al., the digitization of news has presented a significant challenge to the conventional understanding of news.²⁵ The emergence of online platforms has allowed non-journalists to reach a mass audience, resulting in the rise of citizen journalism. This has called into question the previously established link between news and journalists, as non-journalists are now able to actively participate in journalistic activities to produce news content. Initially confined to blogging, citizen journalists have been afforded a more prominent platform through social media, allowing them to post first-hand information, photos, videos, and narratives about newsworthy events. Journalists have also adapted to this trend and increased their presence on social media, using it to break stories and interact with their audiences. X, in particular, has become a valuable tool for journalists to disseminate details about breaking news events. Moreover, social media has challenged the traditional conventions of news presentation, as tweets – despite often being limited to 140 characters – are now considered news, especially if they originate from someone in authority.

The distribution of news content is a complex process, significantly facilitated by social media. One critical challenge posed by social media is its tendency to blur the traditional delineations between types of information sources. For instance, news articles can be published by a news organization and then reach an individual through a dedicated news site, the news organization's Facebook site, or a shared posting on social networks. This means that social media users have to navigate a complex landscape of information shared by multiple sources, which can be perceived as "a set of layers with various levels of proximity to the reader". While receiving information from socially proximate sources can contribute to legitimizing the accuracy of information shared on social networks, users often fail to verify the information that they share.

Additionally, social media amplifies the bandwagon heuristic, as each post comes with its own popularity ratings. Posts receiving many likes, shares, or comments are more likely to attract attention from others, and are therefore even more likely to be further liked, shared, or commented on. Popularity on social media thus becomes positive feedback loop facilitating the spread of unverified information. More recently, the development of so-called news bots has further automated this feedback loop, adding what the unsuspecting news reader might interpret as a stamp of legitimacy to the information. These challenges have significant implications for distributing and consuming news content in the digital age, and warrant further scholarly attention.

These developments have had a significant impact on the distribution and consumption of news and the broader understanding of what actually constitutes news. The definition of news is extremely broad, encompassing not only accounts of recent and significant events, but also portrayals of anything unique or deviant that affects people and captures their attention. Journalism is often regarded as the source of news, and is expected to provide "independent, reliable, accurate, and comprehensive information". Since the primary objective of journalism is to empower citizens with the information they need to participate meaningfully in a democratic society, journalists must prioritize truth above all else. Conforming to specific standards such as accuracy and objectivity is thus an essential part of the professional definition of journalism.

²⁵ TANDOC-LIM-LING (2018)

This responsibility brings a certain degree of power, and journalists occupy an influential position in society that can lend credibility to their reports.

Nevertheless, news is a social construct, and journalists often exercise subjective judgment when selecting which pieces of information to include or exclude from their reports. As a result, news is susceptible to external forces, including governments, audiences, advertisers, and the preferences of the journalists themselves. News is also unique in that it is sold to audiences whose data is then sold to advertisers in turn, making it subject and vulnerable to market forces. Despite these challenges, news is widely expected to contain factual and authentic information. For instance, a landmark survey of American journalists distinguished between those who see their work as reporting on factual reality and those who focus more on creating symbolic or interpretive media content.²⁶ Journalists are often described as ‘making the news’, but this means they are selecting, shaping, and presenting events — not inventing or fabricating them.

The definition of ‘fake news’ and how it differs from ‘legitimate news’ continues to be a topic of much discussion. Generally, news is expected to be truthful and accurate, so the very term ‘fake news’ appears to be an oxymoron. The word ‘fake’ is often used to describe something inauthentic: a copy, a forgery, or a counterfeit. A research study on identifying fake websites identified two types: spoof sites mimicking existing websites, and concocted sites attempting to appear as legitimate commercial entities.²⁷ Another study on fake online reviews emphasized the importance of intention in determining what is considered fake. The study classified fake reviews as deceptive reviews intended to mislead consumers in their purchasing decisions, often written by reviewers having little to no experience with the products or services being reviewed.²⁸

Wardle proposed the most basic classification system during an election season (see *Figure 1*):²⁹

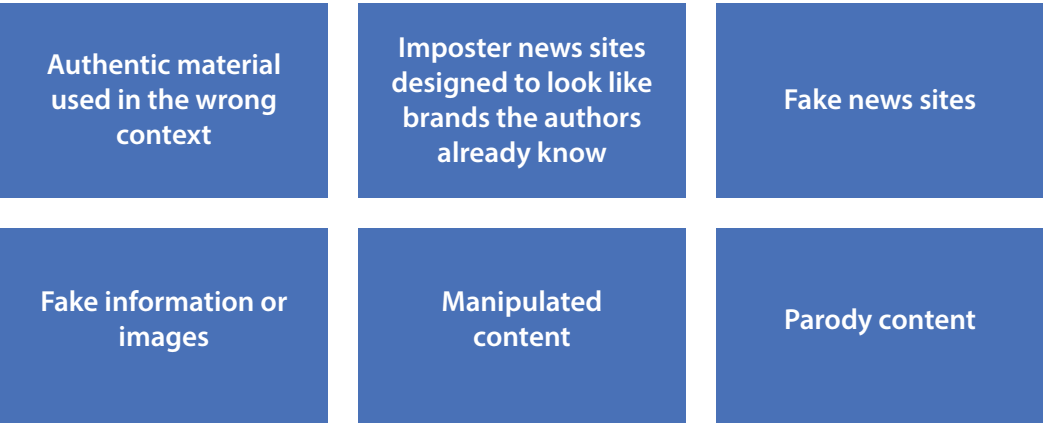


Figure 1. Classification of fake news (adapted from WARDLE³⁰)

²⁶ GRAVES (2016)

²⁷ ZHANG– HONG–CRANOR (2007)

²⁸ HARRISON-WALKER–JIANG (2023)

²⁹ Claire WARDLE: 6 types of misinformation circulated this election season. *Columbia Journalism Review*, 18 November 2016. https://www.cjr.org/tow_center/6_types_election_fake_news.php

³⁰ Ibid.

Another approach is to classify the various types of false and misleading news sources (see *Figure 2*)³¹:

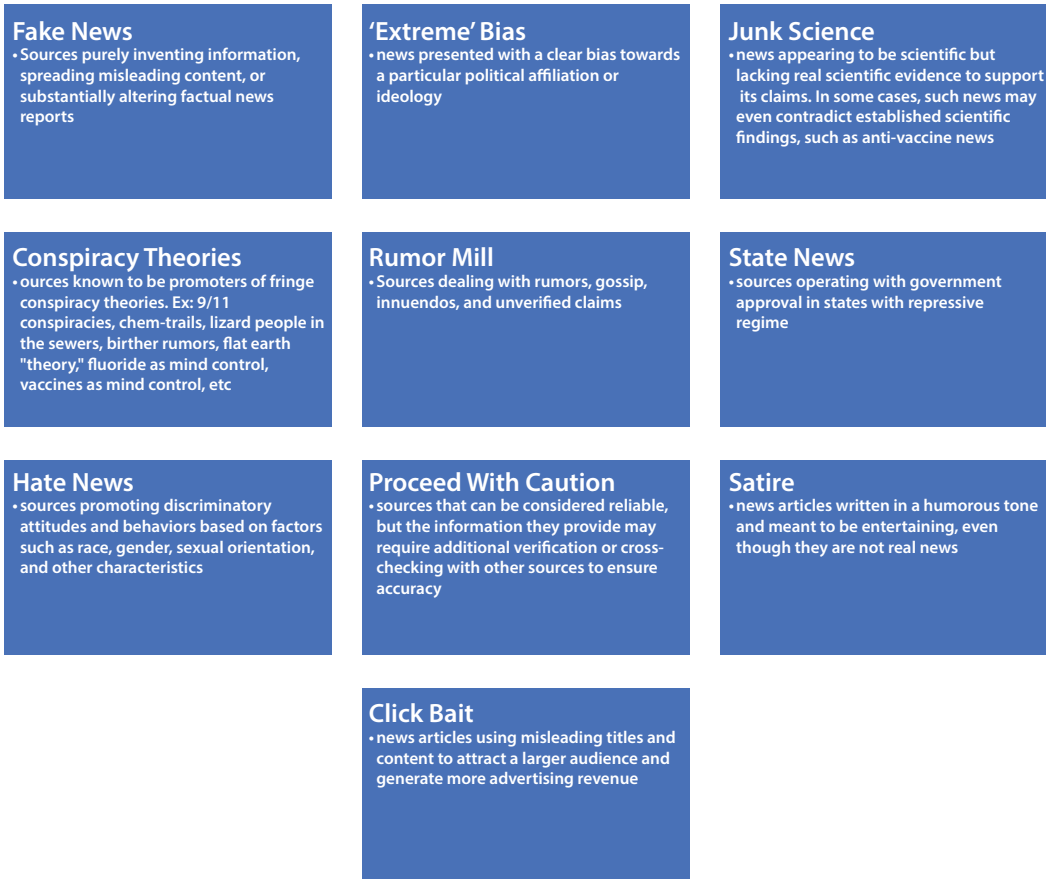


Figure 2. *Types of false or misleading news sources (adapted from Zimdars³²)*

Yet another typology distinguishes the following categories, as shown in *Figure 3*.

³¹ Melissa ZIMDARS: *False, misleading, clickbait-y, and satirical "News" sources*. 2016. https://docs.google.com/document/d/10eA5-mCZLSS4MQY5QG5ewC3VAL6pLkT53V_81ZyitM/preview?tab=t.0

³² Ibid.



Figure 3. *Types of fake news (adapted from TANDOC–LIM–LING [2018])*

According to Lazer et al., fake news is “fabricated information that mimics news media content in form but not in organizational process or intent. Fake-news outlets, in turn, lack the news media’s editorial norms and processes for ensuring the accuracy and credibility of information. Fake news overlaps with other information disorders, such as misinformation (false or misleading information) and disinformation (false information that is purposely spread to deceive people).”³³

Recent empirical evidence challenges the prevailing assumption that political partisanship and motivated reasoning are to blame for the spread of fake news.³⁴ Instead, a lack of careful reasoning, relevant knowledge, familiarity, and source heuristics underlies poor truth discernment. Furthermore, a significant discrepancy exists between individuals’ beliefs and their social media sharing behaviors, mainly attributable to inattention rather than the intentional dissemination of misinformation. However, effective interventions can be developed for encouraging social media users to prioritize accuracy, such as nudging them to engage in critical thinking and leveraging crowdsourced veracity ratings to enhance social media ranking algorithms.

The pervasive influence of fake news on social media – including both personal opinions and scientific consensus on topics such as COVID-19, climate change, and evolution – has become a significant concern in contemporary society³⁵. Its effects cut across various domains, including politics, journalism, and digital media. According to Aggarwal and colleagues,³⁶ fake news – political, scientific, or belief-oriented – significantly affects individuals, especially those who are politically inclined or looking to spread their ideas widely. The spread of fake news on social media can influence how governments, organizations, and individuals respond to events and issues, leading to polarized beliefs and values. This phenomenon has raised serious questions about the credibility, quality, and verification of information on social media platforms. Despite the growing body of research on fake news and social media, a comprehensive and in-depth review of its impact on society has yet to be conducted.

³³ LAZER et al. (2018)

³⁴ PENNYCOOK–RAND (2021)

³⁵ OLAN et al. (2022)

³⁶ AGGARWAL et al. (2012)

In spite of efforts to promote fact-checking and verifying news, individuals seeking information are unlikely to abandon their old ways and beliefs in favor of new ones. This poses a challenge in decreasing the spread of fake news based on unverified and misleading information. Researchers have developed various fact-checking tools for identifying and distinguishing factual discrepancies, but these tools face practical challenges to their ability to reduce the spread of fake news. This is because fact-checking often clashes with people's deeply held values and beliefs, making it harder for them to accept corrections or recognize the dangers of fake news. Thus, there is a need for further research in this area to inform policy and practice, combat the spread of fake news, and promote the dissemination of reliable information.

Facebook's fourth quarterly report for 2020 revealed the extent of fake news on social media platforms. In three months, Facebook eliminated over 12 million COVID-19-related fake news posts, as well as 1.3 million fake profiles.³⁷

Given the prevalence of fake news on social media, it is crucial to prioritize methods for filtering it out effectively. News dissemination today is characterized by text, audio, visual, and video components. However, verifying the accuracy of news can pose a significant challenge, given the scale, diversity, and heterogeneity of data in the online environment. Developing an inclusive and systematic tool capable of detecting the accuracy of news in this context has proven elusive. This is where text analysis assumes a position of utmost importance, owing to the dominant role of text in the news. In this regard, the discipline of Natural Language Processing serves as the primary point of reference for the analysis of text content. The detection of intentionally misleading news is based on an analysis of numerous examples of both fraudulent and truthful news.

In *Figure 4*, Gungor Yildirim classifies various fake news detection models. One such model is expert-fact-checking, which involves a small group of professionals from various disciplines who can determine whether a piece of news is genuine or fake.

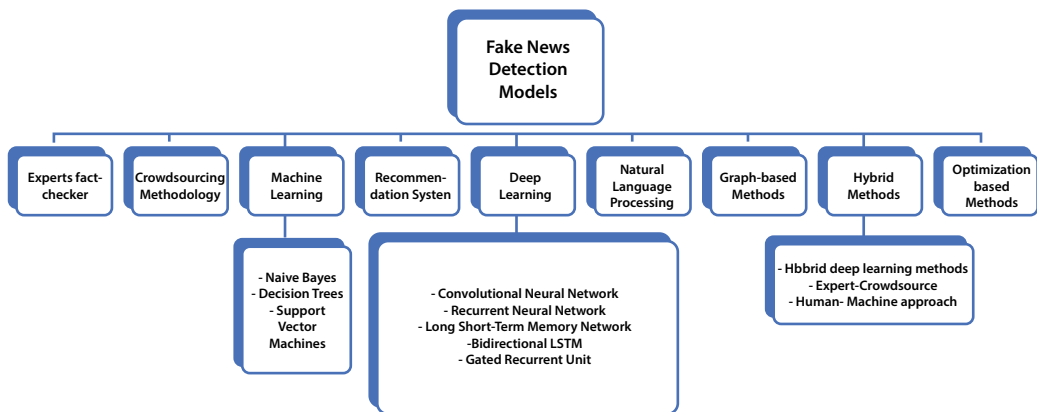


Figure 4. The classification of different fake news detection models (adapted from YILDIRIM [2023])

³⁷ Guy ROSEN: How We're Tackling Misinformation Across Our Apps. *Meta Newsroom*, 22 March 2021. <https://about.fb.com/news/2021/03/how-were-tackling-misinformation-across-our-apps/>

This technique offers a high level of accuracy, along with ease of management. However, the expert-fact-checker approach is time-consuming – particularly when verifying large amounts of information – due to the limited number of experts, and the need for manual annotators. As an alternative, crowdsourcing relies on the collective approval of individuals or groups. The crowd comprises ordinary people from diverse backgrounds who may have limited knowledge of specific news sites, leading them to flag unfamiliar sites as untrustworthy. The strength of crowdsourcing methods lies in their pluralistic nature, utilizing individuals with diverse knowledge.

Nevertheless, the reliability of classification must be verified by determining the number of users affected by the event. Previous research indicates that machine learning algorithms may be capable of detecting fake news, but their training requires a large number of example cases for these models to function effectively. That said, AI-based systems can be effective for automatically verifying reality and detecting fake news. In studies, AI and human analysis have been used to correctly block and tag social media accounts and content flagged as fake in crucial situations and events. These solutions also involve collaboration with fact-checking organizations to carry out intensive manual processes.

In recent year, researchers have conducted many studies with the goal of developing automated solutions to detect fake news, using a variety of different methodologies. However, it is essential to note that while progress has been made to solve the problem, optimal solutions have not yet been found. The significant disadvantages of these methods include their black-box nature, computational efficiency problems, and low interpretability. Additionally, these models require a high degree of optimization in order to acquire suitable values for their parameters.

Among these automated solutions, recommendation systems validate news content which is considered verified and trustworthy and then recommend news articles that are ready for consumption. Collaborative filtering recommendation methods recommend news content based on comments and ratings from other readers/users. Deep learning methods using deep neural network models have also been widely used to determine the authenticity of news content. Additional developments include multimodal models capable of evaluating visual and textual data together, as well as advanced multitask deep learning models able to consider different factors. While the deep models are flexible and can adapt to complex interaction patterns, these methods also pose their own challenges, such as the volume of required data, the inherent complexity of the models, and difficulties with interpretability. The network theory approach will be discussed in more detail later in this study.

According to a study conducted by Vosoughi et al., fake news, especially that related to politics, tends to spread faster and reach a wider audience than legitimate news.³⁸ This study found that fake news was absorbed more deeply across all observed information categories, and in some cases even surpassed the dissemination of real news. Interestingly, it also revealed that humans spread fake news faster than botnets.

The dissemination of fake news has been a growing concern in recent years, with particular instances being motivated by financial gain through pay-per-click ads. Additionally, distributors of healthcare products have been known to share pseudoscientific news to promote their products, as evidenced by the promotion of vitamin C for treating oncological

³⁸ VOSOUGHI–ROY–ARAL (2018)

diseases or coronavirus symptoms. During the COVID-19 pandemic, a photo of a doctor was replaced with that of a porn actor, Johnny Sins, wearing a white lab coat, leading to a polarized response from the public.³⁹ While some individuals found it humorous and shared it as a joke, others believed the doctor in the photo had endorsed the fake quotes, underscoring the importance of identifying the motivation behind sharing fake news.

2.3. Misinformation, Disinformation and Malinformation

Although the terms fake news, misinformation, and disinformation are frequently used interchangeably, it is crucial to differentiate between them, because the motivation for sharing is a key factor. Wardle expanded his taxonomy of fake news in 2017, describing the types of misinformation and disinformation (see *Table 1*).⁴⁰

Type	Description
Satire or Parody	No intention to cause harm, but has the potential to fool
Misleading Content	Misleading use of information to frame an issue or individual
Imposter Content	Impersonating legitimate sources
Fabricated Content	100% fake original content, designed to deceive and do harm
False Connection	Headlines, visuals, or captions that are not supported by the content
False Content	Genuine content shared with false contextual information
Manipulated Content	Genuine information or imagery that is manipulated to deceive

Table 1. 7 Types of Mis and Disinformation (adapted from WARDLE⁴¹)

Nowadays, however, a third concept is increasingly used to differentiate between motivations: malinformation (see *Figure 5*).

³⁹ Saturday Down South: Ole Miss accidentally puts porn star on cups honoring healthcare workers. *SDS.com*, 10 September 2021. <https://www.saturdaydownsouth.com/ole-miss-football/ole-miss-accidentally-puts-porn-star-on-cups-honoring-healthcare-workers/>

⁴⁰ Claire WARDLE: Fake news. It's complicated. *Medium*, 16 February 2017. <https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79>

⁴¹ Ibid.

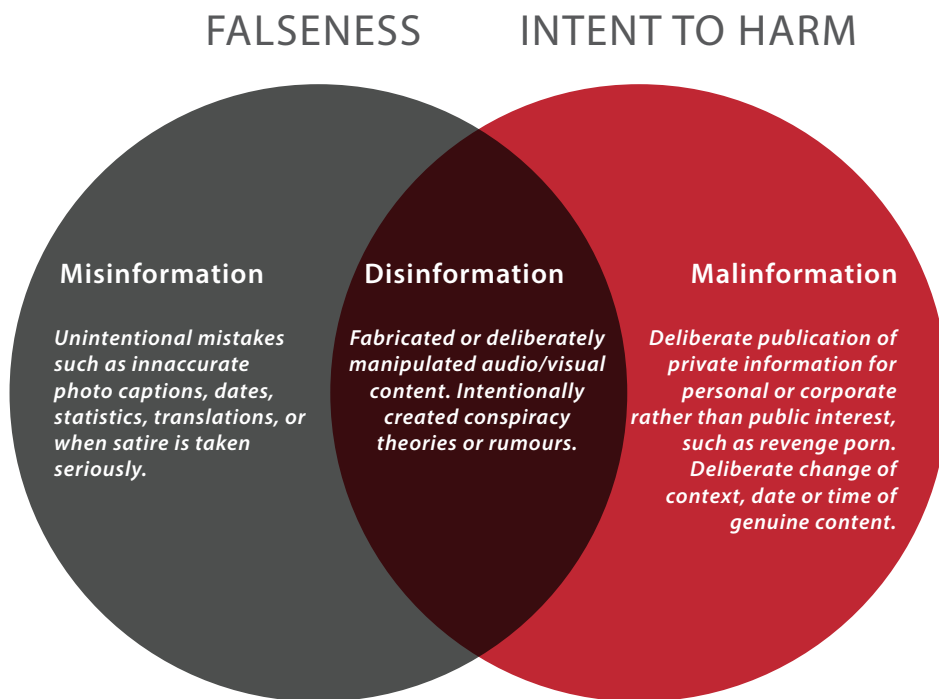


Figure 5. Types of Information Disorder (Source: WARDLE [2018])

Within the context of misinformation, it is noteworthy that the person spreading fake news does not necessarily harbor any malicious intent. Indeed, individuals who are unaware that a piece of information is false may feel compelled to disseminate it on social media when endeavoring to be helpful. This phenomenon can also be attributed to other human vulnerabilities, including the desire to present oneself as well-informed and intelligent. Often, individuals are motivated to share a given piece of news hastily, rather than taking the time to verify its accuracy. A pertinent example of this is the rapid dissemination of news of a celebrity's sudden passing, or a message that seeks to raise awareness of a particular issue (such as the case of Facebook's purported use of users' photos if they fail to copy and paste a specific template text into their news feed).

As opposed to misinformation, disinformation refers to false information that is deliberately created or disseminated with the intention of causing harm. Such disinformation is typically motivated by political, financial, psychological, or social factors. However, the concept of malinformation involves a significant difference. It refers to accurate information that is shared with the intent of causing harm, including the disclosure of private or sensitive information for the purpose of defamation, or to otherwise harm an individual's reputation. Notably, unlike disinformation, malinformation does not necessarily involve the propagation of falsehoods. Instead, it typically entails disseminating messages taken out of context, or those which were once valid in a particular context but are no longer applicable in the present. This may also include misinterpretations of incidents that may be accurate in a different context. One classic example of such misinformation can be found in the old Soviet jokes from Yerevan radio conversations, which were characterized by misinterpreta-

tions and out-of-context information. For instance, a caller to Yerevan radio supposedly once inquired whether the rumor of Mercedes cars being given away in Leningrad was true. The announcer confirmed the news but with a twist: the location was Stalingrad, the vehicles were Moskviches, and they were being looted rather than given away.

The term ‘infodemic’ was coined by David J. Rothkopf in 2003, during the SARS pandemic. It refers to overwhelming amounts of information that can have catastrophic consequences, much like an epidemic, leading to increased illnesses and deaths.⁴² The COVID-19 infodemic has significantly eroded people’s trust in science and reinforced pseudoscientific theories.⁴³ Social networking site algorithms have amplified this trend and facilitated the emergence of absurd conspiracy theorists.⁴⁴ These individuals range from flat-earthers to anti-vaxxers who believe in unfounded theories such as the 5G network causing the coronavirus, and microchips being implanted in individuals through vaccination to monitor their activities.⁴⁵

Although conspiracy theories and misinformation are frequently used interchangeably, they are not equivalent concepts. A conspiracy theory is a suggested interpretation of events that identifies a small group of individuals (the conspirators) who act covertly for their gain in opposition to the public interest. In contrast, misinformation refers to information that is inaccurate or deceptive. While misinformation frequently supports conspiracy theories, this is not necessarily always the case.⁴⁶ In the Hungarian language, a notable distinction should be made between conspiracy theories and ‘konteó’⁴⁷. Even though both terminologies are often used interchangeably in colloquial language, it is important to note that ‘konteó’ originally had its own distinct connotation. Tibor Tóth, also known as tibur, is the author of the popular and once-famous Konteó blog, and is credited with coining the term ‘konteó’. According to Tóth, there exists a clear distinction between the two concepts, which he defines as follows:

“Individuals who blindly adhere to conspiracy theories are characterized by their indiscriminate and uncritical acceptance of a broad range of outlandish claims, from the supposed transformation of the English royal family into reptilian creatures to the existence of Nazi bases on the far side of the moon.

Conversely, the ‘konteó people’ approach these theories with a skeptical and discerning mindset, striving to unveil the hidden motivations and underlying mechanisms behind them. While we aim to uncover the truth, we do not subscribe to the notion that the destiny of humanity hinges on the actions of secretive organizations such as the Bilderberg Group. I think debating and discussing various theories is a challenging and intellectually stimulating activity. Engaging with like-minded individuals can lead to thought-provoking and enriching conversations.”⁴⁸

⁴² Z. KARVALICS (2020)

⁴³ HUA–SHAW (2020)

⁴⁴ ALLINGTON et al. (2021)

⁴⁵ ROMER–JAMIESON (2021)

⁴⁶ ENDERS et al. (2023)

⁴⁷ The term is based on a Hungarian portmanteau of the English term ‘conspiracy theory’.

⁴⁸ BÁNYÁSZ (2016)

Research has shown that social media users are more likely to believe various conspiracy theories about COVID-19, which reinforce anti-vaccination and anti-masking stances, among others.

Figure 6 shows the relationship between these concepts.

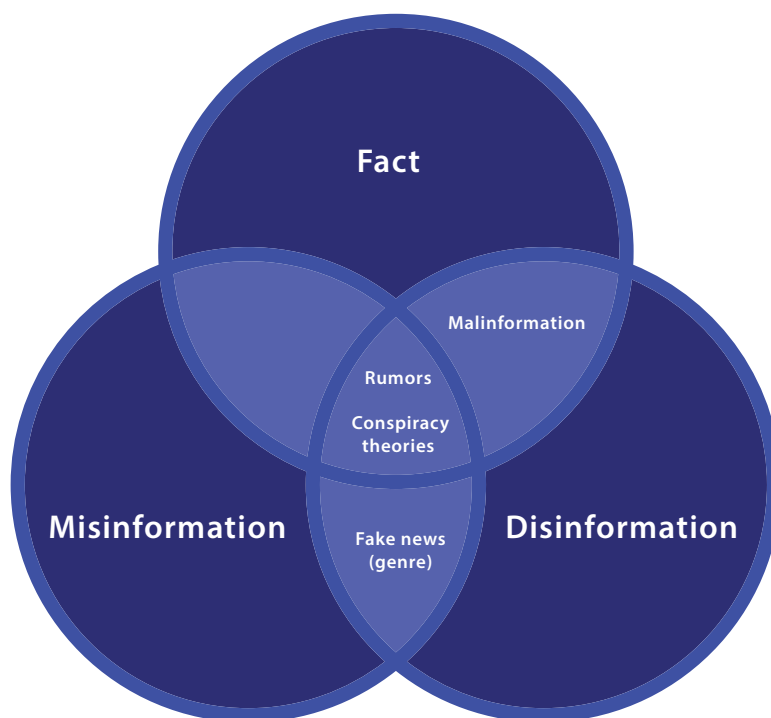


Figure 6. *The relationship between the main concepts (source: Insights blog⁴⁹)*

Scholars have approached the issue of fake news from various perspectives and explored various topics. One such area of focus centers on how governments and international organizations can regulate fake news, particularly disinformation disseminated online. The primary argument in favor of these efforts is to safeguard against threats to democracy.⁵⁰

Accordingly, the European Union has introduced two additional terms to describe the phenomenon of fake news: disinformation and misinformation. The Action Plan for Democracy in Europe, as communicated by the European Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions, provides the following definitions of these related terms:

1. information influence operation refers to a coordinated endeavor, undertaken by either domestic or foreign entities, to deceive a targeted audience by employing disinformation and suppressing independent sources of information;

⁴⁹ Insights blog: Misinformation vs Disinformation. *Taylor & Francis Insights blog*, n.d. <https://insights.taylorandfrancis.com/social-justice/misinformation-vs-disinformation/>

⁵⁰ JUNGHERR-SCHROEDER (2021)

2. foreign interference in the information space denotes a collection of deceitful and coercive methods intended to obstruct and impede the unimpeded formation and expression of the political will of individuals perpetrated by either foreign state actors or their agents.⁵¹

In brief, the phenomena of fake news, disinformation, misinformation, and malinformation all present specific security risks, encompassing the following:

1. misleading and unregulated information that may endanger societal safety;
2. diminishing social reliance on the credibility of the media;
3. external intervention in economic, political, and social processes and groups;
4. imperiling transparent and uncorrupted democratic political processes, such as by manipulating election campaigns;
5. enabling decisions based on inaccurate data that can have detrimental impacts on both individuals and societal security;
6. intensifying social conflict, anarchy, and extremism.⁵²

2.4. Psychological Operations

Identifying and eliminating fake news hinges on the rhetorical question raised in Seneca's play *Medea*, commonly referred to as *cui prodest*, which is an inquiry into who the potential beneficiaries of a given piece of news would be. However, the challenge of eliminating fake news lies in the methodological difficulty of establishing with adequate certainty that the published information is indeed disinformation.

Fake news campaigns have been used to sway political decision-making processes. These campaigns – known as psychological operations or PSYOPs – involve conflicting parties attempting to influence specific target groups through cognitive means. This tactic has been used since ancient times, and is even documented in *The Art of War* by Sun Tzu, who famously stated that “the supreme art of war is to subdue the enemy without fighting.”⁵³ Today, this principle is often associated with hybrid warfare.⁵⁴

Psychological operations are designed to influence the cognitive dimension of a selected target group. The diversification of psychological operations objectives can be identified through a number of vital aspects, such as:⁵⁵

1. influencing the cognitive, affective, and behavioral components of the opposing side;
2. strengthening the support of friendly and loyal populations, in order to attain political and military objectives;

⁵¹ On the European democracy action plan. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM/2020/790 final. *European Commission*, 3. 12. 2020. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0790>

⁵² DOBÁK (2022)

⁵³ TZU (2007)

⁵⁴ HOFFMAN (2007)

⁵⁵ HAIG (2018)

3. securing the support and cooperation of uncommitted or ambivalent target group(s);
4. mitigating the impact of adversary psychological operations on its resources, and on groups requiring protection.

The success of such operations is primarily contingent on the attacker's precision in defining the target group, the content of the message conveyed, and ensuring that the delivery method remains uncompromised. It is important to note that the target group is not necessarily a hostile entity, but may also be the population of the agent's own country or the agent's allies. In a way, virtually every political campaign may be considered a PSYOP.

According to the Information Operations Doctrine of the Hungarian Defense Forces, the conceptual interpretation of psychological operations is centered on the notion that PSYOPS is a method utilized to influence the attitudes, opinions, and behavior of a targeted group to achieve predetermined PSYOPS objectives, as agreed upon by the supervisor. To attain these objectives, a range of PSYOPS activities are designed to evoke or strengthen the desired behavior of the target group, which, in turn, will facilitate the accomplishment of the long-term objectives which have been defined.

Regarding NATO, it is crucial to note that all PSYOPS plans and activities must align with the strategic guidelines for the operations and objectives specified in the Operation Plan. It is also essential for PSYOPS to maintain direct control over the specific 'content' and its dissemination, as well as the target group. Effective psychological operations require significant resources, including adequate intelligence, language support, dissemination mechanisms (such as graphics, print, broadcast, radio, telephone, television, physical press, and voice), appropriate technological tools, and human factors.

When selecting topics for the targeted messages or content, the attacker must choose credible subjects focusing on the vulnerabilities of the target group. The primary objective is to ensure that the message/content is credible, tailored to the target group, and influential. It should also support the attacker's objectives, and establish a reasonable and realistic course of action for the target group.⁵⁶

Regarding the planning of psychological operations, three critical methods are used, based on the classical interpretation of the concept.⁵⁷ These methods are as follows.

1. The strategy of *reflexive control* is intended to influence the decision-making process of an enemy commander. This method involves extensive reconnaissance, as well as penetrating the decision-maker's information systems. Disinformation is then planted to manipulate the information environment in such a way that the enemy decision-maker is influenced to make choices that serve the attacker's interests.
2. The concept of a *social virus* involves infecting a society from within, with people in a particular position spreading fake news. These individuals may be undercover attackers or thought leaders who can be recruited to influence opinions even under a 'false flag'. The spread of fake news typically targets individuals who are ideologically manipulable, politically aggrieved, and not objectively well-informed about current events. Networks can be used to identify thought leaders and influencers,

⁵⁶ HAIG-VÁRHEGYI (2005)

⁵⁷ BÁNYÁSZ et al. (2019)

allowing for the preparation of psychological operations, as well as their detection from a defensive perspective.

3. Conventional methods, such as leaflets or media platforms, as well as more unique methods and tools, can influence the perception of reality and reduce the psychological stability of a person or group. Artificial intelligence research represents one of the most significant branches of such tools. For instance, 'DeepFake' technology utilizes machine deep learning to deliver audiovisual content in which a person's face or voice is superimposed on real-time content. This technology makes it difficult to spot fake news even for the most experienced person, and technological advances are likely to further reinforce this trend.

To better understand psychological operations, it is essential to delve further into certain concepts. Among these concepts, one of the most prominent is propaganda. However, it is essential to note that propaganda here merely serves as a tool that can be employed in various psychological operations. Similar principles apply to fake news, disinformation, misinformation, and malinformation, which can all serve as means to achieve specific psychological objectives.

In the vernacular, psychological operations and propaganda are frequently conflated, yet propaganda is merely a constituent element of psychological operations and, contrary to colloquial terminology, not a prerequisite for advancing a particular political ideology.⁵⁸ While the term PSYOPs is often used interchangeably with propaganda, the latter term carries a negative ideological connotation due to the propaganda factories of the former Nazi and Soviet regimes. Consequently, NATO-related communications and literature prefer the term 'targeted communication'.

Three categories of propaganda can be defined:

1. *White propaganda* typically involves a recognized intermediary, and conveys verifiable information from a reliable source. This form of propaganda frequently utilizes humor and caricatures to belittle and undermine the opposing party.
2. *Black propaganda* is intentionally misleading, and commonly uses false-flag methods in order to delude the intended audience. Such propaganda may create the impression that the source is the government rather than the adversary, thereby obscuring their true intentions.
3. *Grey propaganda* emanates from an unknown source and aims to demoralize the adversary by disseminating false information that targets the enemy. This type of propaganda is intended to lower the morale and fighting spirit of the opposing forces. Grey propaganda is often generated from the 'hinterland', and may touch upon sensitive topics, such as the destruction of soldiers' families or the infidelity of their loved ones.

Propaganda can be defined as reinforcing an exaggerated point of view by disseminating true or false information to advance a particular interest. The primary aim is to persuade a target audience with a specific, significant bias. Propaganda may be employed for political, economic, or military purposes, to demoralize the citizens of other countries, to impact

⁵⁸ MILLER (2015)

‘soft’ targets, to reach large numbers of individuals, or to engender a desired attitude. Historical examples demonstrate that the deliberate, controlled dissemination of information is a crucial tool in propaganda campaigns. Among its constituent elements, we can identify the promotion of uncritical acceptance of simple statements, broad formulations, familiar language, slogans, catchphrases, symbols, illustrations, appeals to desire, and various elements of prestige.⁵⁹

It is possible to distinguish between various categories of propaganda, including but not limited to the following:

1. manipulation of mental influence;
2. the bandwagon technique;
3. the plain folk technique;
4. fear-mongering propaganda;
5. testimonial/reporting propaganda;
6. the false dilemma poser;
7. slogan-based propaganda;
8. statement-type propaganda;
9. the use of fragments of text.⁶⁰

Following the 2016 US national election, Russia has faced recurrent allegations of exploiting cyberspace, mainly social media, to manipulate the internal decision-making processes of foreign nations, primarily by spreading false information. These events have significantly contributed to the recent surge of literature on the subject.⁶¹ Russia has been actively involved in the field of psychological operations since the 2008 Russo-Georgian war, with its intensive deployment of PSYOPs dating back to the 2014 Ukrainian crisis. The RAND

Corporation attributes the success of Russian propaganda to four key factors:

1. dissemination across numerous channels;
2. rapid and frequent dissemination;
3. absence of objectivity;
4. inconsistency.⁶²

3. The future

Artificial intelligence has become inextricable aspect of fake news, as noted previously. For several years, it has been employed in various capacities by the purveyors of fake news, such as for the botnets used for content dissemination. However, AI has also been utilized for the

⁵⁹ DOBÁK (2022)

⁶⁰ Jennifer BETTS: Examples of propaganda done with different tactics. *YourDictionary*, 19 May 2021. <https://www.yourdictionary.com/articles/examples-propaganda-tactics>

⁶¹ FIGUEIRA–OLIVEIRA (2017); JANG–KIM (2018).

⁶² Christopher PAUL – Miriam MATTHEWS: The Russian “Firehose of Falsehood” propaganda model. Why it might work and options to counter it. *RAND.org*, 11 July 2016. <http://www.rand.org/pubs/perspectives/PE198.html>

automated filtering of fake news, identifying anomalies, and mapping patterns. Overall, the advent of generative AI has brought about a paradigm shift in this domain.

In late November 2022, the OpenAI ChatGPT platform was made available to the general public, instantly garnering much attention and triggering significant competition among various AI developers. This hype was likely fueled by Microsoft's investment of \$10 billion in OpenAI. Since then, artificial intelligence has become a ubiquitous buzzword, permeating modern discourse. However, the general understanding of the term is often muddled, which we believe poses a challenge given that public sentiment towards AI can significantly impact its acceptance and utilization. Exaggerated expectations or fears of job displacement can influence the adoption and deployment of technology. Based on Gartner's assessment for 2023, the current position of GenAI on the Hype Cycle is at the Peak of Inflated Expectations, and is anticipated to attain the Plateau of Productivity within the next 5 to 10 years.⁶³

What are the terms most frequently conflated?⁶⁴ The first critical concept is *general artificial intelligence*, which denotes the capacity of a computer system to surpass human performance in any cognitive endeavor. This form of AI is often depicted in science fiction movies, where robots possess self-awareness and act autonomously to achieve their objectives. It is anticipated that the level of general artificial intelligence will be realized by the year 2050. At present, however, the highest level of AI that has been attained is only *narrow AI*, which involves the ability of a computer system to perform specific tasks with greater efficiency than a human. All real-world AI technologies, such as self-driving vehicles and personal digital assistants, fall under this category. Even when the AI appears to be engaging in autonomous, real-time thinking, it is merely coordinating multiple narrowly defined processes and making decisions within a predetermined framework. The AI's 'thinking' capability lacks any awareness or emotion. By contrast, artificial super-intelligence can be defined as a computer system that can surpass humans in nearly all domains, including scientific creativity, general intelligence, and social skills.

It is crucial to differentiate machine learning from deep machine learning. Machine learning is the method computer systems employ for achieving artificial intelligence, which involves utilizing algorithms to identify patterns in data, constructing a data model, and generating predictions based on the model. Machine learning models are trained on subsets of data, and if the data utilized for training the model accurately represents the complete data set to be analyzed, the algorithm will generate more precise outcomes. When a machine learning model is properly trained to perform its task quickly and accurately, it attains narrow AI. Deep machine learning, on the other hand, is an advanced form of machine learning that employs algorithmic networks inspired by the architecture of the human brain, referred to as neural networks. A deep neural network comprises embedded neural nodes; each response leads to new interrelated questions. Deep learning necessitates large data sets, with training sets typically having millions of data points. Following training

⁶³ Lori PERRI: What's New in Artificial Intelligence From the 2023 Gartner Hype Cycle. *Gartner.com*, 17 August 2023. <https://web.archive.org/web/20231218145239/https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle>

⁶⁴ Microsoft Azure: What is Artificial Intelligence? *Microsoft Azure*, n.d. <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-artificial-intelligence>

with these extensive data sets, deep neural networks can handle significantly greater ambiguities than smaller networks. Consequently, they are particularly well-suited for image recognition tasks, where the AI must identify the edges of shapes to discern the content of images. Deep learning is utilized to train AI systems capable of surpassing human capabilities in complex games such as chess.

Generative Artificial Intelligence falls under the category of Large Language Models, which are trained on massive amounts of textual data. These models enhance their abilities with the input data sets derived from diverse text databases such as social media sites and news portals. The noteworthy trait of ChatGPT is its capacity to produce text in real time, mimicking an actual human conversation. Additionally, with the aid of big data, it can accurately predict the next word in a sentence, commonly known as ‘next word prediction’. In essence, the language model learns from a massive corpus of text and retains information regarding the frequency and relation of words, sentences, and paragraphs to generate coherent and meaningful text.

The advent of ChatGPT has generated significant buzz, yet it has also given rise to a degree of confusion among the general populace. While ChatGPT is capable of answering a broad range of questions, and can even write code in Python, it is essentially a language model that generates responses based on common combinations and collocations of words. It is also worth noting that pre-4.0 models were not connected to the internet, and were only trained on a subset of data up until 2021. Nevertheless, the key differentiating factor of ChatGPT is its ability to generate personalized, contextually appropriate responses in real time, using human-like language. Its versatility makes it a valuable tool for handling a wide range of topics, and this feature has made the technology more relatable and accessible to the general public.

The successful deployment of ChatGPT has spawned the development of several comparable platforms. There are numerous generative AI-based platforms currently in existence on the internet, and they are capable of solving a multitude of problems. For instance, individuals can write software in Python without possessing any programming knowledge, or create a script to perform a task within an Excel spreadsheet, generate a PowerPoint presentation, and plot the results on a diagram. Generative AI-based platforms can be classified into various categories, such as text-based, image-based, voice-based, programming-based, and video-based.

Generative artificial intelligence also poses several cybersecurity risks, as even individuals with limited IT knowledge can create tools capable of inflicting significant harm. Several significant analyses and reports have surfaced on this topic, including one from the RAND Corporation⁶⁵, as well as the Cyber Threat Report disseminated by the Hungarian National Cyber Security Center Institute, highlighting the cybersecurity challenges posed by generative artificial intelligence.⁶⁶

This technology has also facilitated the proliferation of fake news, as text-based platforms can quickly generate grammatically correct fake news content. By creating fake profiles with

⁶⁵ MARCELLINO et al. (2023)

⁶⁶ Nemzeti Kibervédelmi Intézet: *A ChatGPT kiberbiztonsági kockázatai. CTI Jelentés*. Budapest, Nemzeti Kibervédelmi Intézet, 2023.

https://nki.gov.hu/wp-content/uploads/2023/03/A-ChatGPT-kiberbiztonsagi-kockazatai_CTI_jelentes.pdf

convincing profile pictures (see *Figure 7*) and appropriate background information, users can generate many fake news stories and distribute them automatically. The attributes of the face to be generated – such as gender, age group and ethnicity – can be freely chosen. The process of image generation does not necessitate any specialized expertise; it is sufficient to simply select the desired options in the online image generator. It is worth noting that, in contrast to Dall-E and other analogous image creation sites, this platform does not require inputting a prompt. The creation of background information, designed to construct a narrative surrounding a generated face, is comparably straightforward. The process involves the selection of age, gender, name, and country of origin, resulting in the immediate generation of fictitious data that is nonetheless clearly linked to the designated country (see *Figures 8–9*). The system also provides geo-coordinates for the fictitious individual's address, which may be utilized to verify the veracity of the stated location (see *Figure 10*).



Figure 7. Example of face generation (own generation, source: *ThisPersonDoesNotExist*⁶⁷)

⁶⁷ <https://this-person-does-not-exist.com/en>



Figure 8. Randomly generated background information, part 1.
(own generation, source: FakeNameGenerator⁶⁸)



Figure 9. Randomly generated background information, part 2.
(own generation, source: FakeNameGenerator)

⁶⁸ <https://www.fakenamegenerator.com/>



Figure 10. *Randomly generated background information, part 3
(own generation, source: FakeNameGenerator)*

In addition, voice-based AI can produce convincing clones of an individual's voice with as little as 10 minutes of audio recordings. The author of this article was able to clone his voice deceptively well using a website called ElevenLab and 40 minutes of audio recordings. Collecting this amount of audio from a public figure is relatively straightforward. As a result, generative artificial intelligence has made it significantly easier for individuals to produce high-quality content that can be leveraged to spread fake news.

As previously discussed, the primary motive behind generating noise is to instill a sense of doubt and suspicion towards almost anything and anyone. This extends to the point where even factual data is viewed as untrustworthy, eroding public trust. The emergence of live voice and facial replication technology, which allows individuals to imitate others' voices or mimic them in video calls, has only heightened this concern. In essence, the prevalence of such advancements raises questions about the reliability of our senses and the extent to which we can rely on them.

References

- ABOKHODAIR–YOO–MCDONALD (2015) = Norah ABOKHODAIR – Daisy YOO – David W. MCDONALD: Dissecting a social botnet. Growth, content and influence in twitter. In: Association for Computing Machinery: *CSCW '15: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, Association for Computing Machinery, 2015. 839–851. <https://doi.org/10.1145/2675133.2675208>
- AGGARWAL et al. (2012) = Rohit AGGARWAL – Ram GOPAL – Ramesh SANKARANARAYANAN – Param VIR SINGH: Blog, blogger, and the firm. Can negative employee posts lead to positive outcomes? *Information Systems Research*, Vol. 23., No. 2. (2012) 306–322. <https://doi.org/10.1287/isre.1110.0360>
- ALLINGTON et al. (2021) = Daniel ALLINGTON – Bobby DUFFY – Simon WESSELY – Nayana DHAVAN – James RUBIN: Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine*, Vol. 51., No. 10. (2021) 1763–1769. <https://doi.org/10.1017/S003329172000224X>
- AZUCAR–MARENGO–SETTANNI (2018) = Danny AZUCAR – Davide MARENGO – Michele SETTANNI: Predicting the Big 5 personality traits from digital footprints on social media. A meta-analysis. *Personality and Individual Differences*, Vol. 124. (2018) 150–159. <https://doi.org/10.1016/j.paid.2017.12.018>
- BÁNYÁSZ (2016) = BÁNYÁSZ Péter: A közösségi média szerepe a lélektani műveletekben az elmúlt időszak válságainak tükrében. *Szakmai Szemle*, 2016/1. 61-81. https://s2prodstorage.blob.core.windows.net/s2cmsblob/Files/Szakmai%20Szemle/SZSZ_2016_1_szam.pdf?sp=r&st=2025-03-25T09:37:58Z&se=2035-03-25T16:37:58Z&spr=https&sv=2024-11-04&sr=c&sig=PAOb3jtCphbbzRgBEutrTUH2uSjRNnF3G8dVaL8Y3Z0%3D
- BÁNYÁSZ et al. (2019) = BÁNYÁSZ Péter – DOBOS László – PALLA Gergely – POLLNER Péter: Lélektani műveletek a közösségi médiában. In: AUER Ádám – JOÓ Tamás (szerk.): *Hálózatok a közszolgálatban*. Budapest, Dialóg Campus, 2019. 111–133.
- BLUMLER–KATZ (1974) = Jay G. BLUMLER – Elihu KATZ (Eds.): *The uses of mass communications: current perspectives on gratifications research*. Sage Annual Reviews of Communication Research Volume III. Beverly Hills, Sage Publications, 1974.
- DOBÁK (2022) = DOBÁK Imre: A dezinformáció – napjaink kiemelt kihívása. *Katonai Jogi és Hadijogi Szemle*, 2022/1. 93–124. https://epa.oszk.hu/02500/02511/00020/pdf/EPA02511_katonai_jogi_szemle_2022_1_093-124.pdf
- ENDERS et al. (2023) = Adam M. ENDERS – Joseph E. USCINSKI – Michelle I. SEELIG – Casey A. KLOFSTAD – Stefan WUCHTY – John R. FUNCHION – Manohar N. MURTHI – Kamal PREMARATNE – Justin STOLER: The relationship between social media use and beliefs in conspiracy theories and misinformation. *Political Behavior*, Vol. 45. (2023) 781–804. <https://doi.org/10.1007/s11109-021-09734-6>
- EŞKİSU–HOŞOĞLU–RASMUSSEN (2017) = Mustafa EŞKİSU – Rumeysa HOŞOĞLU – Kyler RASMUSSEN: An investigation of the relationship between Facebook usage, Big Five,

- self-esteem and narcissism. *Computers in Human Behavior*, Vol. 69. (2017) 294–301. <https://doi.org/10.1016/j.chb.2016.12.036>
- FIGUEIRA–OLIVEIRA (2017) = Álvaro FIGUEIRA – Luciana OLIVEIRA: The current state of fake news: challenges and opportunities. *Procedia Computer Science*, Vol. 121. (2017) 817–825. <https://doi.org/10.1016/j.procs.2017.11.106>
- GRAVES (2016) = Luca GRAVES: *Deciding What's True: The Rise of Political Fact-Checking in American Journalism*. New York, Columbia University Press, 2016.
- GREENWALD (2014) = Glenn GREENWALD: *No palce to hide. Edward Snowden, the NSA, and the U.S. Surveillance State*. New York, Metropolitan, 2014.
- GROVER–KAR–DWIVEDI (2022) = Purva GROVER – Arpan Kumar KAR – Yogesh DWIVEDI: The evolution of social media influence. A literature review and research agenda. *International Journal of Information Management Data Insights*, Vol. 2., No. 2. (2022) Art. No. 100116. <https://doi.org/10.1016/j.jjime.2022.100116>
- HAIG (2018) = HAIG Zsolt: *Információs műveletek a kibertérben*. Budapest, Dialóg Campus, 2018.
- HAIG–VÁRHEGYI (2005) = HAIG Zsolt – VÁRHEGYI Iván: *Hadviselés az információs hadszíntéren*. Budapest, Zrínyi, 2005.
- HARRISON–WALKER–JIANG (2023) = L. Jean HARRISON–WALKER – Ying JIANG: Suspicion of online product reviews as fake. Cues and consequences. *Journal of Business Research*, Vol. 160. (2023) Art. no. 113780. <https://doi.org/10.1016/j.jbusres.2023.113780>
- HE et al. (2014) = Qiwei HE – Cees A.W. GLAS – Michal KOSINSKI – David J. STILLWELL – Bernard P. VELDKAMP: Predicting self-monitoring skills using textual posts on Facebook. *Computers in Human Behavior*, Vol. 33. (2014) 69–78. <https://doi.org/10.1016/j.chb.2013.12.026>
- HINDS–WILLIAMS–JOINSON (2020) = Joanne HINDS – Emma J. WILLIAMS – Adam N. JOINSON: “It wouldn’t happen to me”. Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, Vol. 143. (2020) Art. no. 102498. <https://doi.org/10.1016/j.ijhcs.2020.102498>
- HOFFMAN (2007) = Frank G. HOFFMAN: *Conflict in the 21st century. The rise of hybrid wars*. Arlington, Potomac Institute for Policy Studies Arlington, 2007.
- HUA–SHAW (2020) = Jinling HUA – Rajib SHAW: Corona Virus (COVID-19) “Infodemic” and emerging issues through a data lens. The case of China. *International Journal of Environmental Research and Public Health*, Vol. 17., No. 7. (2020) Art. no. 2309. <https://doi.org/10.3390/ijerph17072309>
- JANG–KIM (2018) = S. Mo JANG – Joon K. KIM: Third person effects of fake news. Fake news regulation and media literacy interventions. *Computers in Human Behavior*, Vol. 80., 2018. 295–302. <https://doi.org/10.1016/j.chb.2017.11.034>
- JUHÁSZ et al. (2015) = Attila JUHÁSZ – Lóránt GYÖRI – Péter KREKÓ – András DEZSŐ: *“I am Eurasian”. The Kremlin connections of the Hungarian far-right*. Budapest, Political Capital, 2015. https://politicalcapital.hu/wp-content/uploads/PC_SD_I_Boll_study_IamEurasian.pdf

- JUNGHERR–SCHROEDER (2021) = Andreas JUNGHERR – Ralph SCHROEDER: Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy. *Social Media + Society*, Vol. 7., No. 1. (2021) <https://doi.org/10.1177/2056305121988928>
- KAPLAN–HAENLEIN (2010) = Andreas M. KAPLAN – Michael HAENLEIN: Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, Vol. 53., No. 1. (2010) 59–68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- KAPOOR et al. (2018) = Kawaljeet Kaur KAPOOR – Kuttamani TAMILMANI – Nripendra P. RANA – Pushp PATIL – Yogesh K. DWIVEDI – Sridhar NERUR: Advances in Social Media Research. Past, Present and Future. *Information Systems Frontiers*, Vol. 20., No. 3. (2018) 531–558. <https://doi.org/10.1007/s10796-017-9810-y>
- KHONDKER (2015) = Habibul Haque KHONDKER: New Media, Political Mobilization, and the Arab Spring. In: James D. WRIGHT (Ed.): *International Encyclopedia of the Social & Behavioral Sciences* [2nd ed.]. Amsterdam, Elsevier, 2015. 798–804. <https://doi.org/10.1016/B978-0-08-097086-8.64135-1>
- LAZER et al. (2018) = David M. J. LAZER – Matthew A. BAUM – Yochai BENKLER – Adam J. BERINSKY – Kelly M. GREENHILL – Filippo MENCZER – Miriam J. METZGER – Brendan NYHAN – Gordon PENNYCOOK – David ROTHSCHILD – Michael SCHUDSON – Steven A. SLOMAN – Cass R. SUNSTEIN – Emily A. THORSON – Duncan J. WATTS – Jonathan L. ZITTRAIN: The science of fake news. *Science*, Vol. 359, No. 6380. (2018) 1094–1096. <https://doi.org/10.1126/science.aao2998>
- LEVAK (2021) = Tomislav LEVAK: Disinformation in the New Media System. Characteristics, forms, reasons for its dissemination and potential means of tackling the issue. *Medijska istraživanja*, Vol. 26., No. 2., (2021) 29–58. <https://doi.org/10.22572/mi.26.2.2>
- LEWANDOWSKY–ECKER–COOK (2017) = Stephan LEWANDOWSKY – Ullrich K. H. ECKER – John COOK: Beyond misinformation. Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition*, Vol. 6., No. 4. (2017) 353–369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- MARCELLINO et al. (2023) = William MARCELLINO – Nathan BEAUCHAMP-MUSTAFAGA – Amanda KERRIGAN – Lev Navarre CHAO – Jackson SMITH: *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0- Next-Generation Chinese Astroturfing and Coping with Ubiquitous AI*. Santa Monica, RAND, 2023. <https://doi.org/10.7249/PEA2679-1>
- MILLER (2015) = David MILLER: Sociology, propaganda and psychological operations. In: Matt DAWSON – Bridget FOWLER – David MILLER – Andrew SMITH (Eds.): *Stretching the sociological imagination*. Basingstoke, Palgrave Macmillan, 2015. 163–188. https://doi.org/10.1057/9781137493644_9
- MUNK (2009) = MUNK Veronika: Sztárság, elméletben. *Médiakutató*, 2009/1. 7–16. https://mediakutato.hu/cikk/2009_01_tavasz/01_sztarsag_elmeletben
- NORMAN (1963) = Warren T. NORMAN: Toward an adequate taxonomy of personality attributes. Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, Vol. 66., No. 6. (1963) 574–583. <https://doi.org/10.1037/h0040291>

- OLAN et al. (2024) = Femi OLAN – Uchitha JAYAWICKRAMA – Emmanuel Ogiemwonyi ARAKPOGUN – Jana SUKLAN – Shaofeng LIU: Fake news on Social Media. The Impact on Society. *Information Systems Frontiers*, Vol. 26. (2022) 443–458.
<https://doi.org/10.1007/s10796-022-10242-z>
- PENNYCOOK–RAND (2021) = Gordon PENNYCOOK – David G. RAND: The psychology of fake news. *Trends in Cognitive Sciences*, Vol. 25., No. 5. (2021) 388–402.
<https://doi.org/10.1016/j.tics.2021.02.007>
- PINXTEN–LIEVENS (2014) = Wouter PINXTEN – John LIEVENS: The importance of economic, social and cultural capital in understanding health inequalities. Using a Bourdieu-based approach in research on physical and mental health perceptions. *Sociology of Health & Illness*, Vol. 36., No. 7, (2014) 1095–1110. <https://doi.org/10.1111/1467-9566.12154>
- PLÉH–BOROSS (2010) = PLÉH Csaba – BOROSS Otilia: *Pszichológiai lexikon*. Budapest, Akadémiai, 2010.
- ROMER–JAMIESON (2021) = Daniel ROMER – Kathleen Hall JAMIESON: Patterns of media use, strength of belief in COVID-19 conspiracy theories, and the prevention of COVID-19 from March to July 2020 in the United States. Survey study. *Journal of Medical Internet Research*, Vol. 23., No. 4., (2021) Art. no. e25215.
<https://doi.org/10.2196/25215>
- SPIER (2017) = Shaked SPIER: What is Social Media: A Critical View. In: Shaked SPIER: *Collective action 2.0. The impact of social media on collective action*. Cambridge, Chandos, 2017. 15–31. <https://doi.org/10.1016/B978-0-08-100567-5.00002-5>
- SPOHR (2017) = Dominic SPOHR: Fake news and ideological polarization. Filter bubbles and selective exposure on social media. *Business Information Review*, Vol. 34., No. 3. (2017) 150–160.
<https://doi.org/10.1177/0266382117722446>
- SYMEONIDIS et al. (2017) = Iraklis SYMEONIDIS – Gergely BICZÓK – Fatemeh SHIRAZI – Cristina PÉREZ-SOLÀ – Jessica SCHROERS – Bart PRENEEL: Collateral damage of Facebook third-party applications: a comprehensive study. *Computers & Security*, Vol. 77. (2017) 179–208.
<https://doi.org/10.1016/j.cose.2018.03.015>
- TAJFEL (1974) = Henri TAJFEL: Social identity and intergroup behaviour. *Social Science Information*, Vol. 13., No. 2. (1974) 65–93. <https://doi.org/10.1177/053901847401300204>
- TANDOC–LIM–LING (2018) = Edson C. TANDOC – Zheng Wei LIM – Richard LING: Defining ‘Fake News’. *Digital Journalism*, Vol. 6., No. 2. (2018) 137–153.
<https://doi.org/10.1080/21670811.2017.1360143>
- TZU (2007) = Sun Tzu: *The Art of war*. [First thus edition.] Las Vegas, Filiquarian, 2007.
- VENKATESH (2000) = Viswanath VENKATESH: Determinants of perceived ease of use. integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, Vol. 11., No. 4. (2000) 342–365.
<https://doi.org/10.1287/isre.11.4.342.11872>

- VOSOUGHI–ROY–ARAL (2018) = Soroush VOSOUGHI – Deb ROY – Sinan ARAL: The spread of true and false news online. *Science*, Vol. 359., No. 6380. (2018) 1146–1151. <https://doi.org/10.1126/science.aap9559>
- WARDLE (2018) = Claire WARDLE: The need for smarter definitions and practical, timely empirical research on information disorder. *Digital Journalism*, Vol. 6., No. 8. (2018) 951–963. <https://doi.org/10.1080/21670811.2018.1502047>
- YILDIRIM (2023) = Gungor YILDIRIM: A novel hybrid multi-thread metaheuristic approach for fake news detection in social media. *Applied Intelligence*, Vol. 53., No. 9. (2023) 11182–11202. <https://doi.org/10.1007/s10489-022-03972-9>
- Z. KARVALICS (2020) = László Z. KARVALICS: *From ‘infodemic’ to the need for global information and knowledge governance*. Köszeg, Institute of Advanced Studies Köszeg, 2020. <https://iask.hu/en/from-infodemic-to-the-need-for-global-information-and-knowledge-governance-by-laszlo-z-karvalics/>
- ZHANG– HONG–CRANOR (2007) = Yue ZHANG – Jason I. HONG – Lorrie F. CRANOR: Cantina. A content-based approach to detecting phishing web sites. In: Carey WILLIAMSON – Mary ELLEN Zurko – Peter PATEL-SCHNEIDER – Prashant SHENOY (Eds.): *Proceedings of the 16th international conference on World Wide Web*. Banff, Alberta, Canada, May 8–12, 2007. New York, Association for Computing Machinery, 2007. 639–648. <https://doi.org/10.1145/1242572.1242659>

The relevance of psychological operations in social media – Findings from empirical research

Part 2

PÉTER BÁNYÁSZ

1. Introduction

This research paper employs case studies to investigate three critical occurrences in the 2020s. These events had a significant impact on the dissemination of false information. The first event was the emergence of a new variant of coronavirus, which caused widespread panic and confusion. Following a series of incidents where scientific investigations were found to be biased or influenced by external parties, there was a significant decline in public trust in the scientific community. As a result, people began to question the validity and accuracy of scientific research, which significantly impacted the public's perception of scientific findings and recommendations. This decrease in trust had far-reaching consequences, affecting everything from medical treatments and vaccinations to environmental policies and technological advancements. The second event was the use of psychological operations during the Russian-Ukrainian conflict, which led to the spread of propaganda and disinformation. The third event was the adoption of generative artificial intelligence (GAI), which enabled the creation of convincing fake news articles, images, and videos.

By conducting a comprehensive analysis of these events, our aim is to gain a deeper understanding of how fake information spreads, and the detrimental effects it can have on our society.

This study investigates the sharing of fake news on social media, and the factors contributing to its spread. To do so, we will analyze social media data to identify the prevalence and patterns of sharing fake news. We will also explore the factors influencing individuals to share fake news, such as their political affiliation, educational background, and online social networks. Furthermore, we will assess the impact of fake news on people's beliefs and behaviors, as well as its broader societal impact.

2. Methodology

The first step in our research was a comprehensive analysis of how search terms related to the various concepts explored in our study are represented in the scientific literature worldwide. In order to accomplish this task, we used the Scopus database, which is widely recognized as Elsevier's primary database of scientific publications. Scopus is a comprehensive database granting diverse and extensive access to over 36,000 titles from more than 11,000 publishers. Its collection spans numerous subjects, from life and social sciences to physical and health sciences. The listed publications include over 34,000 peer-reviewed journals that delve into topics within top-level

subject fields. Scopus offers three types of sources, including book series, journals, and trade journals, all of which are immensely useful for conducting research.

By utilizing this valuable resource, we were able to explore the indexed publications and their metadata,¹ gaining a deeper understanding of the nuances of the research in our field. Using this approach, we obtained a more comprehensive and well-rounded overview of the research landscape, enabling us to produce more robust and insightful results for our study.

In the course of our inquiry, we analyzed specific search terms on Scopus, encompassing subjects such as social media, psychological operations, misinformation, disinformation, malinformation, and fake news. By gathering a substantial amount of data, we conducted a comprehensive trend analysis that involved examining the primary disciplines and the institutions with publications on the topic at hand. To gain a deeper understanding, we conducted a thorough study of the associated keywords using network analysis. Our goal was to identify the most pertinent research areas through keyword frequency analysis. We are confident that our approach provided valuable insights into the subject matter; the more frequently certain keywords occur, the more relevant the associated research areas.

In the second phase of our study, we performed sentiment analysis on various terms linked with COVID-19 and the Russian-Ukrainian war. The focus of sentiment analysis lies in evaluating the emotional impact of shared content. To accomplish this, we employed SentiOne (sentione.com), a content-based web analytics platform spanning 70 languages worldwide. Its purpose is to crawl and analyze content disseminated and perceived via social media and other online channels. Presently, the platform monitoring tool oversees over 20,000,000,000 mentions and collects data from eight distinct sources, such as portals, blogs, X (formerly Twitter), Facebook, Instagram, video, forums, and review sites. The mentions are categorized into statements and articles, with statements being automatically classified as either positive, neutral, or negative via SentiOne's unique, proprietary algorithm. The sentiment analysis offered by the platform is based on research conducted by John R. Crawford and Julie D. Henry, who analyzed the Positive and Negative Affect Schedule (PANAS).

In order to examine the dissemination of particular instances of false information pertaining to the Russian-Ukrainian conflict, we integrated sentiment analysis with a network theory methodology.

3. Results

3.1. Scientometrics

The study of dynamics is an integral part of scientific exploration. According to Yan, scientific communities may grow or shrink in size, new species or diseases may be discovered, and new re-

¹ The metadata of a publication refers to a vast amount of information, going well beyond merely the title and author names. In fact, metadata is a rich source of details that provides a complete picture of a research work and its context. It includes factors such as author affiliations, citation count, publication date, and much more. This information is crucial for researchers, scholars, and other stakeholders as it helps them to understand the significance and impact of a publication beyond its content. In other words, metadata offers a holistic view of the research and its relevance in the broader academic and scientific community.

search topics may emerge.² Some topics continue to be studied on an ongoing basis, while others may disappear over time. It is crucial to understand research dynamics in order to comprehend the evolving cognitive structures of research domains. Previous research studies used bibliographic data to effectively identify research specialties. However, these studies merely provided a static view of the cognitive structure of research domains, which did not fully capture research dynamics. To address this, a new study proposes using topic modeling techniques, by introducing two sets of topic attributes: topic continuity and topic popularity.³ By applying these attributes to a library and information science publication dataset, we can better explore the dynamic research landscape of any field. In academic research, topic modeling techniques can be used to cluster papers, journals, or authors using probabilistic models. A topic can be defined as a probability distribution over terms in a vocabulary.

Regarding Shamszadeh et al., scientometrics, also known as science mapping, is frequently employed along with information visualization to investigate a large body of bibliographic materials and measure diverse scientific activities.⁴ Scientometricians often utilize techniques from scientometrics, information visualization, and text mining to analyze the evolution of various scientific fields. Such analyses allow for mapping the research output in a particular field, which can aid researchers and funding agencies in identifying underexplored areas and making more efficient decisions related to a given field of research. In addition, this makes it easier to identify collaborative interactions or clusters, which promotes further collaboration. A fundamental aspect of scientometric analysis is bibliometric mapping/analysis: by extracting relationships between publications, bibliometric mapping offers a rapid method of summarizing and visualizing the inherent structure within a set of publications. These visualizations, or maps, can be used not only to examine the history of scientific research in a particular field, but also to identify potential future research directions and opportunities for collaboration.

In our study, we analyzed the occurrence of six terms within the Scopus literature database. The following six terms are delineated within the literature summary:

1. social media;
2. fake news;
3. misinformation;
4. disinformation;
5. malinformation;
6. psychological operation.

It is imperative to note that the term analysis was restricted to publications published no later than the end of 2023. This decision was based on two principal rationales. Firstly, we contend that comparing complete, concluded years is conducive to accurately analyzing the data. Secondly, Scopus frequently lists publications several years in advance that journals have approved but not yet published.

² YAN (2014)

³ MIYATA et al. (2020)

⁴ SHAMSADEH–ASGARY–NOSRA (2019)

Table 1. displays the frequency of publications for each respective term.

Keyword	Document
Social media	175,136
Misinformation	16,549
Fake News	9,171
Disinformation	5,749
Psychological Operation	228
Malinformation	42

Table 1. The occurrence of the terms under consideration (adapted from Scopus)

The diagram depicted in *Figure 1* portrays the chronological distribution of publications with regard to the keywords explored. It is important to note that the semantics of certain terms have undergone significant transformation over the years. For instance, the term ‘social engineering’ has evolved to describe a type of attack that exploits human susceptibility, whereas its original meaning, coined in 1842, was simply to influence the societal norms and attitudes of specific groups. The original concept closely resembled contemporary psychological operations; however, it differed fundamentally in that physical violence was deemed an acceptable means for achieving its objectives.

The meaning of the very term ‘social media’ has also changed since its inception. Its first appearance in Scopus dates back to 1941, in an article by Trow W.C. titled “Educational psychology – individual or social?” published in the *Journal of Consulting Psychology*.⁵ With the evolution of the concept and the emergence of social media platforms, there has been an exponential surge in the use of the term, particularly since 2010. It is expected that by the end of 2023, a total of 175,136 publications will be listed in Scopus for the keyword ‘social media’.

⁵ Trow (1941)

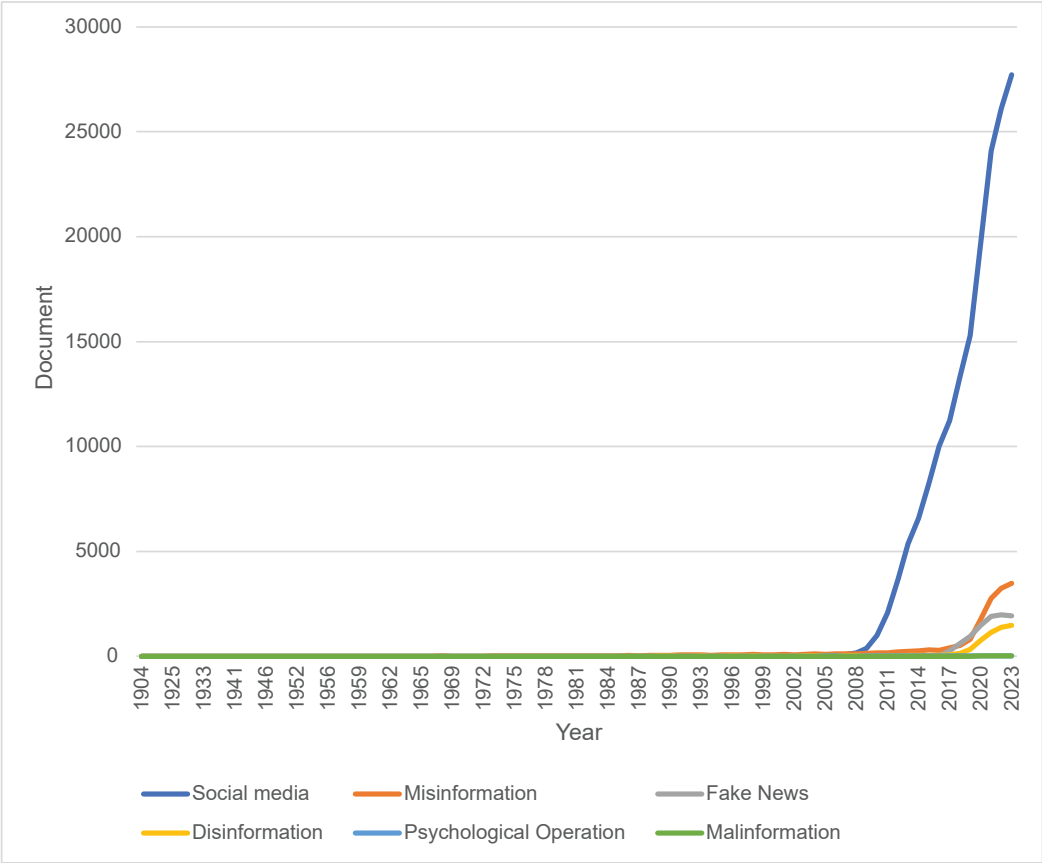


Figure 1. Distribution of publications over time (adapted from Scopus)

Combating fake news is a highly intricate issue that can only be tackled through an interdisciplinary and multidisciplinary approach. Consequently, it is well worth scrutinizing the scholarly disciplines in which publications relevant to our research terms have been published. Accordingly, the disciplines associated with the keywords have been enumerated in *Table 2*. It should be noted that certain publications may have been allocated to multiple disciplines, resulting in a higher count than the number of publications assigned to the said term. The term ‘social media’ itself is a good example of this, with 300,296 instances per discipline, in contrast to the total of 175,136 publications featured within the database.

Area	Social media	Misinfor- mation	Fake News	Disinfor- mation	Psychological Operation	Malinfor- mation
Computer Science	61057	3738	4751	1629	54	17
Social Sciences	57774	4949	3552	2996	100	25
Engineering	21604	1306	1729	515	59	1
Arts and Humanities	14736	1522	1166	753	38	1
Mathematics	11468	804	1166	304	13	2

Area	Social media	Misinformation	Fake News	Disinformation	Psychological Operation	Malinformation
Decision Sciences	10214	578	926	279	10	3
Medicine	41345	6223	660	1016	17	8
Business, Management and Accounting	20938	557	526	257	19	3
Psychology	10985	1645	420	257	41	0
Physics and Astronomy	3008	225	310	66	0	0
Economics, Econometrics and Finance	7233	316	226	152	10	2
Materials Science	1883	170	212	43	1	0
Energy	2818	152	183	48	1	2
Environmental Science	6614	542	159	169	7	4
Multidisciplinary	3202	463	121	146	1	0
Neuroscience	2488	326	100	73	10	0
Biochemistry, Genetics and Molecular Biology	3368	552	81	111	6	0
Agricultural and Biological Sciences	2743	369	62	49	1	0
Chemical Engineering	732	83	55	17	2	0
Nursing	4456	574	52	64	0	0
Chemistry	480	82	42	19	0	0
Earth and Planetary Sciences	2254	157	41	48	2	0
Health Professions	3418	372	39	60	0	0
Immunology and Microbiology	1663	468	33	79	0	0
Pharmacology, Toxicology and Pharmaceutics	2395	390	31	68	0	0
Dentistry	698	80	11	6	0	0
Veterinary	720	133	4	16	0	0
Undefined	2	11	0	0	0	0

Table 2. *Distribution of keywords by discipline (adapted from Scopus)*

Within the domain of scientometrics, we are particularly interested in analyzing keyword distribution. Our working hypothesis posits that the frequency of a given keyword within a particular research area is directly proportional to its perceived relevance. Authors typically

select keywords which they believe best encapsulate the content of their publications. As such, if other authors also employ those exact same keywords, this gives a strong indication of current research trends. While keywords can also be examined concerning their temporal relevance, our study employs a network theoretical approach.

Albert-László Barabási's *Networked* was a pivotal work that sparked the emergence of network science as a unique field across many disciplines. A core postulate of the field is that every network is inherently complex, possessing diverse properties and a wide range of degrees of freedom. Network analysis is thus a powerful tool for understanding complex systems, as it helps uncover hidden patterns and relationships from large amounts of unstructured data. It benefits public institutions and their leaders in making informed decisions that can impact their organizations and society as a whole. By mastering a networked approach, individuals can adeptly execute complicated processes, resulting in more efficient outcomes. Decision-making should be objective and based on quantifiable characteristics, allowing officials and decision-makers to make informed use of impact assessments and pre-decision analyses for the benefit of their constituents.⁶

To conduct our keyword analysis of the data gathered from Scopus, we leveraged the capabilities of VOSViewer, a software program designed to construct and visualize bibliometric networks.⁷ Such networks can encompass various entities, including publications, researchers, or journals, and may be based on various relations, such as citations, co-citations, bibliographic coupling, or co-authorship. VOSViewer also boasts text mining capabilities facilitating the creation and visualization of co-occurrence networks comprising significant terms extracted from the scientific literature.

The study focused on three distinct dimensions:

1. network visualization;
2. overlay visualization;
3. density visualization.

The network visualization shows items denoted by their labels, with circular shapes as the default representation. The size of the label and the circular shape is determined by the item's weight, with larger weights yielding larger dimensions. Labels may be hidden in some instances, to avoid label overlap. The color of an item corresponds to the cluster to which it belongs, while lines indicate links between items. The proximity of two keywords in the visual representation indicates their relatedness based on co-citation links, with closer keywords indicating stronger associations.

The overlay visualization exhibits identical features to the network visualization, except for item coloration. Items may have one of two colors in the overlay visualization. If an item possesses a score, the color assigned to the item corresponds to its score, ranging from blue (lowest score) to green to yellow (highest score). A color bar is displayed in the bottom right corner of the visualization, but only if item colors are score-driven, illustrating the mapping between scores and colors.

The visualization for item density follows a similar pattern of representing items with their labels, much like in the network and overlay visualization. Each point in the item den-

⁶ SZÓCSKA et al. (2019)

⁷ VAN ECK – WALTMAN (2010)

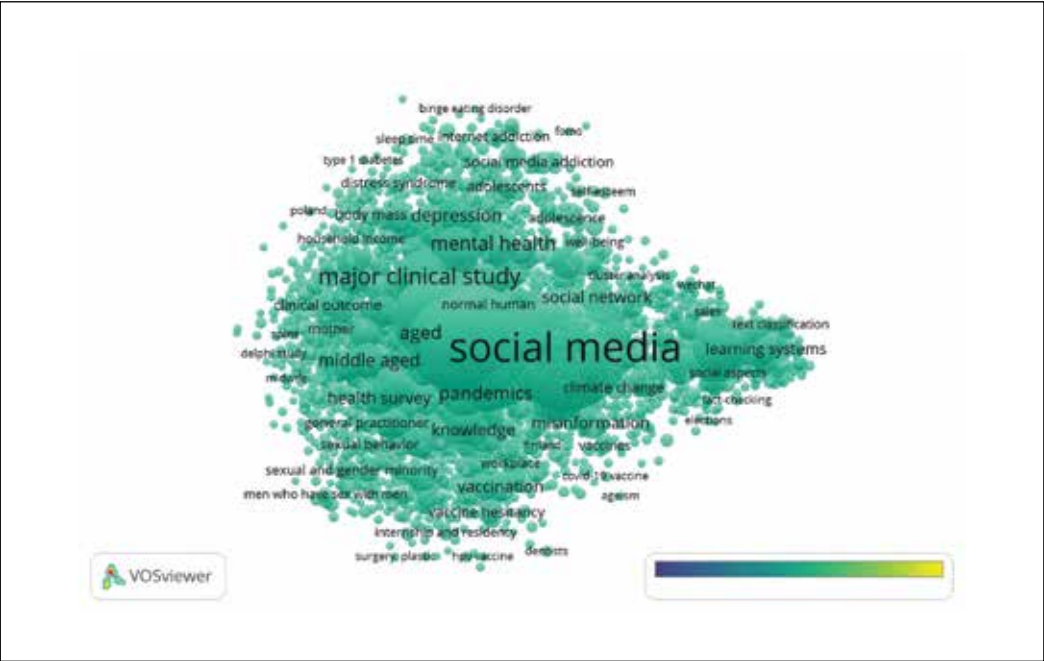


Figure 3. Keywords related to social media – overlay visualization (adapted from Scopus)



Figure 4. Keywords related to social media – density visualization (adapted from Scopus)

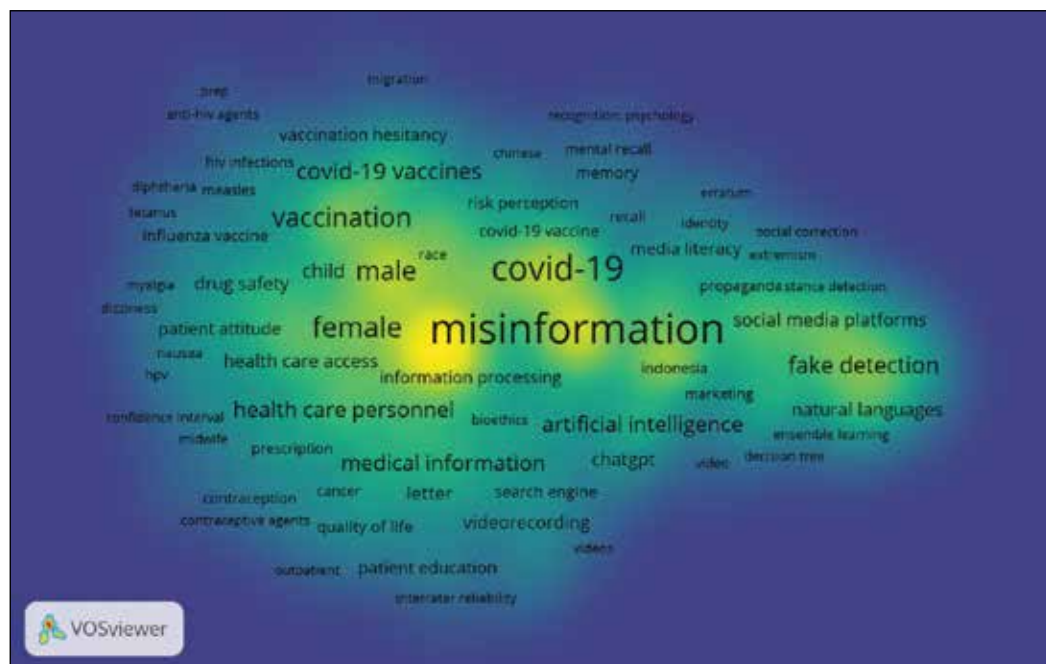


Figure 7. *Keywords related to misinformation – density visualization (adapted from Scopus)*

Fake news-related keywords are depicted in *Figures 8–10*. These keywords can be categorized into eight distinct clusters, as shown in *Figure 8*.

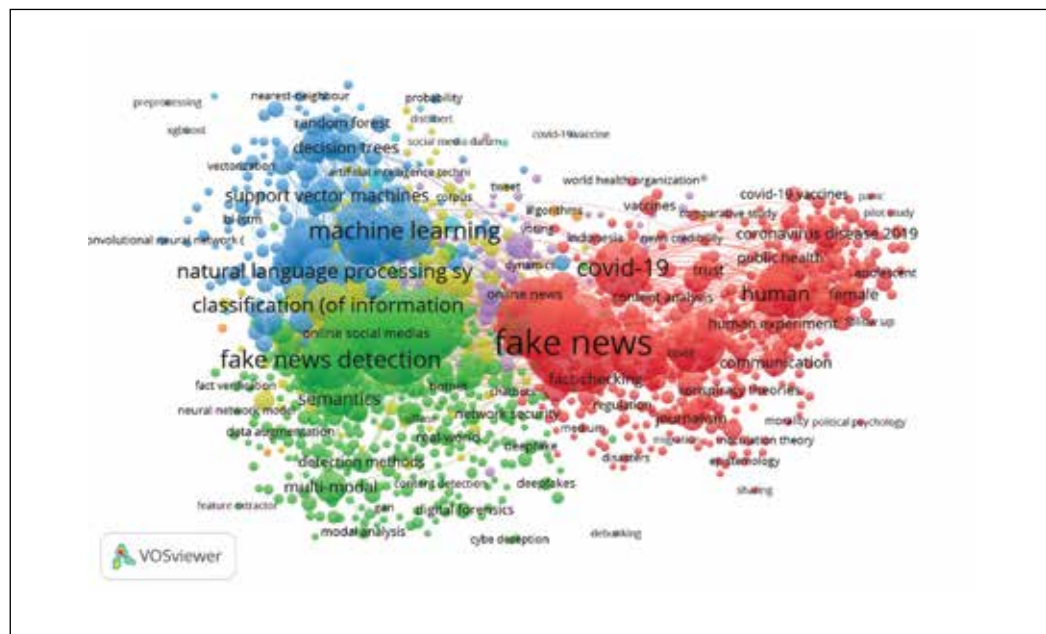


Figure 8. Keywords related to fake news – network visualization (adapted from Scopus)

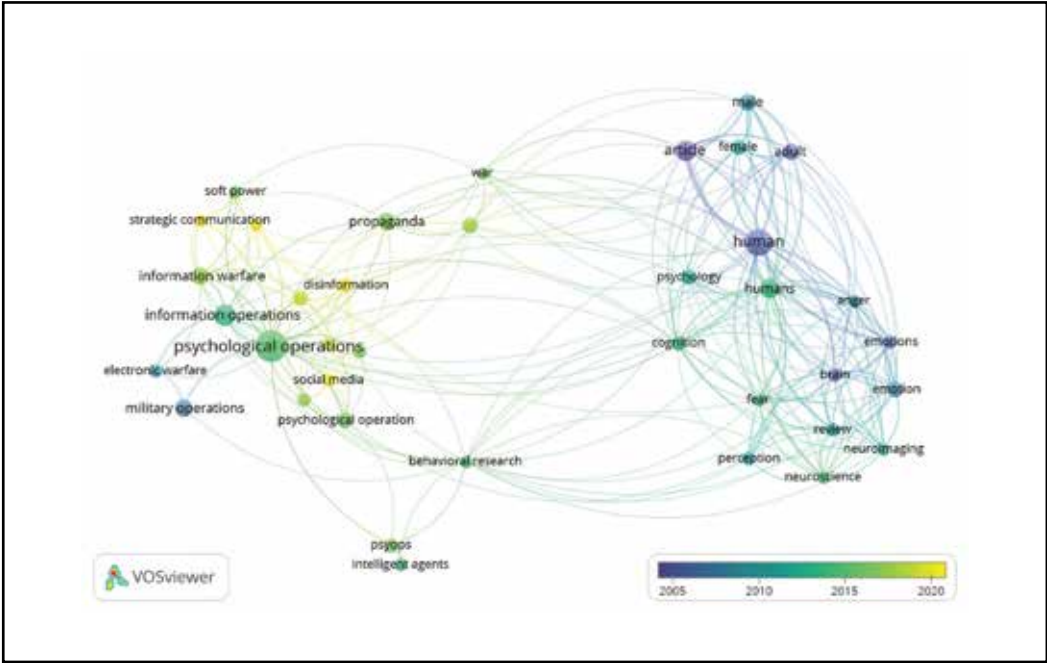


Figure 15. Keywords related to psychological operations – overlay visualization (adapted from Scopus)

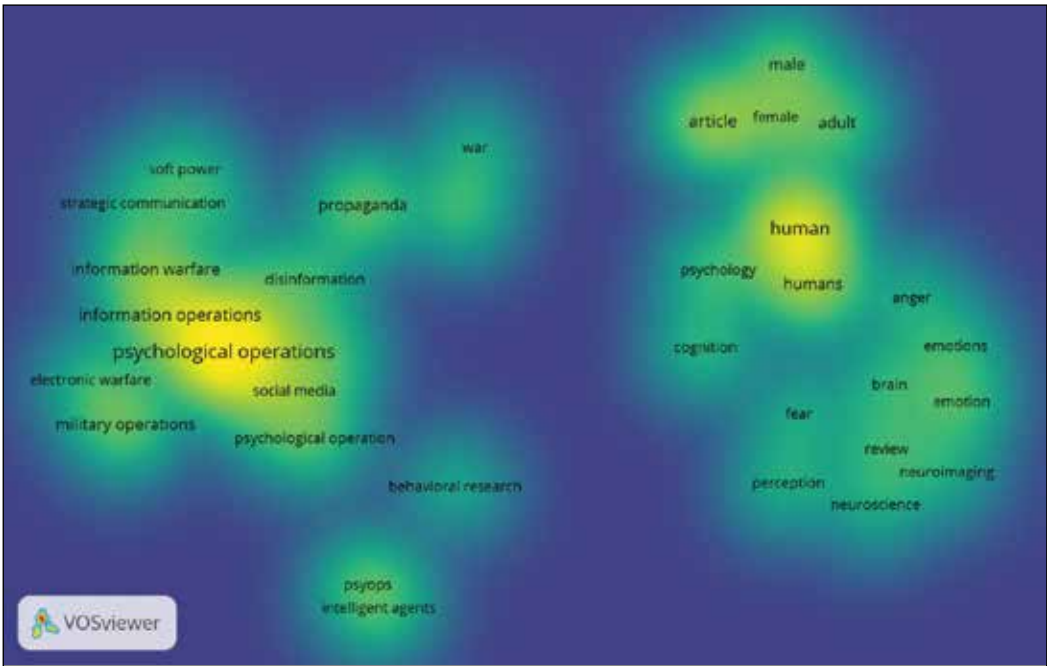


Figure 16. Keywords related to psychological operations – density visualization (adapted from Scopus)

Finally, malinformation-related keywords are depicted in *Figures 17–19*. These keywords can be categorized into three distinct clusters, as shown in *Figure 17*.

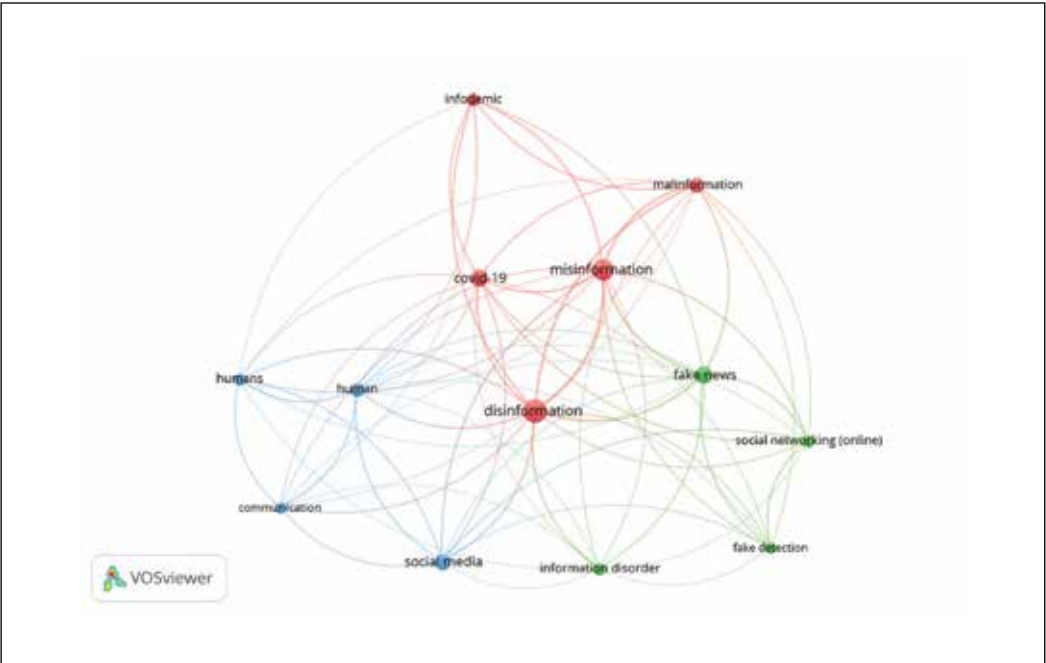


Figure 17. Keywords related to malinformation – network visualization (adapted from Scopus)

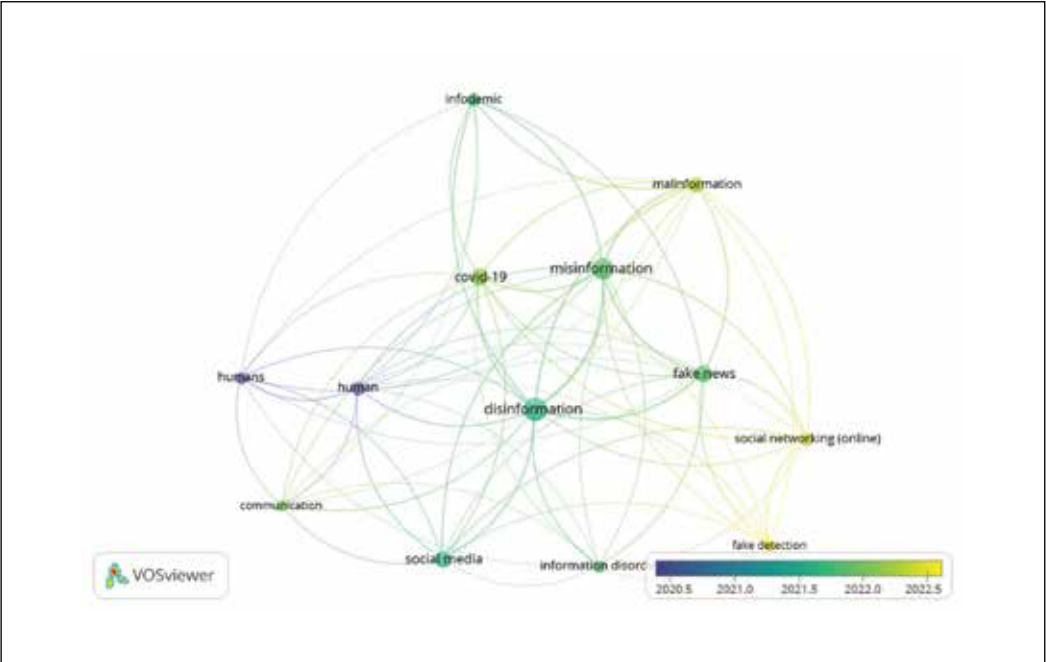


Figure 18. Keywords related to malinformation – overlay visualization (adapted from Scopus)

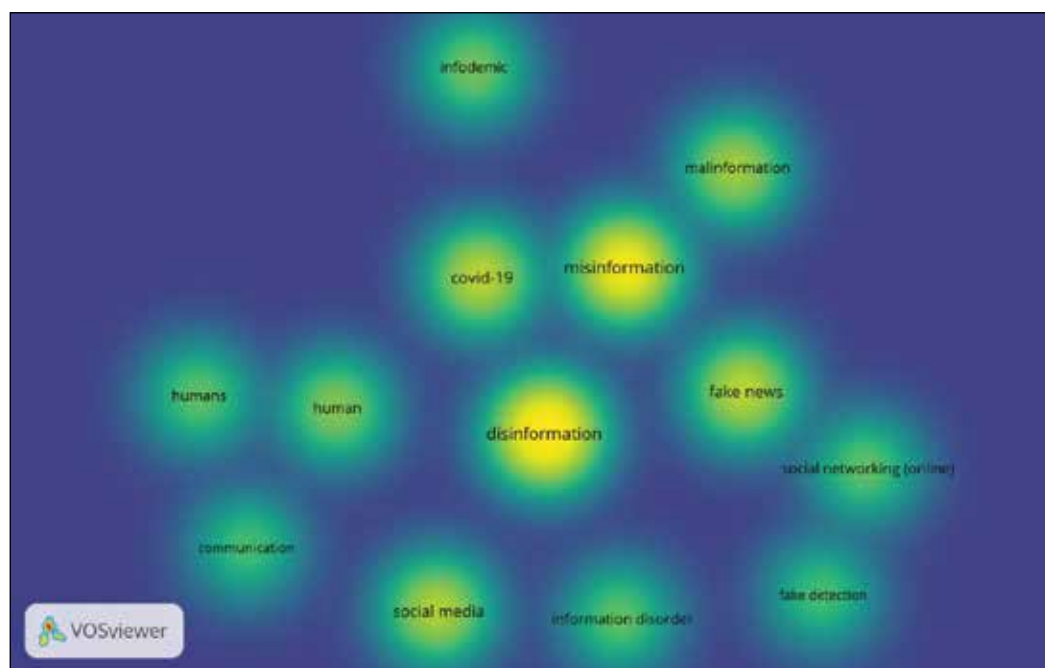


Figure 19. Keywords related to malinformation: density visualization (adapted from Scopus)

3.2. COVID-19

As of the time of writing, the COVID-19 pandemic has persisted in China for over four years. Despite extensive research and investigation, the origins of this novel coronavirus remain shrouded in mystery, providing fertile ground for a host of related conspiracy theories. In the present study, we adopt a neutral stance regarding the virus's origin, concurring with Péter Marton's assertion that its source is ultimately immaterial to its significant impact on individuals and societies worldwide. Among the pandemic's far-reaching consequences, perhaps the most noteworthy has been the erosion of public trust in scientific institutions and practices.

Massive, global fake news campaigns had already existed before 2020, often making rather outlandish claims. One need only think of Pizzagate from 2016, focusing on a supposed cabal of satanic, pedophile, human-trafficking cannibal Democrat politicians in America. Donald Trump's 2016 presidential campaign saw the spread of many similar conspiracies, which have now grown into distinct movements combining multiple conspiracy theories, such as QAnon or the State Deniers. The proliferation of conspiracy theories in Hungary has also increased, with a large number of individuals subscribing to such beliefs. As of January 2024, many of these individuals have not only made the news for their non-compliance with police arrests, but have also been found to be preparing for a *coup d'état* while accumulating weapons. These ardent supporters of conspiracy theories have found a collective identity during the COVID-19 pandemic, where instead of their normal state of hostility to each other, they have united to form a new, shared paradigm. This con-

vergence of conspiracy theorists is believed to have been facilitated by a combination of factors, including the anxiety generated by the pandemic, the messaging from official state and non-governmental organizations, and the characteristics of social media. The World Health Organization (WHO) has recently faced a decline in trust, exacerbated by its initial communication during the COVID-19 pandemic that wearing masks was not essential. The decision was based on the limited supply of masks and the concern that informing people of the need for masks would result in panic buying, leading to a shortage for frontline workers. The production of masks rapidly increased, however, allowing the WHO to announce that mask-wearing was now mandatory. This announcement was then resisted by some who rejected the need for masks based on conspiracy theories, turning it into a political issue. In 2021, the WHO named the new coronavirus variant 'Omicron' instead of 'B.1.1.529', in order to avoid any association with Chinese President Xi Jinping.

In January 2020, a post surfaced on social media claiming that the proliferation of 5G technology was the cause of the novel coronavirus.⁸ This claim, in turn, triggered an upsurge in anti-Chinese sentiment. In April of the same year, this 5G-related misinformation resulted in the vandalization of 77 5G transmitting stations in the UK by individuals apprehensive about the novel coronavirus, in a supposed attempt to impede its spread.⁹ It is vital for researchers and policy-makers to accurately understand the patterns and attitudes of citizens with regard to this matter, as they can leverage this understanding to take more well-informed decisions and tackle imminent risks, such as the link between anti-vaccination and fabricated news.

To achieve this objective, diverse search terms were employed to scrutinize the affective expressions of online discourse on coronavirus vaccination, and the associated disinformation.

Our methodology began with sentiment analysis, which entailed analyzing the openly accessible online content linked to various keywords from 1 January 2020 to 31 July 2021. In this investigation, we investigated the following search terms that were employed in the analysis:

1. 'COVID' AND '5G';
2. 'COVID' AND 'Bill Gates';
3. 'COVID' AND 'microchip'.

To execute the search, we utilized the AND operator for the search terms, which returns only results that include both search terms. The use of quotation marks is key for multi-word expressions, as it enables all the terms within the quotation marks to be considered an indivisible part of the search term. Without quotation marks, searching for the term 'Bill Gates' would yield all results containing either Bill or Gates, rather than the exact phrase. The search for the keywords was conducted globally, across 30 languages.

It is important to distinguish between mentions and reach. The latter refers to the number of times a page is viewed or reloaded, such as when a user revisits their news feed or

⁸ Motoko RICH: As Coronavirus Spreads, So Does Anti-Chinese Sentiment. *The New York Times*, 30 January 2020. <https://www.nytimes.com/2020/01/30/world/asia/coronavirus-chinese-racism.html>

⁹ Corinne REICHERT: 5G coronavirus conspiracy theory leads to 77 mobile towers burned in UK. *CNET*, 7 May 2020. <https://www.cnet.com/health/5g-coronavirus-conspiracy-theory-sees-77-mobile-towers-burned-report-says/>

clicks on a news article multiple times. As shown in *Table 3*, the reach rates are remarkably high. This phenomenon can be attributed to the constant repetition of propaganda messages, which can eventually lead to their acceptance by the target audience.

	Number of shares			
	Mentions	Reach	Positive	Negative
COVID and 5G	1 156 535	~4,4 billion	76 842	140 949
COVID and Bill Gates	1 414 097	~6,3 billion	100 946	187 062
COVID and microchip	145 452	~449 million	10 068	14 599

Table 3. Distributions of the analyzed keywords globally (adapted from SentiOne)

As shown in *Figure 20*, the number of shares linked to COVID and 5G increased considerably during the spring of 2020, reflecting the anxiety articulated in the previously cited article of 15 April 2020. This was the time when a group of British citizens set fire to 5G towers, in a supposed attempt to impede the spread of COVID-19. The figure predominantly shows neutral emotions, however, which can be attributed to two factors. Firstly, a sizable proportion of the shares are merely re-shares, devoid of affective context to evaluate, thereby rendering them neutral as per the algorithm. Secondly, a common reason for the neutral classification is when a share comprises almost equal numbers of positively and negatively labeled terms, which the algorithm identifies as neutral.

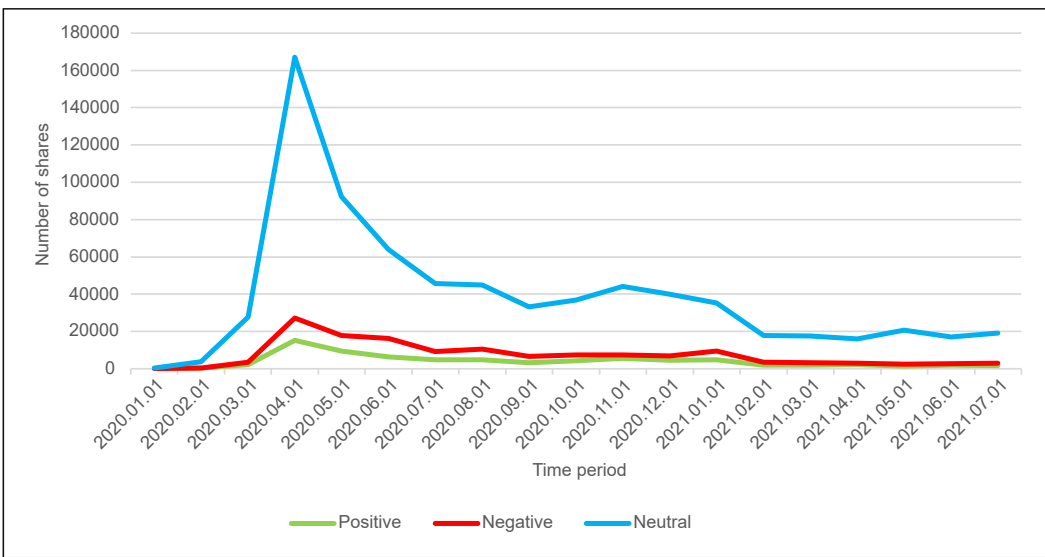


Figure 20. Number of mentions of COVID and 5G globally (adapted from SentiOne)

Regarding *Figure 21*, it may not come as a surprise that the largest volume of shares associated with the topic in question is linked to the UK, which was a notable exception during April 2020, as shown in *Figure 8*. In the period under analysis, the UK had 317,313 shares, whereas Hungary ranked 11th globally with 12,498 shares.

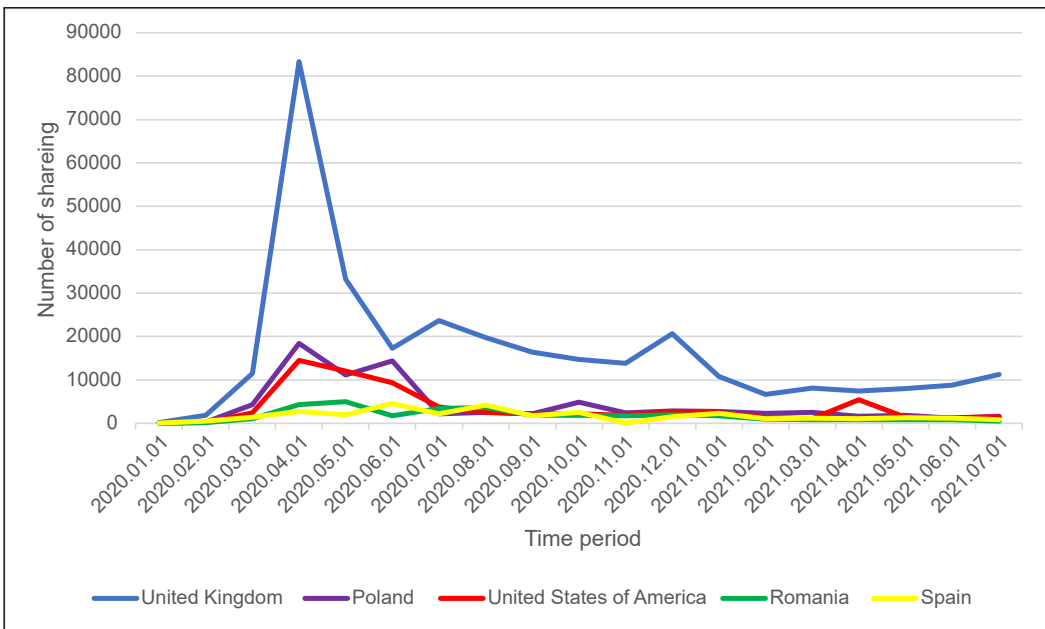


Figure 21. Number of mentions of COVID and 5G globally, by country (adapted from SentiOne)

The graphical representation in *Figure 22* depicts the top ten sources, with the online version of the Hungarian newspaper HVG holding a notable position among them, owing to the extensive coverage of Hungary-related content. This information is beneficial for devising counter-campaigns, as it provides insights into the attitudes of various sources. Analyzing such data makes it possible to tailor targeted campaigns, keeping in mind the unique characteristics of the audience in a particular region. This research holds significant academic value, as it highlights the importance of having a nuanced understanding of the media landscape for effective communication strategies.

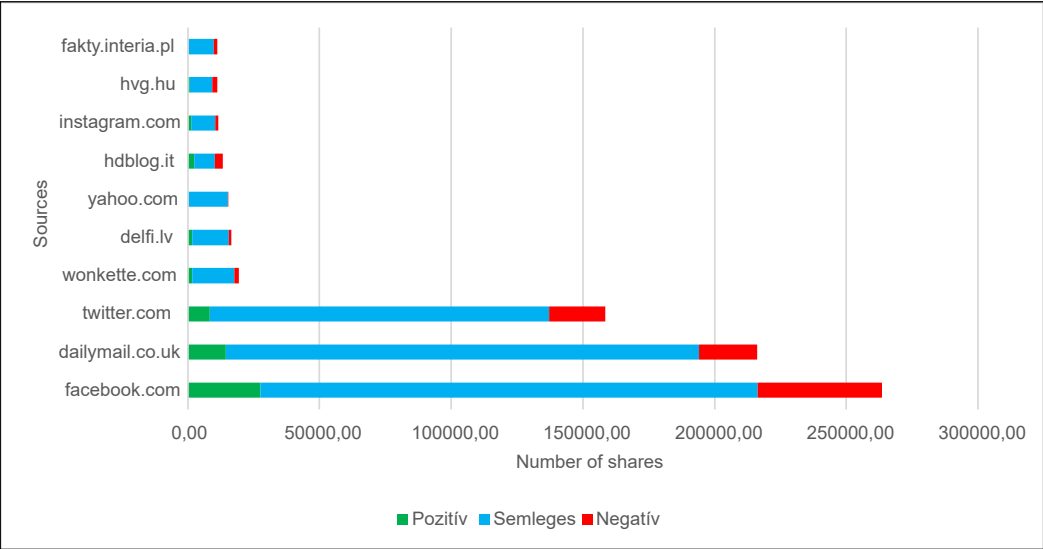


Figure 22. Number of mentions of COVID and 5G, by source (adapted from SentiOne)

Regarding the other two keywords, we have not examined the sources in the context of this study; in the case of 5G, our aim was simply to illustrate its relevance.

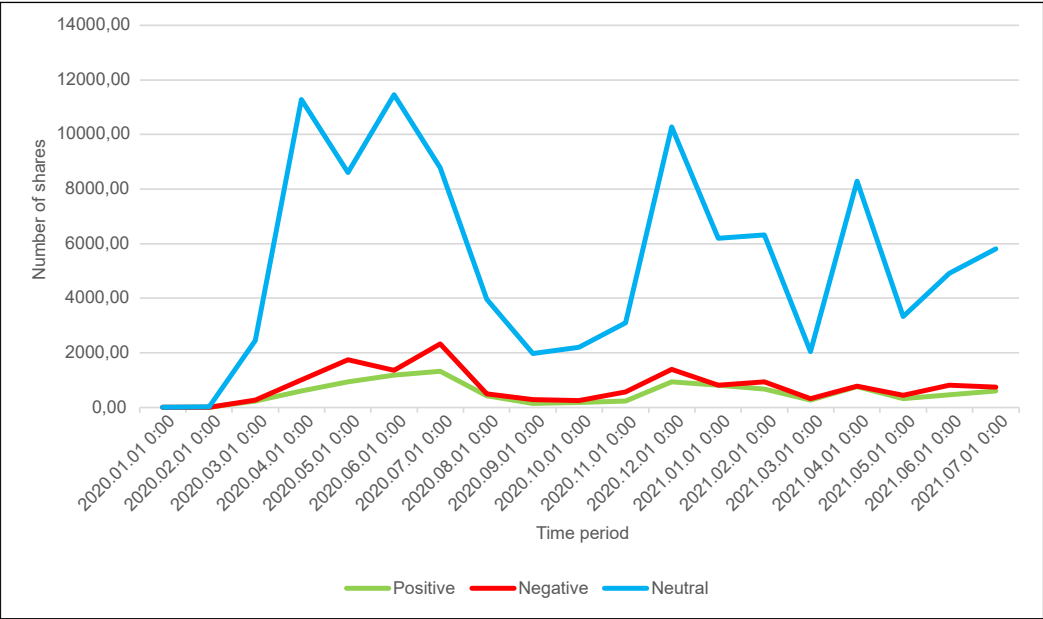


Figure 23. Number of mentions of COVID and microchip globally (adapted from SentiOne)

During the period under analysis, the terms ‘COVID’ and ‘microchip’ were mentioned 145,452 times in a fluctuating pattern, as shown in *Figure 23*. Notably, there was a significant

surge in April 2020, which coincided with the propagation of false information concerning 5G technology.

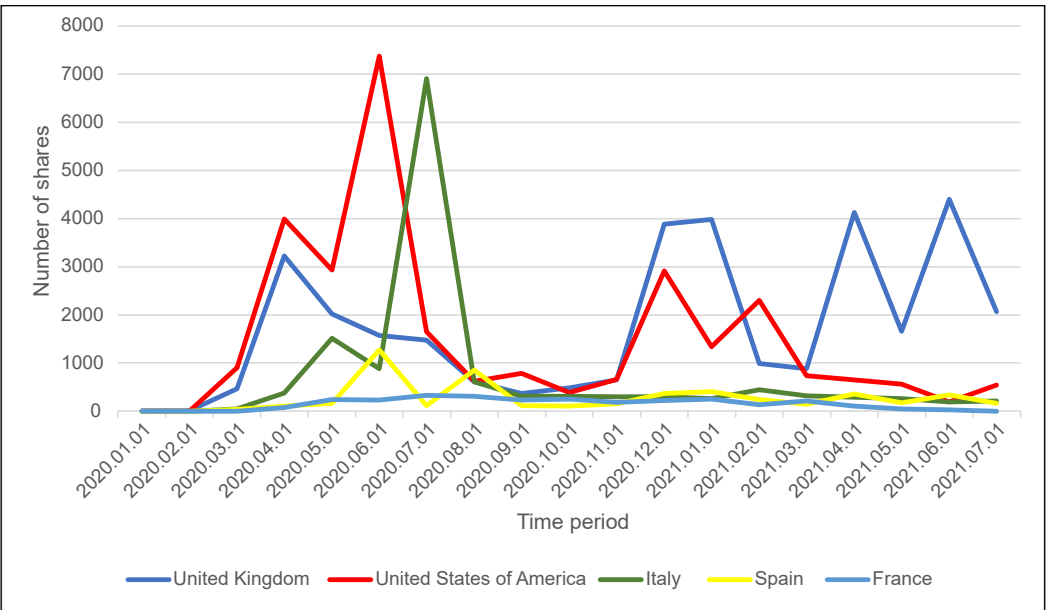


Figure 24. Number of mentions of COVID and microchip globally, by country (adapted from SentiOne)

Breaking down the distribution by country, the United Kingdom remains at the top of the list, while the United States is second (see Figure 24).

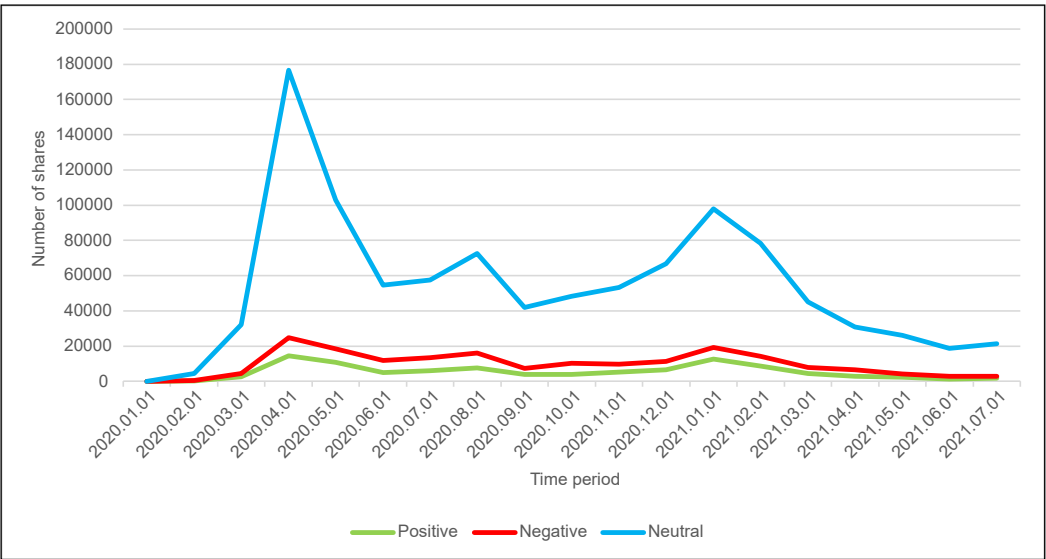


Figure 25. Number of mentions of COVID and Bill Gates globally (adapted from SentiOne)

Figure 25 illustrates Bill Gates as a recurrent protagonist in numerous conspiracy theories, garnering 1,414,097 shares during the period under analysis. Specifically, Gates has been targeted as a scapegoat for COVID-related issues due to a 2010 presentation, and his name has been linked to vaccine development.

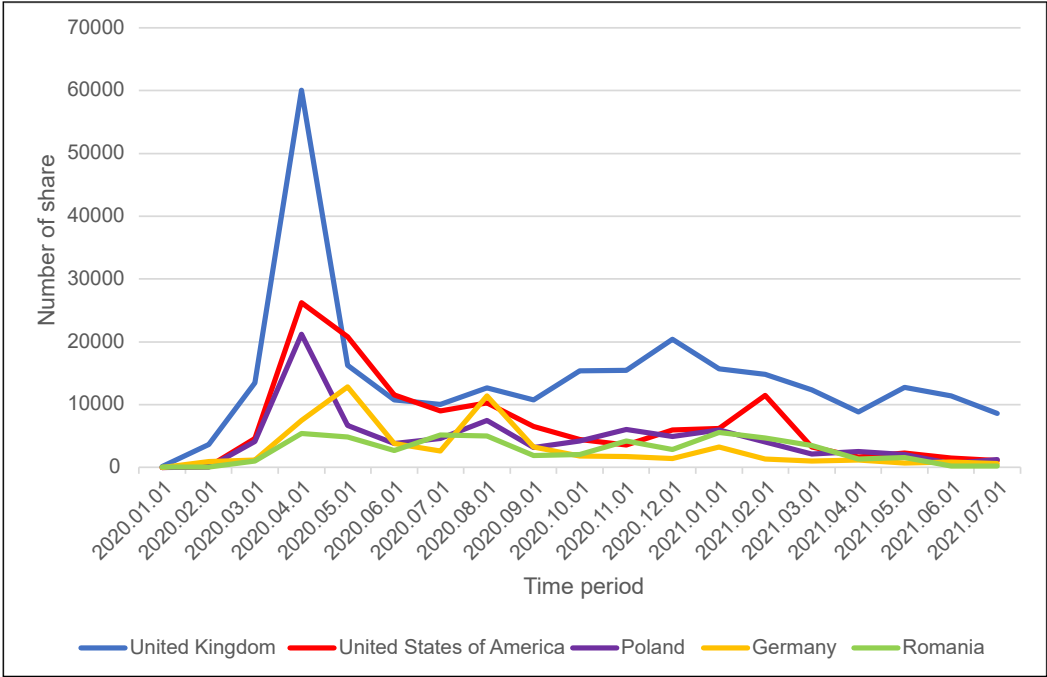


Figure 26. Number of mentions of COVID and Bill Gates globally, by country (adapted from SentiOne)

As expected, the United Kingdom maintains its lead regarding sharing activity for the two aforementioned topics, while the United States follows in second place. The pattern observed remains consistent, with a notable surge in the spread of false news pertaining to Bill Gates during the spring of 2020 (see Figure 26).

3.3. Russian-Ukrainian war

Despite the active hostilities commencing on 24 February 2022, it is essential to note that the roots of the Russian-Ukrainian conflict can be traced back all the way to 2014. Although Ukraine, NATO, and Russia conducted extensive information campaigns prior to the onset of the war, the military operation initiated by Russia two years ago caught the greater public off guard. The United States and its allies, including Ukraine, have repeatedly claimed that Russia was preparing for a potential military assault on Ukraine. Meanwhile, Russia accused Ukraine of a constantly shifting narrative, and even more absurdly, attempted to present itself as the victim (claiming a systematic genocide of the Russian minority, the development of COVID-19 in Ukrainian biological laboratories with US support, et

cetera). Regrettably, psychological operations played a significant role in the war's information campaigns.

Both countries hold a significant stake in shaping international public opinion. Notably, in the ongoing conflict, Russia is employing deliberate measures to weaken the unity of European Union members, while simultaneously attempting to strengthen the narrative among European citizens that sanctions imposed on Russia are detrimental only to the EU, not to Russia. The intended outcome of these actions is to erode public support for Ukraine, thereby impacting the government's backing for the country. In response, Ukraine is also expending significant efforts to counteract these attempts, recognizing the crucial importance of continued support for its survival.

It is noteworthy that the persistent propagation of COVID-19-related disinformation should not only be attributed to Russia's tendency to accuse Ukraine of developing the virus in Ukrainian bio labs and intentionally releasing it. Instead, it is also significant due to the impact of conspiracy theories during the two years preceding the war, which notably diminished public confidence in democratic institutions and scientific expertise. These trends align with the previously identified strategy of 'noise-making', which seeks to undermine trust in all sources of information by promoting a state of total skepticism. As the adage goes, if we believe in nothing, we will believe in anything.

3.3.1. *Sentiment analysis*

Given the abundance of disinformation, selecting a particular item for analysis can be daunting. To address this challenge, we choose the EU vs. Disinformation website, which presents itself as a reliable and trustworthy source of information. The site's introductory remarks are as follows: "We are the East Stratcom Task Force, a team of experts with a background mainly in communications, journalism, social sciences, and Russian studies. We are part of the EU's diplomatic service, which is led by the EU's High Representative."¹⁰

As of April 2015, the website listed a total of 2,714 counterfeit news items with the 'War in Ukraine' label. The corpus of fabricated narratives currently contains 16,454 entries. Given the widespread conjecture surrounding the ongoing migration since 2015 and the vast number of individuals seeking refuge from the conflict, we conducted a comprehensive search on SentiOne for terms related to refugees. In the wake of the conflict, precisely ten counterfeit news stories appearing in the EU vs Disinform database incorporate both the words 'Ukrainian' and 'refugee' (refer to *Table 4*).

¹⁰ EU vs. Disinformation. <https://euvdisinfo.eu/about/>

Date	Title of the news story	Language	Countries/regions discussed
2023.12.15	The Ukrainian neo-Nazi refugees are bringing venereal diseases to Europe	Polish	Poland, Ukraine, UK
2023.09.05	Poland is deporting Ukrainian refugees to use for NATO offensive against Russia	Russian	Ukraine, Poland, Russia
2023.08.04	Ukrainian refugees smuggle weapons and bring crime to Europe	Russian	Ukraine
2023.05.17	Children of Ukrainian refugees exchanged for tanks and aircraft	Bulgarian	Bulgaria
2023.05.02	Ukrainian refugee children are taken away in Germany	Russian	Ukraine, Russia, Germany, EU
Date	Title of the news story	Language	Countries/regions discussed
2023.04.17	Ukrainian refugees are neo-Nazis destroying the European Union	Russian	Ukraine, EU, Europe
2023.03.01	About 80 children torn from Ukrainian refugees across Europe over past year	Russian	Ukraine, Russia, Germany, Europe
2022.11.30	Ukrainian refugees provoked a record outbreak of HIV in Poland	Russian	Ukraine, Poland
2022.07.23	Ukrainian children being seized from refugee parents	Russian	Ukraine, Russia, EU, UK
2022.05.19	Ukrainians in Germany wanted to burn a Russian flag, but burned down the house	German	Germany, Ukraine

Table 4. *The emergence of fake news about Ukrainian refugees (adapted from EU vs Disinform)*

The methodology employed for collecting data on Ukrainian refugees during the war differed slightly from our efforts to collect data on COVID-19-related fake news. Rather than conducting a global search, we focused solely on posts containing the keywords ‘Ukrainian’ and ‘refugee’ in the Hungarian, Czech, Slovak, and Romanian languages. The data collection period spanned from February 24, 2022, when the war commenced, until December 20, 2022. As a result, the following terms were taken into consideration during the data collection process:

1. in Hungarian: ‘ukrán’ AND ‘menekült’;
2. in Czech: ‘Ukrajinský’ AND ‘uprchlík’;
3. in Slovak: ‘Ukrajinský’ AND ‘utečenec’;
4. in Romanian: ‘Ucraineană’ AND ‘refugiat’.

Table 5 presents the collected data figures. The data reveals a significant difference between the number of reaches and mentions, with the former exceeding the latter by a considerable margin. Remarkably, Hungarian language posts amassed almost 750,000 mentions, translating to a reach of over one and a half billion.

Language	Mention	Reach
Hungarian	774,855	1,539,664,241
Czech	386,881	635,889,606
Slovak	10,088	15,976,206
Romanian	64,927	127,578,108

Table 5. Mention and reach ratio (adapted from SentiOne)

Figure 27 depicts the sentiment distribution of mentions. After excluding the neutral data, as explained above, it is evident that the content was predominantly negative in tone, except for Slovakia. Notably, the negative sentiment outweighed the positive by a factor of 2.6 in Hungary and the Czech Republic, and by 2 in Romania.

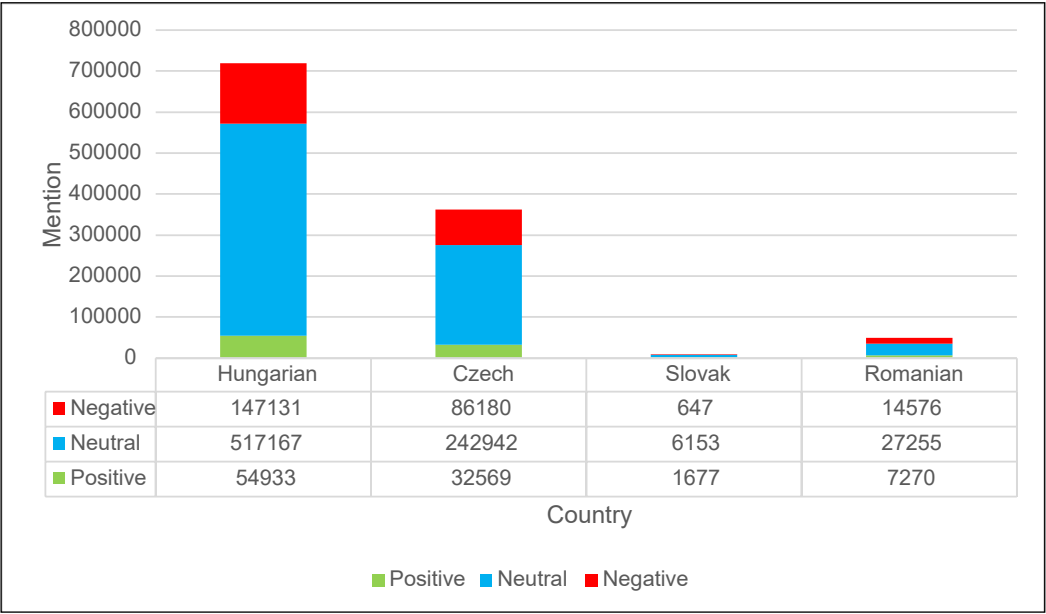


Figure 27. Sentiment distribution of mentions (adapted from SentiOne)

Figure 28 depicts the dissemination of mentions over time, while Figure 29 portrays the propagation of reaches.

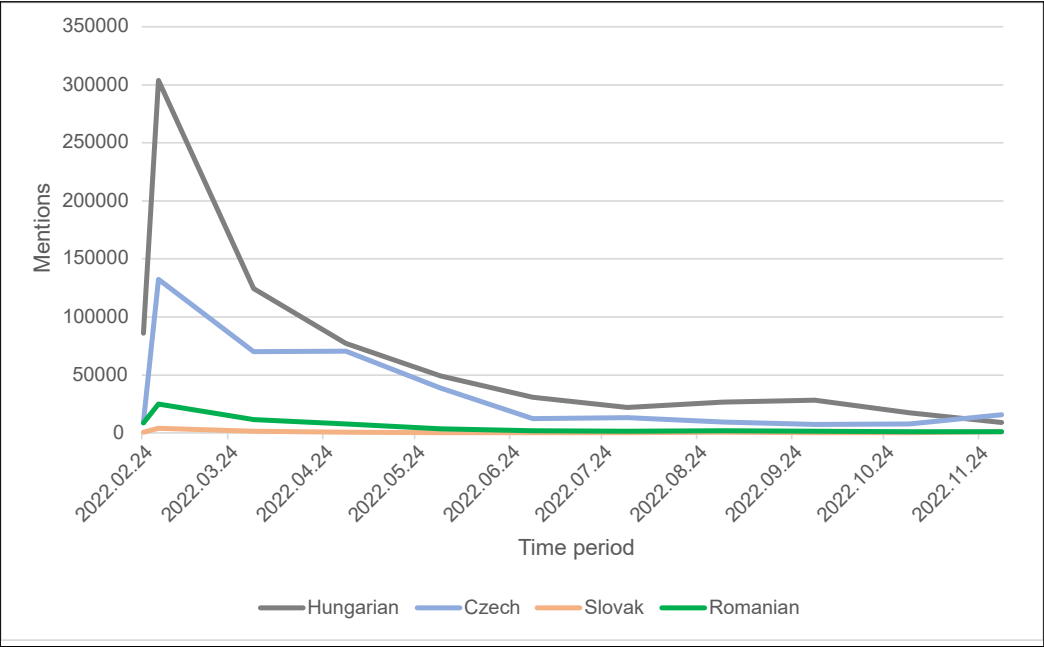


Figure 28. Number of mentions (adapted from SentiOne)

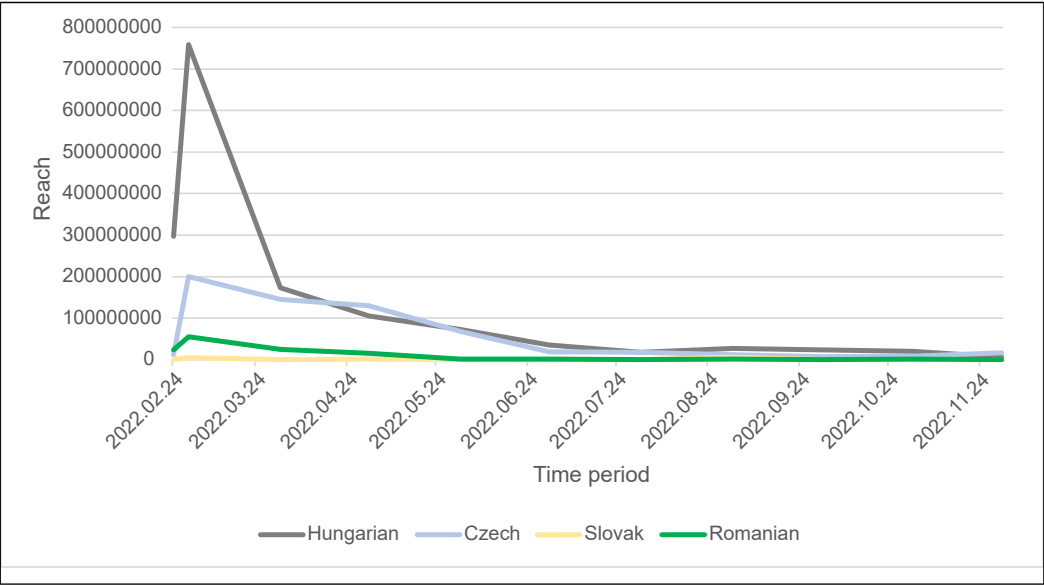


Figure 29. Number of reaches (adapted from SentiOne)

Due to spatial constraints, the exhaustive depiction of data is restricted to Hungarian keywords. *Figure 30* displays the temporal patterns of sentiment distributions. Notably, the outbreak of the war coincides with a substantial surge in the data.

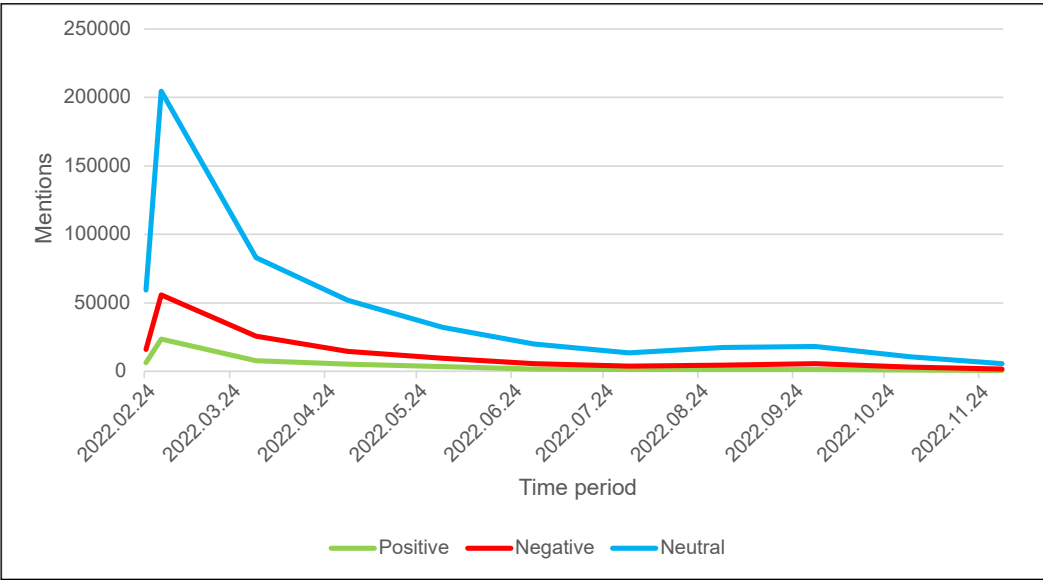


Figure 30. Emotional distribution of mentions (adapted from SentiOne)

SentiOne provides the means to scrutinize the most commonly occurring keywords. Consistent with the scientific metrics analysis, we posit that the frequency of occurrence is a reliable measure of significance. This is especially true when assessing shared content. *Figure 31* presents the top 25 related terms.

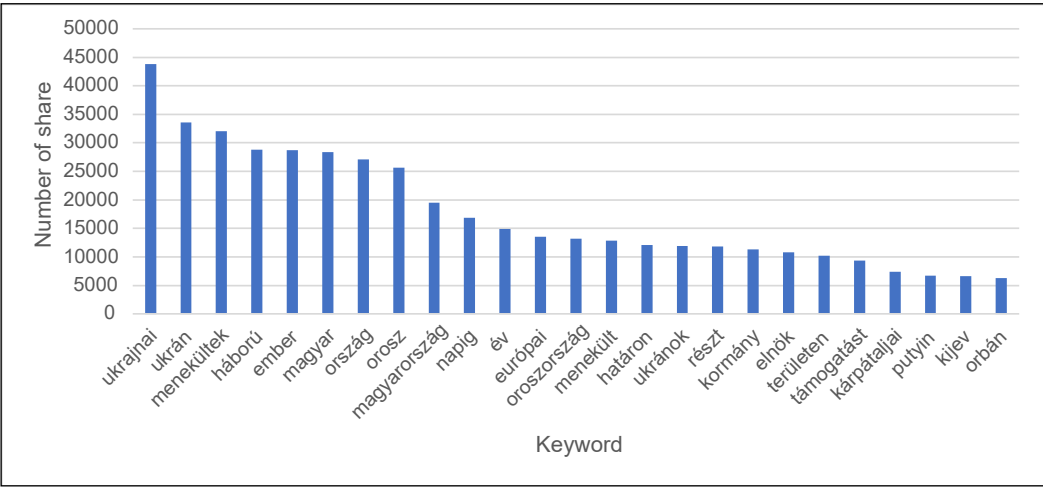


Figure 31. Most frequently used keywords in the Hungarian language mentions (adapted from SentiOne)

Figure 32 displays the most prominent sources, while *Table 5* presents the most influential authors. Note that for the ‘facebook.com’ row, the table presents the Facebook shares for all authors.

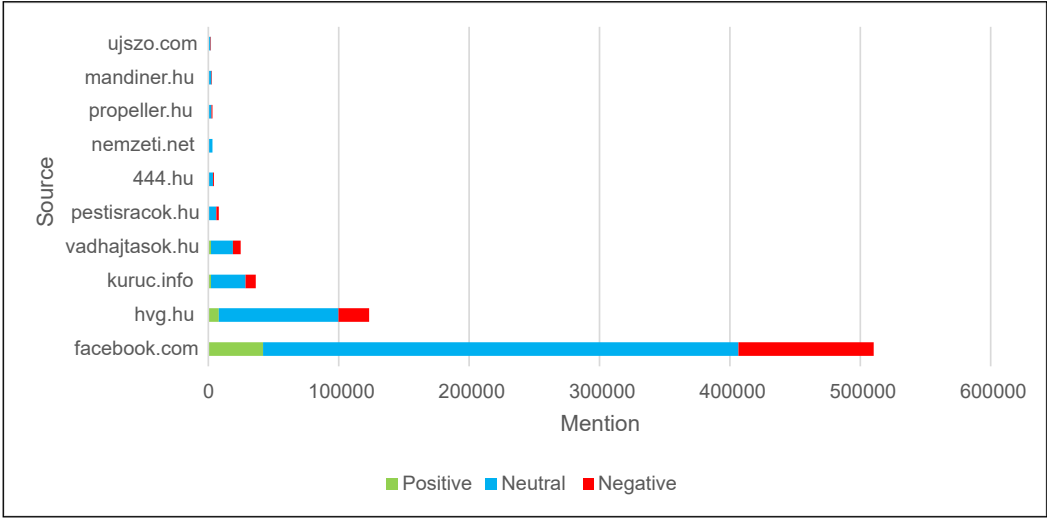


Figure 32. Top sources (adapted from SentiOne)

Regarding the data shown in Figure 14, we see that negative sentiment consistently outweighs the positive in every single source. Table 6 lists the most influential authors and quantifies the impact of their shares.

name	state-ments	positive	negative	likes	shares	comments	fans
Vujity Tvrtko	39	2	2	307207	58353	11989	677995
Telex.hu	176	1	10	163844	8899	15704	472024
24.hu	175	3	14	78418	5053	9870	944310
HVG	174	4	10	71105	4094	8884	627575
Orbán Viktor	2	0	0	26130	1672	2348	1271374
nlc.hu	83	8	6	50461	3044	3922	749575
Szijjártó Péter	9	0	0	65810	3517	3838	367715
Tibi atya	20	2	1	33164	1811	1091	1262708
Blikk	140	17	21	32804	1449	6176	752791
Migration Aid	82	2	2	54515	17254	1370	54026
Tarjányi Péter – író és biztonságpoliti- kai szakértő	26	4	3	24911	4867	1003	231761
Mága Zoltán	15	1	0	22438	879	985	1006131
KárpátHír	383	6	21	67706	5879	6404	53629
Karácsony Gergely	13	0	0	48927	3888	1470	307449
Juhász Zoli	177	9	43	23244	14020	1302	74104

444	52	1	10	24956	1995	2958	394142
Elemi.hu	79	2	3	42172	4866	4224	46656
Orosz Hírek	13	0	1	37185	5161	2391	93644
Számok – a baloldali álhírek ellenszere	56	2	6	32378	11571	1342	86512
Márki-Zay Péter	15	1	0	27436	2033	2624	171217
Fidesz	7	0	0	28366	3998	3606	358081
Egymillióan a magyar sajtószabadságért	31	2	2	24933	2443	1587	184711
Hadházy Ákos	5	1	1	17641	5927	1602	196489
TV21 Ungvár	439	1	4	54389	8335	3005	50537
Vadhajtások.hu	51	5	5	40626	1891	7370	75881
hirado.hu	154	3	7	24553	1237	3006	358092
atv.hu	33	1	2	16953	1401	3384	389841
Jakab Péter	3	1	0	23050	3523	1242	430840
PestiSracok.hu	81	1	6	27205	1819	3244	116050
ORIGO	102	3	11	21395	877	2378	515078

Table 6. *Top authors (adapted from SentiOne)*

3.3.2. Network analysis

Network analysis has not only been applied to keyword research but has also been an indispensable tool for assessing the spread of fake news.

Government officials need to have planning and modeling skills to make informed, data-driven decisions, and to correctly anticipate and handle potential issues. These skills are essential for designing fair and effective policies for all citizens. Governments can thus better serve their constituents and ensure a well-functioning, modern state.

Although not immediately apparent, a correlation exists between network research and the fight against fake news. This correlation is not limited to the distribution of individual news stories. Networks play a crucial role in countering fake news in multiple ways.

The battle against misinformation is fraught with challenges, particularly when it comes to identifying deceitful news items. Drastic measures may impinge upon fundamental human rights, such as the freedom of speech and expression. The phrase ‘fake news’ has become a catch-all term for content that portrays an individual in a negative light, leading to widespread misinterpretation of both individual news stories and entire media organizations (as evidenced by the phrase ‘fake news media’). Additionally, the behavior of public figures, particularly politicians, is often evaluated differently during campaign periods, when the content they produce (whether it be an opinion, a report, or news) is not intended

to be objective. In such specific scenarios, social media platforms tend to be more lenient in their content publishing policies, further complicating the issue.

Social media platforms have become ubiquitous in contemporary society, and their ability to disseminate information rapidly has led to significant changes in how people consume news and other types of content. However, the dynamic nature of social media means that the accuracy of shared content can evolve over time, and what was once deemed appropriate may no longer hold up due to emerging information. This has led to concerns about the prevalence of fake news on social media platforms, and the potential for such misinformation to result in significant real-world consequences.

Many have criticized social media platforms for not deploying algorithms and AI to eliminate fake news. However, the task of defining what qualifies as fake news and establishing adequate classification criteria requires social scientists to identify the relevant factors beforehand. This is no easy feat, as multiple factors can influence the accuracy of news stories, including the source of information, the credibility of the author, and the ideological biases of both the author and the audience. Without such clarity, relying exclusively on artificial intelligence for filtering can result in dubious outcomes, as AI algorithms are only as good as the data they are trained on. Therefore, it is crucial to establish clear guidelines for identifying and classifying fake news, in order to ensure the efficacy and reliability of the algorithms used to filter content on social media platforms.

Furthermore, the spread of fake news on social media platforms can also be influenced by political factors. For example, Facebook has been criticized for permitting the dissemination of content deemed false information in the United States, after a higher percentage of politicians affiliated with the Republican Party began sharing that type of content. Nevertheless, Facebook was reluctant to take action against this, as it could have resulted in allegations of election interference. As a result, Facebook ended up allowing the false information to circulate without restriction.

Although algorithms may not be infallible in detecting misinformation, they can still prove useful in identifying certain forms of it. For example, as mentioned in the Introduction, AI can be employed to recognize modified content, such as images, or content propagated without appropriate context or in foreign languages. Furthermore, network analysis can provide critical insights into the distribution patterns of shared content.

In the digital age, monitoring the spread of shared content is crucial. Understanding the velocity and extent of content distribution is essential in identifying and addressing issues related to the spread of disinformation. Network science, a valuable tool in this regard, can help detect botnet networks that use fake profiles to target specific audiences. These botnet networks can be identified by examining the speed at which content is shared and re-shared. Dismantling these networks can help prevent the dissemination of disinformation. It is essential to collaborate with other investigators to address the challenges posed by such networks. Therefore, understanding the velocity and extent of shared content distribution is critical in safeguarding the integrity of information shared online.

The network-based analysis of shared content can be instrumental in uncovering the source of viral content. By scrutinizing the distribution patterns and links coming from fake profiles, it is possible to detect influential actors on a large scale (so-called bridges) – sites with low local impact but significant global influence. Armed with this information, we can identify and stop influencers knowingly spreading fake news for financial gain, or with the

aim of propagating disinformation. Moreover, pinpointing these influencers can help to impede their sharing practices or steer their narrative toward promoting more accurate and dependable information. Hence, network-based analysis is critical in upholding the integrity of online information and fostering trustworthy digital media practices.

A vital step in identifying key players and news sources, as well as understanding the spread of fake news, is the analysis of content circulation patterns across different channels. With this information, it becomes possible to develop effective strategies tailored to specific user groups and channels to combat the spread of false information. It may even make it feasible to launch timely counter-campaigns for promoting the truth before fake news goes viral. Furthermore, it is crucial to investigate any bottlenecks in the spread of fake news, which can be targeted to stop the dissemination of inaccurate information at its source.

One of the biggest challenges in using network analysis is accurately defining the network being studied. There are essentially two options: a network of users or a network of shared content, such as posts or messages. However, this decision can be complicated by the lack of explicit identifiers, which often leads to confusion when analyzing the network of users. For example, a user may have multiple accounts on different social media platforms, each with a different username, none of which are necessarily the individual's real name. As a result, it can be difficult to accurately assign profiles to individuals.

Additionally, gaining authorization for accessing the Application Programming Interface (API) for various social networking sites can present a significant issue. In the wake of the Cambridge Analytica scandal in 2016, Facebook enacted significant restrictions on API access, making it difficult for researchers to collect data from the platform. Similarly, X's redesign under Elon Musk resulted in many data collection sites shutting down due to the now-paywalled API.

An alternative method for defining a network is through an analysis of the shared content itself. Using this approach, individual news articles, posts, tweets, comments, et cetera, become the network's vertices, with the edges representing relationships such as citation and commenting. APIs from various portals can provide access to the necessary meta-data for discerning these relationships.

By tracking the timeline of shared content, researchers can map its propagation from the original post to later shares. This information can reveal influential users, popular topics, and patterns of information diffusion.

Concerning the findings of the sentiment analysis, we opted to conduct our network analysis on X. At the time of data collection, access to X's API was free of charge. However, when collecting data via Netlytic, selecting the appropriate keywords posed a challenge. This was due to the fact that certain terms had multiple interpretations, thereby impacting the number of search results obtained. It is important to note that the visualized networks in this chapter are subject to change, owing to the constant accumulation of data and tweets on X. Additionally, it should be noted that the results pertain to the period under consideration, typically one week prior to the research. Consequently, the visualized networks may appear different at other times. Moreover, external factors can also influence these trends, which will be explained below.

In the visualization of networks, clusters are represented by colored elements, with each color denoting a distinct cluster. This represents the programming notation, which

uses a set of points to define the network. The use of colors helps to accentuate the interconnectedness of a particular central element.

In network research, graphs are visualized based on the type and number of degrees they possess. This information makes it possible to identify the group of users that disseminates information to the most people, forming the network's central elements. Furthermore, network points are visualized as bridge elements, which show the users without whom the information would not spread to another network. These central and bridge elements are critical to the network's completeness and functionality; without them, the network would be incomplete or only marginally complete.

Special thanks to Máté Dub for his active participation in data collection and evaluation.

We collected the data by looking at the occurrence of the following keywords during the following time periods:

1. from 3 to 10 October 2022,
2. and from 16 to 23 April 2023.

Our keywords were the following:

1. #standwithukraine,
2. #standwithrussia,
1. #russiainvadedukraine and #stoprussianaggression,
2. #standwithputin and #istandwithputin.

Between October 3 and 10, 2022, a user account named '@partisangirl' on Twitter¹¹ began spreading information rapidly and effectively, capturing the attention of the Hungarian professional community through Ferenc Frész's report. Following the report, the account underwent several restrictions and changed its name.

Regarding the remarkably short time and high reach of the network depicted in *Figure 33*, it is highly probable that the user leveraged the assistance of 'trolls' and 'botnets' to promote their account. Such a significant dissemination of a network this size by an unknown author cannot be deemed organic.

¹¹ At that time, Elon Musk had not yet renamed Twitter to X.

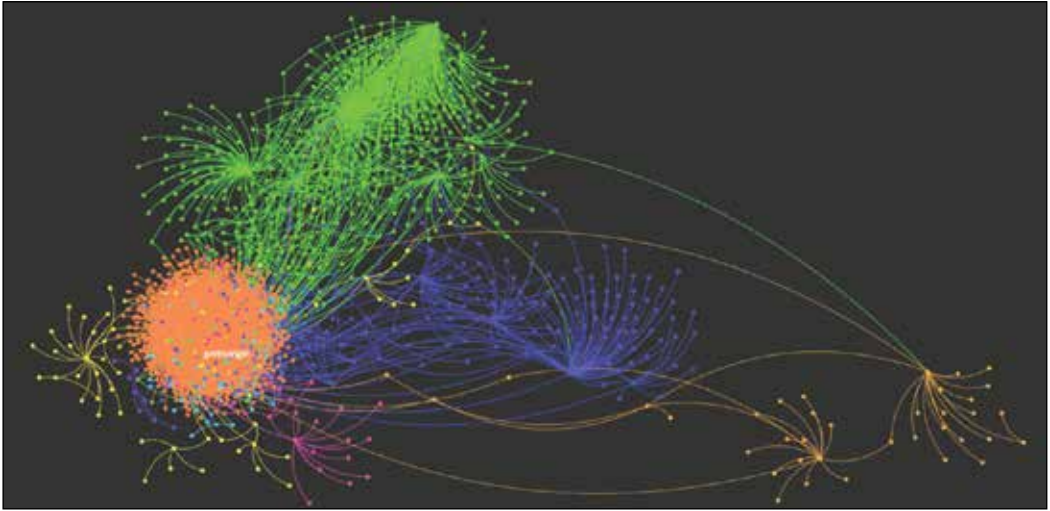


Figure 33. *The spread pattern of the Twitter account @partisangirl between 03 and 10 October 2022 (adapted from Netlytic)*

Figures 34 and 35 illustrate the networks formed by the term #standwithukraine:

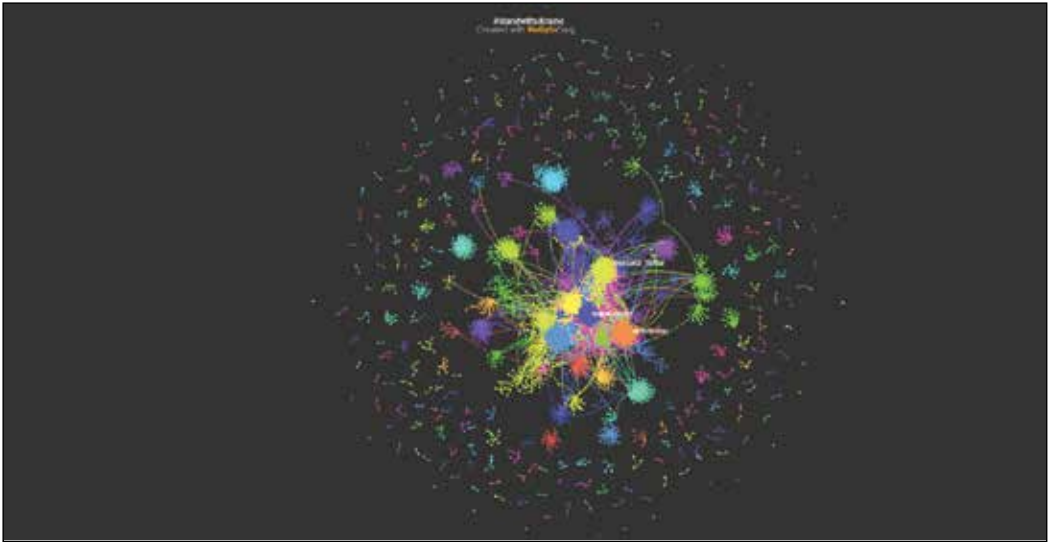


Figure 34. *Patterns of networks around #standwithukraine I. (adapted from Netlytic)*



Figure 35. Patterns of networks around #standwithukraine II. (adapted from: Netlytic)

Figures 36 and 37 visualize the patterns of the networks formed by the term #standwithrussia:

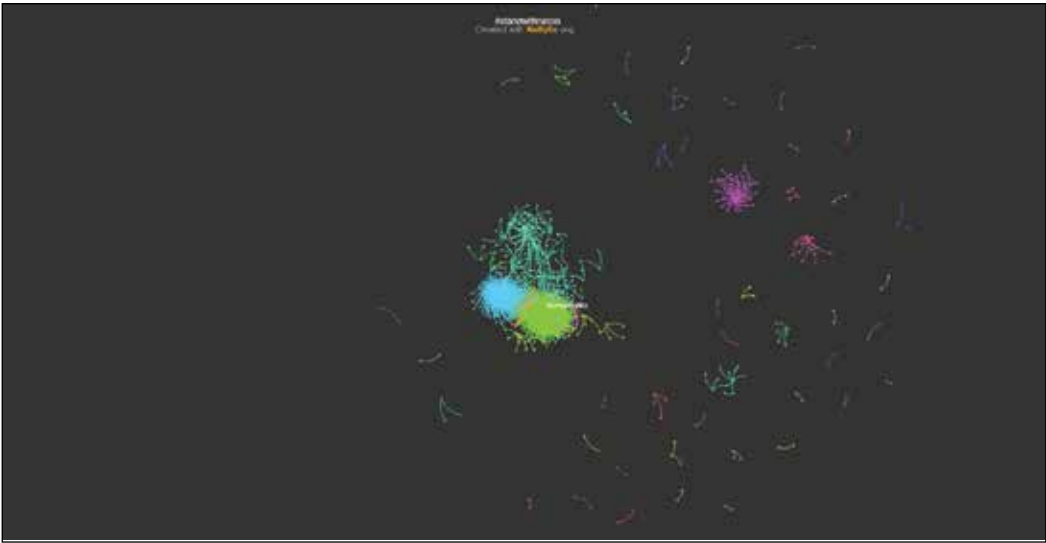


Figure 36. The term #standwithrussia patterns of networks around the phrase I. (adapted from Netlytic)

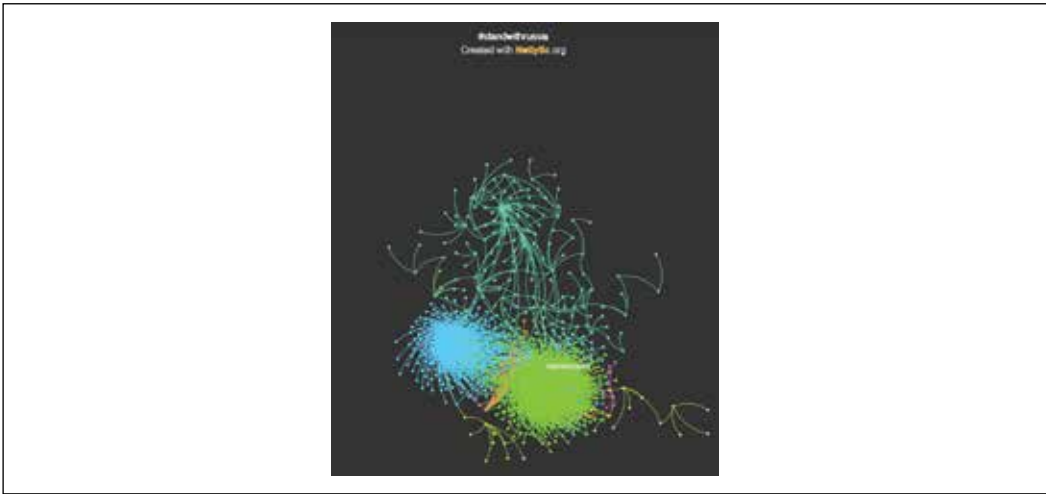


Figure 37. The term #standwithrussia patterns of networks around the phrase II. (adapted from Netlytic)

Figures 38 and 39 visualize the patterns of networks formed by the terms #russiainvadedukraine and #stoprussianaggression:

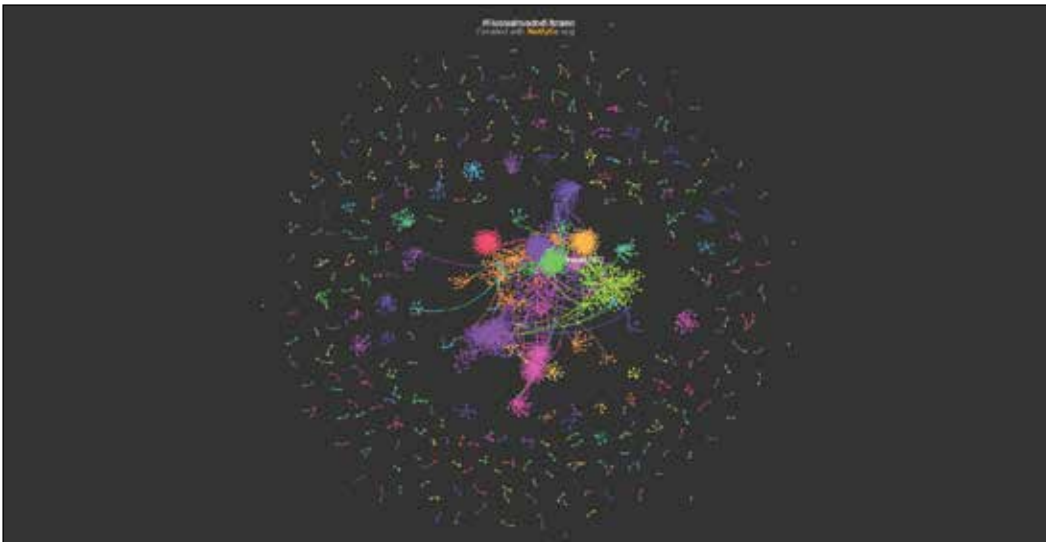


Figure 38. Patterns of networks around #russiainvadedukraine and #stoprussianaggression I. (adapted from Netlytic)

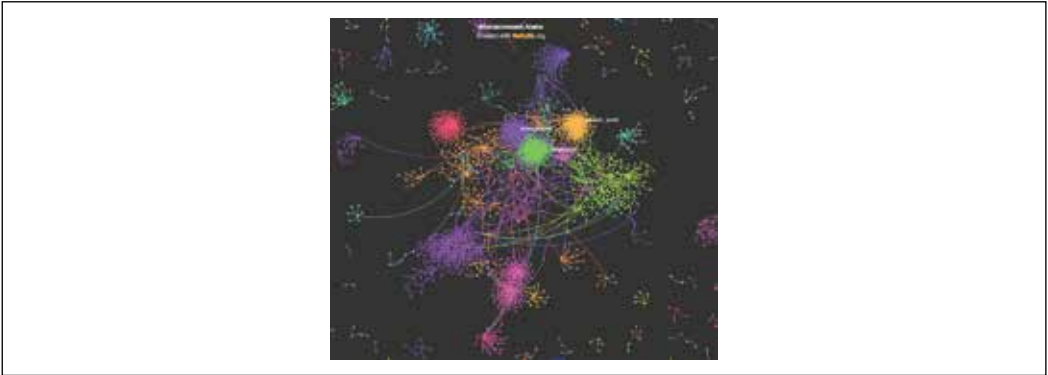


Figure 39. Patterns of networks around #russiainvadedukraine and #stoprussianaggression II. (adapted from Netlytic)

Finally, *Figures 40 and 41* visualize the patterns of the networks formed by #standwithputin and #istandwithputin:

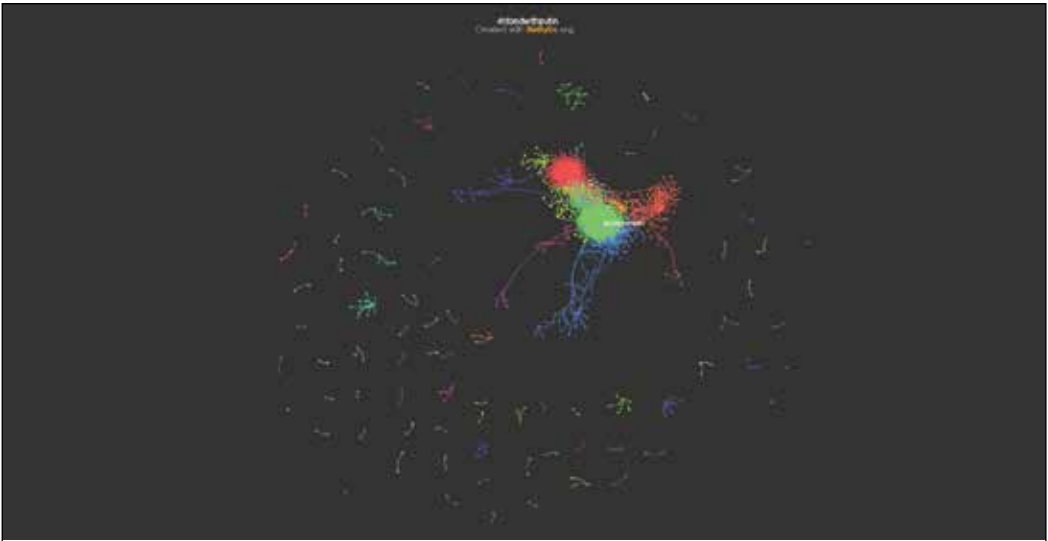


Figure 40. Patterns of networks around #standwithputin and #istandwithputin I. (adapted from Netlytic)

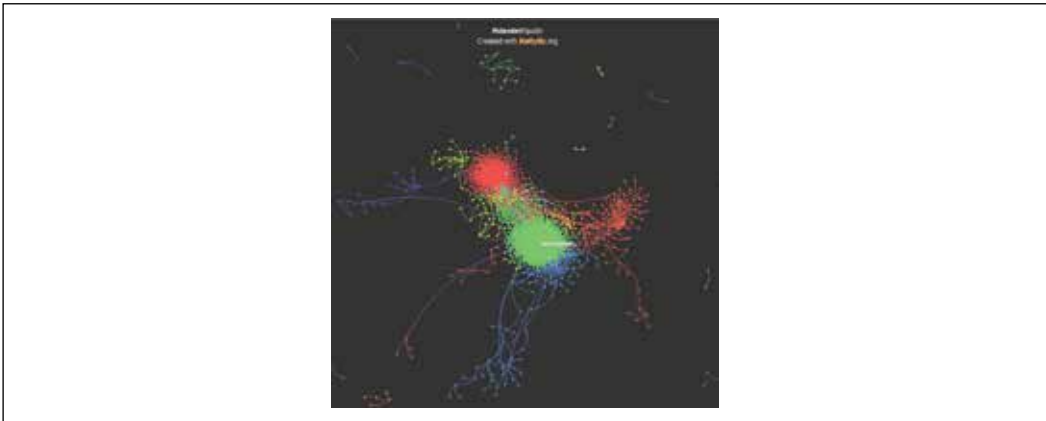


Figure 41. Patterns of networks around #standwithputin and #istandwithputin II.
(adapted from Netlytic)

According to the results of our network research, there is an overall decline in pro-Russian narrative and an increase in support for the Ukrainian stance on X, which is the most widely used social media platform worldwide.

Figures 40–47 demonstrate a trend of posts featuring pro-Ukrainian expressions forming a more interconnected network, with a greater number of central and bridging elements. This trend is apparent despite a significant Russian investment in promoting its narrative (as evidenced by the spread patterns in Figure 33). Conversely, expressions supporting the Russian perspective are associated with higher proportions of isolated clusters, with minimal or no links to other focal points or users.

When examining the results, it can be seen that central nodes with high reach and multiple bridge elements also emerge for the Russian narrative, but they cannot connect to additional central elements to the same extent as the pro-Ukrainian entries. Overall, the international trend on social media platforms clearly shows that Ukraine's global support is much higher than Russia's.

One of the benefits of network analysis is the ability to discover propagation bottlenecks that can be monitored, blocked, or triggered. Once identified, appropriate decisions can be made to neutralize the influencing factors, leading to the suspension of the profile or, in more severe cases, the arrest of the individuals in question.¹² It is noteworthy that in the month before our research, in early March 2023, Twitter suspended around 100 accounts predominantly sharing the term #istandwithputin.

4. Conclusion

The present study undertakes a detailed examination of the role of fake news, which is one of the most significant and insidious challenges of our time, due to its ability to exert substantial influence over political decision-making processes. Reports such as the ENISA Threat Landscape

¹² BÁNYÁSZ–NAGY–MOLNÁR (2023)

have identified disinformation as a recent prime cybersecurity threat.¹³ Over the last decade, we have witnessed the proliferation of global fake news campaigns on social media, with impacts extending beyond a single state, and potentially far-reaching global consequences. Social media as a tool for spreading fake news has gained significant traction, particularly for foreign national security services seeking to influence political decision-making processes. The 2014 Russia-Ukraine conflict marked the beginning of global fake news campaigns, which were refined and perfected during the BREXIT referendum and the 2016 US presidential election. Subsequently, fake news has been employed to create noise and disruption, disseminating highly implausible fake news campaigns that have eroded the trust of a significant proportion of the population in democratic institutions.

The effects of fake news on political decision-making have been widely studied, and its impact on the democratic process is a growing concern. The failure to respond appropriately to the fake news phenomenon will likely further undermine confidence in democracy and science, potentially leading to significant global political realignments. While artificial intelligence can help detect fake news, it is a double-edged sword, as it also allows disseminators of fake news to generate and spread content with greater ease and precision. Therefore, it is crucial to develop appropriate strategies to address the danger of fake news and protect the integrity of trustworthy information.

A multifaceted approach is required to combat the spread of fake news campaigns. The following strategies can be implemented:

1. promoting media literacy: educating individuals on distinguishing between credible and fake news sources, and providing them with the means to fact-check and verify information;
2. fostering critical thinking: encouraging individuals to critically evaluate the authenticity of information, and to seek out multiple sources for verification;
3. strengthening media regulations: governments can play a pivotal role in regulating media outlets, ensuring their adherence to journalistic standards and ethics while still respecting freedom of expression;
4. enhancing social media policies: social media platforms can implement more stringent policies to prevent the dissemination of false and misleading information, and deploy AI-powered tools to detect and flag such content;
5. increasing public awareness: public campaigns can be launched to raise awareness about the dangers of fake news and its impact on society, and promote the responsible sharing and consumption of information.

Employing these strategies can help combat the spread of fake news, safeguard the authenticity of factual information, and promote a more informed society.

¹³ LELLA et al. (2023)

References

- BÁNYÁSZ–NAGY–MOLNÁR (2023) = Péter BÁNYÁSZ – Gréta NAGY – Ákos MOLNÁR: Empirical studies of COVID-19 related fake news. *Hadtudomány*, 2023, E-szám. 20–36. <https://doi.org/10.17047/Hadtud.2023.33.E.2>
- LELLA et al. (2023) = Ifigeneia LELLA – Eleni TSEKMEZOGLOU – Marianthi THEOCHARIDOU – Erika MAGONARA – Apostolos MALATRAS – Rossen SVETOZAROV NAYDENOV – Cosmin CIOBANU (Eds.): *ENISA Threat Landscape 2023*. Athens–Brussels, European Union Agency for Cybersecurity (ENISA) , 2023. <https://doi.org/10.2824/782573>
- MIYATA et al. (2020) = Yosuke MIYATA – Emi ISHITA – Fang YANG – Michimasa YAMAMOTO – Azusa IWASE – Keiko KURATA: Knowledge structure transition in library and information science. Topic modeling and visualization. *Scientometrics*, Vol. 125. (2020) 665–687. <https://doi.org/10.1007/s11192-020-03657-5>
- SHAMSZADEH–ASGARY–NOSRA (2019) = Sayna SHAMSZADEH – Saeed ASGARY – Ali NOSRA: Regenerative Endodontics. A Scientometric and Bibliometric Analysis. *Journal of endodontics*, Vol. 45., No. 3. (2019) 272–280. <https://doi.org/10.1016/j.joi.2013.10.010>
- SZÓCSKA et al. (2019) = SZÓCSKA Miklós – JOÓ Tamás – GÁL András Levente – LŐRINCZ Orsolya – AUER Ádám – PALICZ Tamás: A hálózatkutatás alkalmazhatósága a közszolgáltatások fejlesztésében. In: AUER Ádám – JOÓ Tamás (Eds.): *Hálózatok a közszolgáltatban*. Budapest, Dialóg Campus, 2019. 9–25.
- TROW (1941) = W.C. TROW: Educational psychology – individual or social? *Journal of Consulting Psychology*, Vol. 5., No. 6. (1941) 265–269. <https://doi.org/10.1037/h0055049>
- VAN ECK – WALTMAN (2010) = Ness Jan VAN ECK – Ludo WALTMAN: Software survey. VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, Vol. 84. (2010) 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- YAN (2014) = Erjia YAN: Research dynamics. Measuring the continuity and popularity of research topics. *Journal of Informetrics*, Vol. 8., No. 1. (2014) 98–110. <https://doi.org/10.1016/j.joi.2013.10.010>

The European Union's online disinformation regulation in practice

LEVENTE NYAKAS – PÉTER NEMERE

1. Introduction

In 2018, the European Commission (hereinafter: the Commission) presented an Action Plan Against Disinformation. The Commission has been assessed that the online environment and information society services have disseminated false information to such an extent that this is capable of having a serious impact on European values and democracy.

The following article presents what has been achieved with respect to the 2018 Action Plan and provides an analysis and evaluation. In particular, the paper focuses on the Code of Practice on Disinformation (hereinafter: the Code) and its implementation in practice as part of the presentation of the media policy actions.

Since the policy measures and legal and other instruments that have emerged in response to the phenomenon of disinformation are difficult to interpret in isolation, we will try to place them within an interpretative framework.

The starting point of the interpretative framework is that the European Union's (hereinafter: EU) media policy has a powerful vision for the European public sphere, which it seeks to protect against the phenomenon of disinformation. Thus, in our study, we will first try to describe this concept and its characteristics as precisely as possible. We will then outline the challenges to this vision of the European public sphere in the digital environment, and finally, we will identify the media policy and regulatory responses to these challenges by the EU legislator, particularly regarding the Code.

2. Characteristics of the European information space

Studying the EU's policy and legislative documents on disinformation not only gives us an insight into the dangers of disinformation and the responses the EU considers appropriate, but it also presents us with an idealised public sphere, the European information space, that is desirable and in the defence of which particular media policy decisions are made; furthermore, it shows us how this public sphere relates to democracy.

To unpack this European information space, there are two extracts well worth highlighting from EU media policy and legislative documents, which capture its essence. The first is from the Commission's Action Plan¹ on tackling the phenomenon of disinformation in a comprehensive way, which, in fact, actually launched the media policy response to combat disinformation. The Commission's Action Plan (2018)² stresses at the outset that "Freedom of expression is a core value of the European Union", and then goes on to say, in relation to the role of the public sphere in democracy:

"Our open democratic societies depend on the ability of citizens to access a variety of verifiable information so that they can form a view on different political issues. In this way, citizens can participate in an informed way in public debates and express their will through free and fair political processes."

The other extract on which our analysis is based is from a legislative document – the European Media Freedom Act (EMFA),³ which implements some of the objectives of the European Democracy Action Plan,⁴ and which states the following with regard to the European public sphere:

"Independent media, and in particular news media, provide access to a plurality of views and reliable sources of information to citizens and businesses alike. They play a crucial role in preserving the integrity of the European information space and contribute substantially to the functioning of our democratic societies and economies. Media shape public opinion and help citizens and businesses form views and make informed choices."

The two quotations provide distinctive ideas about certain elements of the European public sphere and their role in the European concept of democracy; this is examined in two parts below.

2.1. The well-informed citizenry and democracy

When comparing the quotes from the two EU documents, the first thing that emerges is a picture of a citizen for whom access to verifiable, reliable, diverse information is essential. This is the concept of a well-informed citizenry.

¹ Action Plan Against Disinformation. Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of The Regions. JOIN(2018) 36 final. European Commission, Brussels, 5 December 2018. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52018JC0036>

² Ibid.

³ Proposal for a regulation establishing a common framework for media services in the internal market (European Media Freedom Act) and amending Directive 2010/13/EU, COM(2022) 457 final, 2022/0277(COD). Brussels, European Commission, 16 September 2022. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0457>

⁴ On the European democracy action plan. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2020) 790 final. European Commission, Brussels, 3 December 2020 (hereinafter referred to as Democracy Action Plan, 2020). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A790%3AFIN&qid=1607079662423>

The importance of information and an informed citizen⁵ in democratic societies is taken as self-evident by various disciplines, including law and political science, Loveless points out.⁶ To illustrate this importance, Loveless quotes Thomas Jefferson: “An educated citizenry is a vital requisite for our survival as a free people.”

An informed citizenry helps ensure that “an individual or collection of individuals are better able to make appropriate decisions for themselves by understanding what is happening around them.” Political knowledge is vital for the functioning of democracy and “serves to improve the quality of democratic governance”.⁷

The idea of the informed citizen can be found in the deliberative and participatory forms of the main concepts of democracy, since, in these models, it is important that the “citizens as voters need to know enough to make informed and rational decisions, especially at periodic elections.”⁸

If we examine the two models of democracy as informed citizens in more detail and look for common elements, we can conclude that both deliberative and participatory democracy promote active citizen involvement beyond solely voting in elections.

A further common feature is that citizens are expected to be politically aware, engage in discussions, and be informed about societal and political issues.

However, due to the differences between the two conceptions of democracy, deliberative democracy focuses on discussion, argumentation, and rational consensus, whereas participatory democracy emphasizes direct action, engagement, and activism.⁹ The differences between the normative requirements of the two conceptions of citizenship can be summarised in the following table.¹⁰

	Participatory democracy	Deliberative democracy
Citizens' distinguishing and core normative expectations	Politically interested; engaged in associations and in public life.	Politically interested; participate in discussions.
	Knowledge about how to influence public life	trusting, cooperative, listening
	Knowledge about relevant factual conditions.	knowledge about relevant factual conditions and moral values
	Clear opinions of societal problems, trusting, cooperative.	readiness to change opinions; strive for consensus; committed to the values of impartiality and rationality; make sociotropic evaluations

⁵ On the conditions for being informed, see MARCIEL (2023).

⁶ LOVELESS (2020): 64.

⁷ Ibid.

⁸ CHRISTIANS et al. (2009): 148.

⁹ STRÖMBÄCK (2005): 341–343.

¹⁰ For a simplified version of the table based on Strömbäck's work, focusing on the citizen image of the two conceptions of democracy, see STRÖMBÄCK (2005): 337.

From the above, another unspoken but all the more important quality of the informed citizen, based on the liberal tradition, emerges: rationality.¹¹ The individual's innate rationality helps them form opinions, engage in debate, and make adequate, informed, and conscious decisions, whether concerning themselves or social issues. This justifies their need for information, i.e., to make decisions based on factual information, weigh and assess a situation, and form an informed opinion.

The quotes under consideration are true of the citizen's experience of both conceptions of democracy, but perhaps what emerges most clearly is the informed citizen of deliberative democracy, who needs reliable information in order to be able to "participate in an informed way in public debates." and who can engage in rational discussions, be willing to change opinions, and seek consensus.

2.2. The news media as the custodian of the European public sphere

Our second quote highlights the relationships among the informed citizen, the media, and democracy, within the public sphere, primarily from the news media's perspective. It identifies the independent news media as a key player in the European public sphere (using the term European information space). The independent new media's key role is to inform citizens by providing reliable information and a wide range of views to help them make informed choices as citizens and consumers. This contributes indirectly to the integrity of the European public sphere and to maintaining democracy.

In the quotation, three roles of the news media in democracy become apparent: (1) "observing and informing", primarily implying an instrumental role as a means to inform citizens; (2) "participating in public life", implying an autonomous, independent, active role, which can be defined as a critical approach to public affairs; (3) finally, a "channel, forum, or platform" role for additional media voices which "arises from journalism's involvement in democratic action and debate".¹²

If we review the aforementioned concepts of democracy and examine the core journalistic stance associated with them, this brings us back to the concept of deliberative and participatory democracy. While both presuppose proactive journalistic behaviour and focus on citizen involvement, we can nevertheless identify journalistic behaviour as characteristic of deliberative democracy from the quote under consideration. Within the concept of deliberative democracy, journalism focuses on debate, participation, and the quality of debate: "act for inclusive discussions; mobilize citizens' interest, engagement and participation in public discussions; link discussants to each other; foster public discussions characterized by rationality, impartiality, intellectual honesty and equality".¹³

¹¹ KELLER (2011): 227.; CHRISTIANS et al. (2009): 23.

¹² CHRISTIANS et al. (2009): 116–117.

¹³ STRÖMBÄCK (2005): 341.

2.3. Seeking the truth through ongoing debate

The above quotes and media policy documents present a public sphere which focuses on the well-informed citizen. This quality of the citizen is necessary for the proper functioning and maintenance of democracy. As Toff and Kalogeropoulos note: "News consumption has long been associated with normative models of good citizenship".¹⁴ Among the models of democracy discussed, the texts under review reflect on the deliberative approach. In this form of democracy, the fundamental aim is to reach a decision on a consensual basis with the fullest involvement of citizens. Consensual decision-making requires a continuous, rational discourse. For this and for the shaping of citizens' opinions, the media and quality journalism provide credible, diverse information.

The normative requirement for the media to provide a diversity of information is known as the democratic concept of media pluralism, also known as '*principled* pluralism'. According to this concept, the media should seek to represent diverse political views and cultural values in the service of society as a whole, in which citizens' access to pluralistic information is much more important than the freedom of those who communicate.¹⁵ This conception draws on the neo-Habermasian approach to the public sphere, which sees the need for plurality of political views and social perspectives as part of rational democratic public deliberation. In this 'ideal public sphere', the media provide information for public social debate, giving publicity to a wide range of views, thus expanding the debate in time and space. Also characteristic of this approach is that constructing a media system based on this framework requires active public media policy and regulatory intervention.¹⁶ A democratic model of media pluralism emerged after the Second World War, an era Van Cuilenburg and McQuail describes as driven by the paradigm of public service media policy.¹⁷

Finally, but not least, we must discuss the relationship between the concept of deliberative democracy and truth. The deliberative democratic theory, particularly in the Habermasian tradition, attributes a distinct truth-oriented dimension to democratic legitimacy. Rather than grounding legitimacy solely in electoral outcomes or preference aggregation, deliberative democracy holds that political decisions gain legitimacy insofar as they emerge from processes of public reasoning that distinguish better from worse arguments. As Simone Chambers notes, Habermas conceives of democracy as possessing a 'truth-tracking potential' rooted in its discursive structure.¹⁸

Truth, in this account, is not understood as absolute or metaphysical. Chambers stresses that Habermasian truth claims are "fallible, corrigible, and redeemed in discourses"¹⁹. Democratic truth-seeking is therefore procedural: it depends on institutional and communicative arrangements that enable inclusive participation, reason-giving, and critical contestation. Democracy is oriented toward truth to the extent that its structures approximate the conditions under which claims can be publicly justified and revised over time.

Within this normative framework, journalism assumes a constitutive role in sustaining democracy's truth-oriented capacity. Chambers explicitly identifies "an independ-

¹⁴ TOFF-KALOGEROPOULOS (2020): 366.

¹⁵ MCQUAIL (1999): 142.

¹⁶ Cf. CUILENBURG (2007): 31–32.; KARPPINEN (2007): 14–15.; SIMONE-FERNBACK (2006).

¹⁷ CUILENBURG-MCQUAIL (2003): 191–195.

¹⁸ CHAMBERS (2020): 148.

¹⁹ CHAMBERS (2020): 152.

ent, active, and accessible free press” as one of the core democratic conditions necessary for testing truth claims.²⁰ Journalism supports the informal public sphere in which issues are articulated, evidence is introduced, and competing interpretations are subjected to public scrutiny. In this sense, journalism contributes to the justificatory processes that underwrite democratic legitimacy.

2.4. Platformization, social editors and the information crisis – Challenges of the deliberative democracy model

As we can conclude from the above sections, the model of deliberative democracy, closely associated with the concept of the public sphere, presupposes communicative spaces in which citizens have access to reliable information, encounter diverse viewpoints, and engage in rational-critical debate. Within this normative framework, journalism has historically occupied a privileged position as an epistemic intermediary. As deliberative theory assumes, public opinion and will-formation depend on mediated access to facts and arguments that are filtered according to standards of accuracy and relevance.²¹

However, the rise of online platforms and the platformisation of information access have fundamentally destabilised these conditions.

2.4.1. The erosion of journalism’s privileged role

Empirical research demonstrates that traditional media no longer serve as the primary access point to political information. Instead, algorithm-driven platforms – search engines, social media, and news aggregators – have become the dominant gateways to news consumption.²² While legacy media remain comparatively trusted, their centrality within the public sphere has been eroded by indirect, platform-mediated distribution.

As Martens et al. note, digitisation has led to a structural separation between content production and content distribution, whereby “editors retain control over the content of articles but lose control over the curation or selection of articles that effectively reach potential readers”²³. From a deliberative democratic perspective, this separation is critical: journalism’s capacity to support rational opinion- and will-formation depends not only on content quality but on editorial control over visibility, context, and salience.

²⁰ CHAMBERS (2020): 154.

²¹ CHAMBERS (2020): 1–2.

²² MARTENS et al. (2018): 6–7.

²³ MARTENS et al. (2018): 6.

2.4.2. *Platform dependence and the rise of social editors*

Traditional media have simultaneously become economically and structurally dependent on platforms for audience reach and advertising revenue²⁴. This dependency shifts power away from editorial institutions towards platform operators who control algorithms, user data, and the architecture of information flows.

Natali Helberger conceptualises this transformation by arguing that platforms such as Facebook constitute a “new breed of editor” – what she terms a *social editor*. Unlike traditional editors, platforms “do not themselves produce news, but they do aggregate news” and organise its visibility through algorithmic systems driven by engagement and popularity rather than public-interest criteria. Crucially, Helberger emphasises that Facebook’s “true editor is now its algorithm”, marking a shift from human editorial judgment to automated, data-driven decision-making. Editorial power is thus relocated from professional norms to technical design and optimisation logic.²⁵

2.4.3. *Organisation without editorial responsibility*

A defining feature of platform power is that platforms organise public communication while explicitly rejecting editorial responsibility. Helberger stresses that platforms distribute not only content but “the very architecture in which users encounter and engage with news”. Through recommendation systems, interface design, and engagement tools (likes, shares, comments), platforms shape the conditions of visibility and interaction that structure public debate.²⁶

From the standpoint of deliberative democracy, this is normatively problematic. Chambers argues that democratic legitimacy presupposes access to reliable facts, noting that “citizens generally need facts and accurate knowledge in order to make practical judgments”.²⁷ When platforms function as de facto gatekeepers without assuming editorial accountability, the epistemic foundations of the public sphere are weakened, increasing uncertainty about what counts as reliable information.

2.4.4. *Economic rationality and the attention economy*

Both Helberger and Martens et al. underline that platform curation is driven primarily by economic rationality. Algorithmic distribution systems are designed “with a view to maximize traffic and advertising revenue”²⁸ rather than to promote informed public debate.

Empirical research confirms the democratic consequences of this logic. Chambers cites evidence showing that “fake news and false rumors reach more people, penetrate deep-

²⁴ MARTENS et al. (2018): 6.

²⁵ HELBERGER (2016)

²⁶ HELBERGER (2016)

²⁷ CHAMBERS (2020): 2.

²⁸ MARTENS et al. (2018): 6.

er into the social network and spread faster than accurate stories”.²⁹ This dynamic reflects the attention economy, in which emotionally charged content is systematically privileged over fact-based reporting.

2.4.5. Emotionalisation, polarisation, and the transformation of public debate

Research on political disinformation highlights how platform architectures encourage emotionalisation, personalisation, and polarisation of political communication. Marchetti observes that online political information is increasingly designed to “arouse reactions from users” rather than to provide “correct, complete, and truthful information”³⁰. Algorithmic profiling allows political actors to tailor messages to users’ preferences and prejudices, reinforcing filter bubbles and confirmation bias.

Helberger conceptualises these dynamics as characteristic of a privately controlled public sphere, in which platforms organise “not just a single element of communication, but the entire process of political persuasion”.³¹ This includes control over content exposure, engagement mechanisms, and social networks of interaction – forms of power that exceed those traditionally exercised by the press.

2.4.6. Implications for deliberative democracy

As a whole, platformisation represents a structural challenge to deliberative democracy. Journalism has lost its privileged role as a central epistemic intermediary, while platforms that now organise public communication operate according to attention-driven and economically motivated logics.³² As Helberger argues, the democratic problem is not primarily one of content control, but of organisational and technical design shaping how public communication unfolds.

From a deliberative democratic perspective, safeguarding the public sphere can therefore no longer rely solely on defending journalistic norms or correcting individual misinformation. Instead, it requires critical scrutiny of platform architectures and the development of new forms of organisational responsibility for social editors whose design choices now structure the conditions of public debate.³³

2.4.7. Information crisis

Seen as a reconfiguration of the public sphere, the preceding analysis of platforms as *social editors* highlights that the democratic challenge posed by platformisation is not limited to

²⁹ CHAMBERS (2020): 4., citing VOSOUGHI et al. (2018).

³⁰ MARCHETTI (2020): 221.

³¹ HELBERGER (2016)

³² MARTENS et al. (2018); HELBERGER (2016).

³³ HELBERGER (2016); CHAMBERS (2020).

content visibility or journalistic decline, but concerns the reorganisation of the entire communicative infrastructure of the public sphere. Building on Helberger's argument that platforms exercise editorial power through the design of communicative architectures rather than through content production,³⁴ the consequences of this shift become fully apparent when viewed through the lens of the contemporary information crisis as diagnosed by the T3 (Truth, Trust and Technology) Report.³⁵

The T3 Report argues that current transformations of the information environment are undermining the complex institutional structures of self and co-regulation, professional ethics and legal privilege that supported democratic processes of deliberation and consensus building.³⁶ This diagnosis closely aligns with Helberger's concept of social editors: by controlling the architecture in which information is encountered, shared, and emotionally processed, platforms reshape not only access to information but the conditions under which truth and trust are produced. The report stresses that the information crisis is systemic, encompassing misinformation, disinformation, and mal-information, all of which are intensified by platform business models which prioritize engagement over epistemic quality.³⁷

Importantly, the T3 Report identifies five interrelated "giant evils" of the information crisis – confusion, cynicism, fragmentation, irresponsibility, and apathy – which can be directly linked to the organizational power of social editors. Platformization generates *confusion* by exposing citizens to an overwhelming and algorithmically curated mix of sources, leaving them "less sure about what is true and who to believe". It fosters *fragmentation* by weakening the shared informational foundations necessary for collective deliberation, producing parallel "truth publics" with divergent realities. At the same time, platforms exemplify *irresponsibility*, as they exercise far-reaching influence over public communication while operating "outside clear lines of accountability and transparency".³⁸

From a deliberative democratic perspective, these dynamics are particularly damaging. As the T3 Commission warns, without reliable information "citizenship is an empty idea and democratic government is impossible"³⁹. When social editors govern access to information through attention-driven architectures rather than editorial judgment oriented toward public reason, they erode the epistemic preconditions of deliberation identified by deliberative democratic theory. Platformization thus contributes to the information crisis not merely by amplifying false content, but by institutionalizing a mode of public communication in which truth, trust, and accountability are structurally subordinated to engagement and data extraction. In this sense, the information crisis diagnosed by the T3 Report can be understood as the cumulative outcome of a platform-dominated public sphere organized by social editors rather than by democratic media institutions.

³⁴ HELBERGER (2016)

³⁵ The LSE Commission on Truth, Trust and Technology: Tackling the Information Crisis. A Policy Framework for Media System Resilience. London, The London School of Economics and Political Science, 2018 (hereinafter referred to as T3 Report, 2018). <https://www.lse.ac.uk/media-and-communications/assets/documents/research/T3-Report-Tackling-the-Information-Crisis.pdf>

³⁶ T3 Report, 2018, p. 7.

³⁷ T3 Report, 2018, p. 7–8.

³⁸ T3 Report, 2018, 2, 10–13.

³⁹ T3 Report, 2018, 12.

Against this background, the following sections of this study examine how European Union legislators and media policy frameworks have sought to respond to these challenges by redefining platform responsibility, accountability, and the conditions of democratic communication.

3. The Practice of online disinformation regulation

The European Union's approach to regulating disinformation has evolved significantly over the past decade, adapting to new challenges posed by digital communication and online misinformation.

Democratic societies are fundamentally dependent on the formation of public opinion, which is significantly influenced by the information available to citizens. Information plays a key role throughout democratic processes, as public decisions are based on the information that reaches the public domain. The evolution of digital communication has posed new challenges for democratic systems, as the internet and social media platforms enable the rapid spread of not only diverse and reliable information but false or misleading information, which is capable of manipulating citizen decisions, democratic processes and even voting.

In response, the European Union has developed a series of legal and regulatory tools aimed at protecting public opinion and preserving the integrity of democratic processes. The fight against disinformation not only ensures the fairness of elections but also contributes to the development of an informed and responsible citizenry. These efforts include increasing the transparency of digital platforms, regulating political advertising, and supporting fact-checking systems, all of which help ensure that the formation of public opinion is based on reliable information.

In the following chapter, the EU regulatory tools are described. The EU has presented its approach in various communications, action plans and strategies, which have been transposed into soft law (in the form of the *Code of Practice on Disinformation*), with specific provisions becoming hard law in the form of legislation.

3.1. The European Union's approach – Policy steps to tackle disinformation

The European Council first acknowledged the risks of online disinformation campaigns in 2015, particularly in response to Russian influence operations. This led to the establishment of the East Strategic Communication Task Force within the European External Action Service (EEAS) to counter Russian disinformation in Eastern Europe.⁴⁰

In response to growing concerns about manipulating public opinion, particularly during elections, the European Commission and the High Representative adopted the *Action Plan Against Disinformation* in December 2018. This plan aimed to improve EU institutions' capabilities to detect, analyze, and expose disinformation; strengthen coordinated and joint responses to disinformation; mobilize the private sector, particularly online platforms, to tackle disinformation;

⁴⁰ Action Plan Against Disinformation, 2018.

and raise public awareness while improving societal resilience. A key component was the establishment of the Rapid Alert System, which has enhanced cooperation among EU Member States and international partners such as NATO and the G7.⁴¹

In April 2018, the European Commission released a communication titled *Tackling Online Disinformation – A European Approach*, emphasizing the need for a comprehensive EU response. At the EU level, this document defines disinformation as verifiably false or misleading information created, presented, and disseminated for economic gain or to intentionally deceive the public. It highlights the role of social media platforms in amplifying disinformation and introduces the EU Code of Practice on Disinformation, a voluntary framework for online platforms to improve transparency and counter manipulation.⁴²

Building upon previous efforts, the European Commission adopted the *Democracy Action Plan in December 2020*, reinforcing commitments to counter disinformation. This plan focused on strengthening the Code of Practice on Disinformation by involving more service providers as signatories, enhancing transparency in political advertising, increasing support for independent fact-checkers and researchers, and implementing stricter measures against foreign interference in elections.⁴³

The EU has engaged major tech companies such as Facebook, Google, and Twitter in self-regulatory measures through the Code of Practice on Disinformation. Platforms are expected to increase transparency in political advertising, close fake accounts, prevent bot-driven amplification of false information, and promote fact-checking initiatives and authentic news sources. The EU has also invested in independent fact-checking networks and media literacy campaigns to enhance public resilience against misinformation. Programs such as *Media Literacy for All* seek to educate citizens on recognizing and critically evaluating information.⁴⁴

A central concern has been protecting electoral integrity from foreign interference. The Rapid Alert System, in conjunction with national electoral authorities, monitors and addresses disinformation threats in real time – a measure included in the Action Plan Against Disinformation.⁴⁵ While the Code of Practice remains voluntary, the EU has hinted at potential regulatory actions if platforms fail to comply.⁴⁶ The Digital Services Act (DSA), introduced subsequently, provides a legal framework to enhance accountability in online content moderation.⁴⁷

As can be seen from the overview above, the European Union has developed a multi-faceted approach to tackling online disinformation, evolving from early strategic communication efforts to more structured regulatory frameworks. Future directions likely involve stricter enforcement of platform responsibilities, greater transparency regarding online po-

⁴¹ Action Plan Against Disinformation 2018, 7.

⁴² Tackling Online Disinformation – A European Approach. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2018) 236 final. European Commission, Brussels, 26 April 2018 (hereinafter referred to as Tackling Online Disinformation – A European Approach, 2018). <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0236>

⁴³ Democracy Action Plan, 2020.

⁴⁴ Ibid.

⁴⁵ Action Plan against Disinformation, 2018.

⁴⁶ Tackling Online Disinformation – A European Approach, 2018.

⁴⁷ Democracy Action Plan, 2020.

litical content, and enhanced international cooperation to combat cross-border disinformation threats.⁴⁸ The EU is striving to balance disinformation regulation with the protection of freedom of expression, so as to ensure a resilient democratic society in the digital age.

In addition to the existing regulatory measures, the European Union has further strengthened its legislative framework against online disinformation through the Digital Services Act (DSA),⁴⁹ the European Media Freedom Act (EMFA),⁵⁰ and the Regulation on Transparency and Targeting of Political Advertising (TTPA)⁵¹ which will be later discussed in detail.

3.2. The Code of Practice on Disinformation

Of the EU tools for tackling online disinformation mentioned in the previous section, we will look in more detail at the Code, as it is specifically designed to ensure the transparency of information flows on online platforms.

The European Union has continued to refine its approach against online disinformation in accordance with its strategy.⁵² These efforts culminated in the evolution of the Code of Practice on Disinformation (Code2018) which included 21 commitments in different areas mentioned in the previous documents.

The self-regulatory system introduced by Code2018 required online platform service providers who committed to its application to publish their self-assessment reports for the first time in October 2019, based on their experiences from the first year of implementation. The application of Code2018 also resulted in specific measures by the stakeholders to combat disinformation; the European Commission (Commission) published these measures on September 10, 2020.⁵³

In its evaluation of Code2018, the Commission highlighted the successes of its application, while also emphasizing the need to further develop the Code2018 in several areas to ensure its full and consistent implementation by the online platform service providers party to it and the Member States. Based on practical experience, on May 26, 2021, the European Commission published guidance outlining the areas in which it deemed necessary to strengthen Code2018.⁵⁴ As a result, negotiations began between the signatory service providers, those wishing to join, and the Commission. While Code2018 marked a pioneering step in self-regulation, its evaluation revealed the

⁴⁸ Ibid., 4.

⁴⁹ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC. OJ L 277, 27.10.2022.

⁵⁰ Regulation (EU) 2024/1083 of the European Parliament and of the Council of 11 April 2024 establishing a common framework for media services in the internal market and amending Directive 2010/13/EU. OJ L 2024/1083, 17.4.2024.

⁵¹ Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising. OJ L 2024/900, 20.3.2024.

⁵² Tackling Online Disinformation – A European Approach, 2018.

⁵³ Code of Practice on Disinformation 2018.

<https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation>

⁵⁴ European Commission Guidance on Strengthening the Code of Practice on Disinformation. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2021) 262 final. European Commission, Brussels, 26 May 2021 (hereinafter referred to as Guidance on Strengthening the Code of Practice on Disinformation). <https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation>

need for further enhancements. This led to the development of the Strengthened Code of Practice on Disinformation (Code2022), which was officially launched on June 16, 2022.⁵⁵

3.2.1. *The Strengthened Code of Practice on Disinformation (Code2022) and its practice*

The Code2022 represents a significant upgrade over its predecessor, incorporating more stringent obligations for signatories and improving monitoring mechanisms. The number of signatories increased to 34 at its inception, including major digital platforms such as Microsoft, Meta, and Google. The revised Code2022 sets 44 commitments and 128 specific measures in key areas.⁵⁶ These were much more detailed and precise than the previous 21 commitments set by Code2018.

The primary commitments include: 1) *demonetization of disinformation* to ensure that advertising revenues do not support the spread of false information; 2) *transparency requirements* to provide meaningful data on political advertising and platform algorithmic operations; 3) *platform integrity measures* to implement mechanisms to detect and mitigate coordinated disinformation campaigns; 4) *user empowerment* to offer tools to help users identify false information and report misleading content; 5) *support for fact-checkers and researchers* to facilitate access to data for academic and independent verification purposes; 6) *crisis response mechanisms* to disinformation surges during crises such as pandemics or geopolitical conflicts.

These commitments are monitored through a multi-tiered approach, involving regular reporting, independent audits, and public transparency tools such as the *Transparency Center*, which provides stakeholders with access to progress reports and compliance records.⁵⁷

A crucial structural addition was the establishment of the Permanent Task Force, which plays a central role in ensuring ongoing improvement and implementation of Code2022.

3.2.2. *Structure and governance of Code2022*

The Permanent Task Force comprises representatives from the European Commission, the European Regulators Group for Audiovisual Media Services (ERGA)⁵⁸, the European Digital Media Observatory (EDMO), the European External Action Service (EEAS), and signatory platforms. The Permanent Task Force oversees compliance reports, which signatories must submit either annually or semi-annually depending on their size and role within the ecosystem.⁵⁹

⁵⁵ The Strengthened Code of Practice on Disinformation 2022. <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>

⁵⁶ Ibid.

⁵⁷ European Commission: Commission presents guidance to strengthen the Code of Practice on Disinformation [Press Release]. Brussels, European Commission, 26 May 2021. <https://digital-strategy.ec.europa.eu/en/news/commission-presents-guidance-strengthen-code-practice-disinformation>

⁵⁸ From 8 February 2025 the European Board for Media Services replaced and succeed ERGA.

⁵⁹ European Commission Guidance on Strengthening the Code of Practice on Disinformation. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2021) 262 final. European Commission, Brussels, 26 May 2021. <https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation>

The Permanent Task Force is divided to several subgroups, such as Monitoring and Reporting, Transparency Center, Integrity of Services, Advertising, Crisis Response or New Signatories subgroups.

3.2.3. *Criticism of the Code*

Experts and scholars have generally viewed the EU's voluntary Code of Practice on Disinformation with the recognition of its intent but significant criticism of its effectiveness. The 2018 Code was acknowledged as a 'logical and necessary first step', yet evaluations after its first year showed 'mixed results' and widespread dissatisfaction among stakeholders.⁶⁰ Analysts noted that the Code's voluntary nature (with no binding obligations or sanctions for non-compliance) led to uneven and inconsistent implementation by online platforms. Platforms committed to improve transparency (e.g. in political advertising), but in practice they were not obliged to implement any specific measures and largely relied on self-reporting.⁶¹ Experts in the field argued that this lack of enforceability undermined the Code's impact: in the first two years of application, compliance was driven mostly by reputational incentives rather than legal necessity. For example, all signatories adopted political ad transparency labels, yet their definitions of 'political' varied widely, resulting in inconsistent application of the commitments across platforms. Another factor which made the task facing experts in the field more difficult was that the Code's provisions to 'empower the research community' remained insufficient – companies did create data repositories (e.g. ad libraries) but retained unilateral control over access, leading the European Commission itself to lament 'episodic and arbitrary'⁶² data-sharing which frustrates independent scrutiny.⁶³ An Institute for Strategic Dialogue report evaluating the 2018 Code's rollout during the EU elections in 2019 highlighted major shortcomings in key areas (political ad transparency, platform integrity measures, and researcher support), concluding that the Code was a step forward but 'insufficient progress' had been made on many commitments. These findings underscored the limits of self-regulation and self-reporting, and prompted calls for "more effective and sustainable regulatory approaches" to combat online disinformation.⁶⁴

Researchers welcomed the 2022 Strengthened Code as a further step – it expanded the number of signatories, commitments and more detailed measures yet many experts maintained a guarded scepticism. The revised Code introduced clearer obligations, but it remained a voluntary instrument, pending integration into the co-regulatory framework of the Digital Services Act. Early academic assessments indicate that the 2022 Code's impact has been limited in practice. A cross-European study of online advertising, for instance, found no significant reduction in major ad networks' relationships with popular misinformation websites af-

⁶⁰ PAMMENT (2020): 1.

⁶¹ SHATTOCK (2021): 1–2.

⁶² European Commission: Assessment of the Code of Practice on Disinformation – Achievements and areas for further improvement. Commission Staff working document, SWD(2020) 180 final. European Commission, Brussels. 10 September 2020. 11. <https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement>

⁶³ SHATTOCK (2021): 1.

⁶⁴ COLLIVER (2020): 2.

ter the new Code's adoption.⁶⁵ Ad networks largely withdrew from very low-traffic, false-news sites, but continued to serve ads on 'highly unreliable' sites that generate the bulk of disinformation traffic.⁶⁶ This suggests that the strengthened commitments did not fully curb the financial incentives underpinning disinformation. Similarly, experts point out that several structural problems persist despite the Code's improvements: platform transparency reports still often lack uniform, quantitative outcome metrics, hampering external evaluation, and certain measures (such as data access for researchers or country-specific reporting) rely on goodwill rather than enforceable tasks.⁶⁷ Indeed, by late 2024, even regulatory bodies like ERGA echoed many scholarly concerns, noting gaps in consistent reporting, data quality, and member-state-level information under the Code's implementation.

3.2.4. ERGA Monitoring Projects

To further strengthen compliance mechanisms, ERGA launched a Monitoring Project to assess the implementation of Code2022 by signatory platforms.⁶⁸ This project aimed to develop a long-term methodology for monitoring digital platforms' adherence to the Code2022's commitments. It was implemented in collaboration with national regulators and an external contractor selected through an open call. The project evaluated measurable key performance indicators (KPI), available data on platform activities, and the resources necessary for effective oversight. Monitoring also included Qualitative Reporting Elements (QRE) and quantitative Service Level Indicators (SLI).⁶⁹

A key component of this initiative involved assessing how different national regulatory authorities can, at the same time, contribute to the evaluation of a common monitoring method that suits their specific digital environments. Therefore, the project also standardized reporting formats to ensure comparability across EU member states. The first Monitoring Procedure of the Code2022 assessed the baseline reports from 16 December 2022 to January 2023, which are publicly available on the Transparency Centre's website,⁷⁰ such as the following reports since then.

The first monitoring round of the Code2022 highlighted both progress and shortcomings in its implementation. In its report,⁷¹ ERGA commended the increased granularity of data reporting, the standardization of reporting templates, and the inclusion of measurable Key Performance Indicators (KPIs). The establishment of a Transparency Centre was seen as a positive step towards public accessibility. However, ERGA identified several weaknesses, including inconsistencies in reporting periods, excessive verbosity, lack of contextualized data, and insuffi-

⁶⁵ PAPADOGIANNAKIS et al. (2025): 4–7.

⁶⁶ Ibid., 9–10.

⁶⁷ Brooke TANNER: Commentary on the EU Code of Practice on Disinformation. *Brookings.edu*, 5 August 2022. <https://www.brookings.edu/articles/eu-code-of-practice-on-disinformation/>

⁶⁸ ERGA: ERGA report on the first year of the Strengthened Code of Practice on Disinformation. 2023. https://media-board.europa.eu/document/download/c7a390d6-3039-44fd-9acb-e6ad61d17b85_en?file-name=ERGA-SG3-report-CoP_June-2023_as-adopted.pdf

⁶⁹ Ibid.

⁷⁰ <https://disinfocode.eu/reports>

⁷¹ ERGA: ERGA report on the first year of the Strengthened Code of Practice on Disinformation. 2023.

cient Member State-level reporting. Additionally, some platform Signatories did not fully comply with reporting guidelines, limiting the ability to assess the Code's effectiveness. ERGA stressed the need for more structured data provision, particularly regarding the performance of automated moderation systems and the impact of AI-generated content. It also called for stronger cooperation among stakeholders and improved access to data for independent research. Going forward, ERGA urged Signatories to enhance data quality, adhere to reporting guidelines more strictly, and ensure timely updates on disinformation policies, emphasizing that these measures are crucial for the Code's success in combating online disinformation.

ERGA reported and published a statement on the progress and the challenges within the Code2022 in December 2024.⁷² Despite progress, several areas still require significant improvement, according to ERGA. A pilot assessment conducted in the spring of 2024 by nine national regulatory authorities identified gaps in the reports submitted by the Code's signatories. Key deficiencies were found in political advertising transparency, researcher data access, fact-checking support, media literacy initiatives, and the lack of country-specific reporting for certain European Economic Area states. Additionally, the uneven engagement of some very large online platforms and very large online search engines with the Code's obligations raises concerns about their commitment to addressing disinformation. ERGA emphasizes the need for stronger enforcement mechanisms, better quality data reporting, and enhanced cooperation with national regulators to ensure compliance. Developing a rapid response mechanism for countering disinformation – especially in politically sensitive periods such as elections – has seen progress, but further refinements are necessary to enhance its effectiveness, according to ERGA.

The findings contributed to refining the EU's broader disinformation monitoring strategy, ensuring that platforms remain accountable and proactive in mitigating the spread of false information.

3.2.5. *Monitoring mechanisms*

The Code2022 has incorporated several additional monitoring mechanisms to ensure oversight. First of all, Signatories publish signatory reports annually or semi-annually where they outline their progress in meeting commitments. Signatories are also subject to independent audits, where independent organizations assess the effectiveness of platform measures in combating disinformation.

The Permanent Task Force comprises a Crisis Monitoring Group, which is a specialized body to coordinate rapid responses to emerging disinformation threats, especially during elections or geopolitical crises.

The Code2022 creates the possibility for data-sharing. Platforms are encouraged to share relevant data with regulators, fact-checkers, and researchers to enhance transparency and improve disinformation detection strategies.

These mechanisms aim to ensure that Code2022 remains a dynamic framework, adapting to evolving threats and maintaining its effectiveness over time.

⁷² ERGA: ERGA Statement on the progress and challenges within the Code of Practice on Disinformation. 2024. https://media-board.europa.eu/document/download/55b7cf51-4cc8-4c7d-a363-c453bd411fdc_en?filename=ERGA-SG3-Statement-CoP_final.pdf

3.2.6. *Future outlook*

The strengthened regulatory framework under Code2022, coupled with enhanced monitoring efforts such as the Monitoring Procedures, marked a significant step forward in the EU's fight against disinformation. However, it does not solve all the challenges or create a legally binding effect. The only power that enhances enforceability is the publicity of the reports, but evaluators recurrently report a lack of sufficient data provided by signatories. In addition, the Code2022 is an 'à la carte' agreement, in which signatories can choose which commitments they are willing to fulfil and which they are not.

Several characteristics can be identified as having given rise to self-regulation in this field. The challenges and costs of oversight and enforcement, the rapidly changing nature of the regulatory subject that necessitates flexible adjustments, and the significantly greater sectoral knowledge of market players operating social media platforms compared to the state – all these factors may encourage states, and in this case, the European Union, to delegate certain responsibilities to self-regulation.

However, when considering self-regulatory systems in this case, we cannot speak of pure self-regulation, given that the Commission and other organizations of the European Union play an active role in the creation and implementation of norms which indicate a move towards co-regulation.⁷³

Overall, ensuring continued compliance and adapting to emerging threats will remain an ongoing challenge. Future developments may include further regulatory measures, increased enforcement capabilities, and the potential transformation of Code2022 into a binding regulatory framework, should voluntary compliance prove insufficient.

3.2.7. *The Future of the Code2022*

ERGA reaffirmed in a statement⁷⁴ its strong commitment to combating disinformation and referred to the conversion of the Code of Practice on Disinformation under the DSA (see point XY.) ERGA will also be subject to transformation under the EMFA. The European Board for Media Service (EBMS) will continue ERGA's role in monitoring the Code.⁷⁵

3.3. The Digital Services Act

The DSA has been in force since 17 February 2024. This regulation creates a comprehensive legal framework for online platforms which provide services within the EU. It has several provisions which are relevant for online disinformation regulation. The DSA addresses disinformation in its preamble several times and identifies it as a societal risk.⁷⁶ It further un-

⁷³ CHYSTOFOROVA–REVIGLIO (2025): 1–28., 5–7.

⁷⁴ ERGA: ERGA Statement on the progress and challenges within the Code of Practice on Disinformation. 2024.

⁷⁵ Ibid.

⁷⁶ DSA Recital (2).

derlines the danger of the manipulative techniques contributing to the spread of disinformation in various cases, such as advertising, or content disseminated by algorithms or targeting.⁷⁷

The DSA prescribes notice and action mechanisms for illegal content removal⁷⁸, transparency requirements for online advertisements, such as regular reporting obligations, traceability of traders and the creation of public repositories of advertisements.⁷⁹ It creates additional obligations for very large online platforms (VLOP) and very large online search engines (VLOSE) to assess systemic risks about the content published and the algorithms used in their services.⁸⁰ The DSA also prescribes due diligence duties for online intermediaries,⁸¹ and independent audits evaluating the compliance with its provisions.^{82 83}

3.3.1. The Codes of Conduct under Art. 45 of DSA

The DSA aims to regulate not just with binding provisions, but also encourages the establishment of self-regulatory measures. In the Preamble the DSA sets the target to rely on the provisions of the Code2022, when introducing the Codes of Conduct.

Changing the Strengthened Code of Practice on Disinformation (Code2022) into a Code of Conduct, as outlined under Article 45 DSA, alters the regulatory landscape for online platforms by creating obligations for the designated online platforms. According to Article 45 1. DSA the Commission and the Board shall encourage and facilitate the drawing up of voluntary codes of conduct at Union level to contribute to the proper application of the DSA. However, it is not clear how these make a difference when the legislation repeatedly uses wording that can be understood as not referring to the creation of legally binding codes of conduct. These codes of conduct will focus on the specific challenges of tackling different types of illegal content and systemic risks, in accordance with Union law in particular on competition and the protection of personal data.

The DSA requires such codes of conduct to set out their specific objectives, include key performance indicators to measure the achievement of those objectives and take due account of the needs and interests of all interested parties, particularly citizens, at Union level. It also requires regular reporting for participants to the Commission, the European Board for Digital Services (Board), or national Digital Service Coordinators (DSC). These bodies also monitor their activities and publish conclusions in reports. In the case of systematic failure to comply with the codes of conduct, the Commission and the Board may invite the signatories to the codes of conduct to take any necessary action.

⁷⁷ DSA Recital (69), (83), (84).

⁷⁸ DSA Art. 14–15.

⁷⁹ DSA Art. 24, 30, 39.

⁸⁰ DSA Art. 26–27, 33, 34–35.

⁸¹ DSA Art. 10–15, 35–37, 44–48.

⁸² DSA Art. 28, 33, 37.

⁸³ NANNINI et al. (2025)

The DSA prescribes specific rules for Codes of conduct for online advertising⁸⁴ and for accessibility⁸⁵. According to the DSA, the Commission shall encourage the development of these codes of conduct by 18 February 2025 and their application by 18 August 2025.

However, the shift towards a stricter compliance mechanism raises critical considerations about balancing regulatory goals with fundamental rights, notably freedom of expression. When assessing the previous reports based on the Code2022, evaluators and researchers often criticized the limited amount of and the quality of data in the reports on content removal. As these reports do not cover individual cases but rather describe general procedures, they do not provide transparency into individual cases and increase the risk of over-regulation: platforms might adopt overly cautious moderation policies to avoid sanctions, potentially leading to increased content restrictions or even censorship.

Therefore, ongoing dialogue and cooperation among regulators, platforms, and civil society stakeholders will be crucial to ensure the Codes of Conduct effectively combat disinformation while safeguarding democratic values and freedom of expression.

3.4. The European Media Freedom Act (EMFA)

The European Media Freedom Act (EMFA) is designed to promote and protect media freedom, pluralism, and editorial independence throughout the European Union. It introduces a unified set of rules to ensure that both public and private media can function without interference from political forces or economic pressure.

The EMFA includes measures to secure editorial autonomy, increase transparency regarding media ownership, and to ensure public service media remain independent. It also reinforces the protection of journalistic sources, including safeguards against surveillance technologies, and introduces clearer rules on content removal by major online platforms. Furthermore, the Act requires openness in state advertising practices and mandates that media market mergers be evaluated for their impact on pluralism and editorial freedom. The European Board for Media Services (former ERGA), comprising national media regulators, has been established to ensure consistent implementation. The Act entered into force on 7 May 2024 and became fully applicable from 8 August 2025.⁸⁶

The EMFA also provides transparency measures and relies on self-regulatory initiatives. Article 18 EMFA lays down specific rules for media service providers on very large online platforms. The very large online platforms shall provide a functionality allowing recipients of their services to declare that they are media service providers, to make declarations on their editorial independence from Member States, political parties or third countries and entities controlled or financed by third countries and that they are subject to the oversight of competent national authorities and adhere to any self- or co-regulatory mechanism governing editorial standards. It also gives the possibility to provide contact details and other legal information. Article 18 imposes control on these declarations as well, by

⁸⁴ DSA Article 46 .

⁸⁵ DSA Article 47.

⁸⁶ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/protecting-democracy/european-media-freedom-act_en

stipulating that very large online platforms shall seek confirmation from regulators on certain matters.

In the following, I highlight the key elements of Article 18. These specific measures may play a particularly important role in the case of media service providers which have submitted a declaration pursuant to paragraph 1 of this Article. These rules shall apply where a provider of a very large online platform intends to take a decision to suspend the provision of its online intermediary services in relation to content provided by a media service provider or a decision to restrict the visibility of such content, on the grounds that such content is incompatible with its terms and conditions. Prior to such a decision to suspend or restrict visibility taking effect, it shall communicate to the media service provider concerned a statement of reasons according to the given provisions⁸⁷ and the DSA for its envisaged decision to suspend or restrict visibility, and give the media service provider the opportunity to reply to the statement of reasons within 24 hours after receiving it. Providers of very large online platforms shall take all the necessary technical and organizational measures to ensure that complaints lodged by media service providers are processed and decided upon as a priority and without undue delay. A media service provider may be represented by a body in the internal complaint-handling process referred to in those articles.

In addition to Article 18 of the EMFA, the following article establishes the framework for a structured dialogue on such cases. Article 19 EMFA prescribes that the Board shall regularly organize a structured dialogue between providers of very large online platforms, representatives of media service providers and representatives of civil society in order to discuss experience and best practices in the application of Article 18, including with regard to the functioning of very large online platforms and their processes for moderating content provided by media service providers; foster access to diverse offerings of independent media on very large online platforms; and monitor adherence to self-regulatory initiatives which aim to protect users from harmful content, including disinformation and foreign information manipulation and interference. The Board shall report to the Commission on the results of the structured dialogues referred to in paragraph 1. Where possible, the Board shall make the results of such structured dialogues publicly available.

Therefore, EMFA, and particularly its Articles 18 and 19, create a framework intended to reinforce self-regulatory initiatives (such as the Code2022), and creates a platform for various key actors to discuss the outcomes of these provisions. It is also crucial that these articles place independent media service providers in a distinguished position, thus recognizing their highly important role in keeping citizens informed.

Some critics claim that such independent media service providers receive ‘privileged treatment’ or ‘media immunity’. However, these measures are designed to balance freedom of expression with transparency and responsibility, and stipulate measures against arbitrary or non-transparent decisions by powerful digital gatekeepers.⁸⁸

⁸⁷ Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services. OJ L 186, 11.7.2019.

⁸⁸ Matteo MONTI: Why online public discourse needs a media privilege: in defence of Article 18 of the EMFA. *EMFA Observatory Blog*, 1 November 2024. <https://cmpf.eui.eu/in-defence-of-article-18-of-the-emfa/>

3.5. Transparency and Targeting of Political Advertising (TTPA)

Last, but not least, the EU adopted a regulation on the transparency and targeting of political advertising. The primary objective of the TTPA is to establish a harmonized framework across the European Union which ensures transparency and accountability in political advertising, particularly in the digital environment. The regulation mandates that political advertisements must be clearly labeled as such, disclosing their sponsors and the amounts spent on them. It also sets strict guidelines on the use of personal data for targeted political advertising, in an effort to prevent manipulation and protect individual privacy. By enforcing these measures, the regulation seeks to uphold the integrity of democratic processes, ensuring that voters can make informed decisions free from undue influence or misinformation. This initiative aligns with the broader European Democracy Action Plan, reinforcing the EU's commitment to safeguarding democratic values in the digital age.⁸⁹ The TTPA will be fully applicable from 10 October 2025.

The European Commission aims to force service providers to comply with the regulation, therefore it has also published guidelines to support the TTPA's implementation.⁹⁰ However, following the adoption of the regulation, Google and Meta both announced that, as of October 2025, they would cease the dissemination of political advertising within the European Union. In their respective statements, the companies referred to the disproportionate complexity of the regulatory framework, the excessively broad scope of the definition of political advertising, and the legal uncertainty surrounding the technical and operational conditions for compliance. In light of these factors and to mitigate regulatory and liability risks, both undertakings decided to withdraw entirely from the EU political advertising market.

Meta argued that the Regulation imposes “unworkable” requirements, prompting what it described as a “difficult but necessary” decision to impose a prohibition on political advertising on its platforms.⁹¹ Google advanced analogous concerns, emphasizing that the definition of political advertising lacks sufficient precision, that no harmonized and comprehensive elections database exists across the Union's 27 Member States, and that the relevant guidance on implementation was issued only shortly before the Regulation's application date. As a result, Google concluded that compliance would entail significant legal uncertainty and operational risk.⁹²

The decision by Meta and Google to unilaterally withdraw from the EU political advertising market has sparked significant criticism from civil society and industry stakeholders, who argue that such self-imposed bans may inadvertently undermine the democratic

⁸⁹ Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising. OJ L 2024/900, 20.3.2024.

⁹⁰ Guidelines to support the implementation of Regulation (EU) 2024/900 on the transparency and targeting of political advertising. Communication from the Commission, C(2025) 6829 final. European Commission, Brussels, 8.10.2025. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=intcom:C%282025%296829>

⁹¹ Meta: Ending Political, Electoral and Social Issue Advertising in the EU in Response to Incoming European Regulation. Meta, 25 July 2025. <https://about.fb.com/news/2025/07/ending-political-electoral-and-social-issue-advertising-in-the-eu/>

⁹² Annette KROEBER-RIEL: *An update on political advertising in the European Union*. Google Blog, 14 November 2024. <https://blog.google/company-news/inside-google/around-the-globe/google-europe/political-advertising-in-eu/>

objectives of the TTPA. Critics contend that by removing major digital platforms from the political advertising ecosystem, a significant channel for legitimate civic engagement and issue advocacy – including communications by non-profit organizations, trade unions, and smaller political actors – is curtailed, thereby narrowing the scope of public discourse rather than enhancing it.⁹³ This concern is amplified in the context of electoral processes, where civic participation and access to information are essential to free and fair elections. Withdrawing paid political communication options risks entrenching disparities in visibility and may disproportionately disadvantage moderate voices.⁹⁴ Furthermore, civil society signatories to open letters have urged both companies to reconsider their bans, emphasizing that withdrawal does not absolve platform operators from their obligations to support democratic discourse and that engagement with EU institutions, rather than retreat, is a more constructive route to aligning corporate policies with regulatory objectives.⁹⁵

These developments underline the practical challenges associated with the enforcement of the TTPA, notwithstanding the efforts of the European Commission to facilitate compliance through the publication of non-binding implementation guidelines.

4. Conclusions

The European Union's regulatory framework for addressing online disinformation fundamentally reflects an idealized conception of the European public sphere, emphasising the importance of a well-informed citizenry as essential for democratic societies. According to EU policy and legislative documents, such as the Commission's Action Plan on Disinformation (2018) and the European Media Freedom Act (EMFA), access to verifiable, reliable, and diverse information is viewed as critical for enabling citizens to participate effectively and rationally in democratic processes. In the EU's view, independent news media play a key role in preserving the integrity of this public sphere, ensuring that citizens receive credible information to engage in informed decision-making and public debate.

Despite the EU's comprehensive strategies, several critical issues persist. Foremost among these is the challenge of ensuring consistent and effective application across platforms and member states. Although substantial efforts have been made to standardize monitoring and reporting procedures through initiatives such as the Code of Practice on Disinformation (Code2022), implementation remains uneven. ERGA identified critical gaps in reporting quality, notably in political advertising transparency, fact-checking support, media literacy initiatives, and country-specific data provision.

⁹³ Ivan VANDERMEERSCH: Meta's Political Ads Ban: Lessons for Democracy, Marketers, and Regulators. *FEDMA.org*, 2 October 2025. <https://www.fedma.org/2025/10/02/metas-political-ads-ban-lessons-for-democracy-marketers-and-regulators/>

⁹⁴ Liberties: Meta and Google, Don't Throw European Voters Under The Bus: Why Self-Imposed Political Advertising Bans Threaten Free and Fair Elections in Europe. *Liberties.eu*, 25 September 2025. <https://www.liberties.eu/en/stories/meta-google-advertising-bans/45513>

⁹⁵ VoxPublic: Open letter to Meta and Google regarding their decision to stop political advertising. *VoxPublic.org*, 29 September 2025. <https://www.voxpublic.org/Open-letter-to-Meta-and-Google-regarding-their-decision-to-stop-political.html?lang=en>

As these areas have been affected by binding legislation since being identified by the ERGA, the next few years may see some improvement in them.

Moreover, the self-regulatory aspect of the EU's approach has clear limitations. The Strengthened Code of Practice on Disinformation (Code2022), despite setting forth stringent guidelines and detailed obligations for signatories, remains inherently voluntary and is characterized as an 'à la carte' agreement. Signatories selectively commit to certain obligations, thus diminishing the overall effectiveness and consistency of compliance.

Data availability and quality continue to pose significant hurdles. Evaluations repeatedly indicate that platforms often provide limited, overly generalized data, so inhibiting precise assessment of moderation efficacy and algorithmic transparency. This data deficit also undermines regulatory effectiveness and public trust in online platforms, while making it difficult to acquire detailed knowledge of their activities.

Furthermore, procedural slowness and complexity impair timely responses to disinformation threats, particularly during politically sensitive periods. Although the establishment of crisis response mechanisms within Code2022 and the DSA represents progress, ERGA highlights ongoing inadequacies which require substantial refinement.

The role of media literacy emerged relatively late in EU initiatives; while initially underestimated, it slowly became recognized as a crucial factor. Recent programs such as Media Literacy for All demonstrate a growing commitment to empowering citizens in critically evaluating information.

The existing EU framework combines binding legislative instruments such as the Digital Services Act (DSA) and the European Media Freedom Act (EMFA) with non-binding yet influential instruments like Code2022. The transformation of Code2022 into a Code of Conduct under Article 45 of the DSA suggests a potential shift towards stricter enforcement mechanisms. However, balancing regulatory rigor with the protection of fundamental rights, notably freedom of expression, remains an ongoing and complex challenge, especially when addressing online disinformation. Ongoing dialogue among regulators, platforms, and civil society is crucial for effectively addressing these objectives.

In conclusion, the EU's multi-layered approach has represented a significant advance in combating online disinformation since 2015. Nevertheless, considerable efforts are necessary to overcome the above-mentioned limitations. Future strategies may require evolving from self-regulatory frameworks towards more binding, enforceable regulations to ensure comprehensive protection against disinformation threats, but the risk of overregulation remains a threat to the freedom of expression.

References

- CHAMBERS (2020) = Simone CHAMBERS: Truth, deliberative democracy, and the virtues of accuracy. Is fake news destroying the public sphere? *Political Studies*, Vol. 69. No. 1. (2020) 147–163. <https://doi.org/10.1177/0032321719890811>
- CHRISTIANS et al. (2009) = Clifford G. CHRISTIANS – Theodore L. GLASSER – Denis McQUAIL – Kaarle NORDENSTRENG – Robert A. WHITE: *Normative theories of the media. Journalism in democratic societies*. Champaign, University of Illinois Press, 2009.

- CHYSTOFOROVA–REVIGLIO (2025) = Kateryna CHYSTOFOROVA – Urbano REVIGLIO: Framing the role of experts in platform governance. Negotiating the code of practice on disinformation as a case study. *Internet Policy Review*, Vol. 14., No. 1. (2025) 1–28.
<https://doi.org/10.14763/2025.1.1823>
- COLLIVER (2020) = Chloe COLLIVER: *Cracking the code. An evaluation of the EU Code of Practice on Disinformation*. London, Institute for Strategic Dialogue, 2020.
https://www.isdglobal.org/wp-content/uploads/2020/06/isd_Cracking-the-Code.pdf
- CUILENBURG (2007) = Jan van CUILENBURG: Media diversity, competition and concentration. Concepts and theories. In: Els de BENS (Ed.): *Media between culture and commerce (Changing Media – Changing Europe Series. Vol. 4.)*. Malta, Intellect Ltd. – Gutenberg Press, 2007. 25–54.
- CUILENBURG–MCQUAIL (2003) = Jan VAN CUILENBURG – Denis MCQUAIL: Media policy paradigm shifts: Towards a new communications policy paradigm. *European Journal of Communication*, Vol. 18., No. 2. (2003) 181–207.
<https://doi.org/10.1177/026732310301800200>
- HELBERGER (2016) = Natali HELBERGER: Facebook is a new breed of editor: a social editor. *LSE Blogs*, 15 September 2016. <https://blogs.lse.ac.uk/medialse/2016/09/15/facebook-is-a-new-breed-of-editor-a-social-editor/>
- KARPPINEN (2007) = Kari KARPPINEN: Making a difference to media pluralism. A critique of the pluralistic consensus in European media policy. In: Bart CAMMAERTS – Nico CARPENTIER (Eds.): *Reclaiming the media. Communication rights and democratic media roles*. Bristol, Intellect, 2007. 9–30.
- KELLER (2011) = Perry KELLER: *European and international media law. Liberal democracy, trade, and the new media*. Oxford, Oxford University Press, 2011.
<https://doi.org/10.1093/acprof:oso/9780198268550.001.0001>
- LOVELESS (2020) = Matthew LOVELESS: Information and democracy. Fake news as an emotional weapon. In: Serena GIUSTI – Elisa PIRAS (Eds.): *Democracy and fake news. Information manipulation and post-truth politics*. London, Routledge, 2020. 64–76.
<https://doi.org/10.4324/9781003037385>
- MARCHETTI (2020) = Gloria MARCHETTI: Political disinformation on the net and the risk to democracy. *International Relations and Diplomacy*, Vol. 8, No. 5. (2020) 220–230.
<https://doi.org/10.17265/2328-2134/2020.05.004>
- MARCIEL (2023) = Rubén MARCIEL: On citizens' right to information. Justification and analysis of the democratic right to be well informed. *Journal of Political Philosophy*, Vol. 31., No. 3. (2023) 358–384. <https://doi.org/10.1111/jopp.12298>
- MARTENS et al. (2018) = Bertin MARTENS – Luis AGUIAR – Estrella GOMEZ-HERRERA – Frank MUELLER-LANGER: The digital transformation of news media and the rise of disinformation and fake news. *JRC Working Papers on Digital Economy*, No. 2. (2018)
<https://ideas.repec.org/p/ipt/decwpa/2018-02.html>
- MCQUAIL (1999) = Denis MCQUAIL: *Media Performance. Mass Communication and the Public Interest*. London, Sage, 1999.

- NANNINI et al. (2025) = Luca NANNINI – Eleonora BONEL – Davide BASSI – Michele Joshua MAGGINI: Beyond phasein. Assessing impacts on disinformation of the EU Digital Services Act. *AI and Ethics*, Vol. 5. (2025) 1241–1269.
<https://doi.org/10.1007/s43681-024-00467-w>
- PAMMENT (2020) = James PAMMENT: *The EU Code of Practice on Disinformation. Briefing note for the new EU Commission*. Washington, Carnegie Endowment for International Peace, 2020. https://carnegie-production-assets.s3.amazonaws.com/static/files/Pamment_-_EU_Code_of_Practice.pdf
- PAPADOGIANNAKIS et al. (2025) = Emmanouil PAPADOGIANNAKIS – Panagiotis PAPADOPOULOS – Nicolas KOURTELLIS – Markatos EVANGELOS: *Before & After. The Effect of EU's 2022 Code of Practice on Disinformation*. New York, Association for Computing Machinery, 2025. <https://doi.org/10.1145/3696410.3714898>
- SHATTOCK (2021) = Ethan SHATTOCK: Self-regulation 2.0? A critical reflection of the European fight against disinformation. *Harvard Kennedy School Misinformation Review*, Vol. 2., No. 3. (2021) <https://doi.org/10.37016/mr-2020-73>
- SIMONE–FERNBAC (2006) = Maria SIMONE – Jan FERNBAC: Invisible hands or public spheres? Theoretical foundations for U.S. broadcasting policy. *Communication Law & Policy*, Vol. 11., No. 2. (2006) 289–292. https://doi.org/10.1207/s15326926clp1102_5
- STRÖMBÄCK (2005) = Jesper STRÖMBÄCK: In search of a standard. Four models of democracy and their normative implications for journalism. *Journalism Studies*, Vol. 6., No. 3. (2005) 331–345. <https://doi.org/10.1080/14616700500131950>
- TOFF–KALOGEROPOULOS (2020) = Benjamin TOFF – Antonis KALOGEROPOULOS: All the news that's fit to ignore. How the information environment does and does not shape news avoidance. *Public Opinion Quarterly*, Vol. 84., Special Issue (2020) 366–390. <https://doi.org/10.1093/poq/nfaa016>
- VOSOUGHI et al. (2018) = The spread of true and false news online. *Science*, Vol. 359., No. 6380. (2018) 1146–1151. <https://doi.org/10.1126/science.aap9559>

Regulation of online disinformation in the Netherlands

RONAN Ó FATHAIGH

1. Introduction

The purpose of this chapter is to provide an overview of the current regulation of online disinformation in the Netherlands.¹ This chapter first describes the applicable regulatory framework in relation to online disinformation, including Dutch government policy, legislation, and self-regulation. It also describes the various institutions involved in tackling disinformation, including their powers, obligations, and procedures. This chapter then describes a particularly notable operation of this regulatory framework, namely a number of landmark judgments from Dutch courts involving online disinformation and online platforms, where the courts have stepped in to order platforms to reinstate accounts closed for disseminating online disinformation.

2. Regulatory background

This section describes the applicable regulatory framework in relation to online disinformation, in particular Dutch government policy, legislation, and self-regulation. At the outset, and before delving into legislation and self-regulation, it is important to emphasize that the Dutch regulatory framework is heavily influenced by the following principle enshrined in the Dutch government's official disinformation policy framework: freedom of expression and freedom of the press “must take centre stage” when adopting policy on disinformation, and “qualifying disinformation as such” and fact-checking are “*not* primary government duties”.² Notably, and for the purposes of this chapter, disinformation is defined by Dutch government policy as “deliberate, mostly covert, dissemination of misleading information with the aim of harming public debate, democratic processes, the open and knowledge economy or public health”.³ Crucially, it is a form of “harmful, but often legal” behavior, and disinformation is “part of

¹ See also, VAN HOBOKEN et al. (2019): 75–10.; and European Regulators Group for Audiovisual Media Services (2021): 40.

² Government-wide strategy for effectively tackling disinformation. Ministry of the Interior and Kingdom Relations, 2022, 1 (emphasis added). <https://web.archive.org/web/20250603202608/https://www.government.nl/binaries/government/documenten/parliamentary-documents/2022/12/23/government-wide-strategy-for-effectively-tackling-disinformation/nl-government-wide-strategy-against-disinformation.pdf>

³ Ibid., 1.

the freedom of speech”.⁴ As such, the Netherlands does not have a specific legislative provision prohibiting the dissemination of disinformation, although the Dutch Criminal Codes does criminalize other types of expression that may include disinformation, such as libel, slander, insult, group insult, and incitement to hatred or discrimination.⁵

2.1. Government-wide strategy for tackling disinformation

Before setting out legislation and self-regulation applicable to disinformation, it is important to begin with the Dutch government’s official policy framework for combatting disinformation. This is contained in the Dutch government’s landmark *Government-wide strategy for effectively tackling disinformation* (Disinformation Strategy), adopted in 2022.⁶ Demonstrating the multi-faceted dimensions to tackling disinformation, the Disinformation Strategy was jointly adopted and sent to Parliament by the Minister for Internal Affairs and Kingdom Relations, the State Secretary on Kingdom Relations and Digitalisation, the Minister of Justice and Security, and the State Secretary for Culture and Media. At the outset, and crucially, the government emphasized that the overarching principle of its policy is that “qualifying disinformation as such” and “fact-checking” are *not* primary duties for the government.⁷ Crucially, the policy reflects how the Netherlands advocates an “alternative to content control” that “safeguards human rights” and effectively counteracts disinformation campaigns.⁸ However, where national security, public health, social, or economic stability are at stake, the government “can act” and “debunk” disinformation.⁹ As mentioned above, the Disinformation Strategy defines disinformation as “deliberate, mostly covert, dissemination of misleading information with the aim of harming the public debate, democratic processes, the open and knowledge economy or public health”.¹⁰

The Disinformation Strategy underscores that an effective approach to tackling disinformation requires government-wide measures centered on fundamental rights, including freedom of expression and freedom of the press. In this regard, there are two main strands to the government’s strategy on countering disinformation, namely: (a) strengthening public debate, and (b) reducing the influence of disinformation. First, on strengthening the public debate, the strategy notes that promoting a “diverse media landscape” is important in order to limit the influence of disinformation; and the government is committed to maintaining confidence in and pluralism of Dutch media, as this “contributes to limiting the breeding ground for the negative effects of disinformation on society”.¹¹ As such, the government has adopted policies to strengthen public broadcasting, local broadcasting, and investigative journalism. Further, the government is strengthening citizens’ resilience to disinformation, with a number of new policies, including the Ministry of the Interior’s “intensification of public communication about

⁴ Ibid., 9.

⁵ See *Wetboek van Strafrecht* [Criminal Code] (2024): Art. 137c, 137d, 137e, Art. 261, and Art. 266.

⁶ <https://web.archive.org/web/20250603202608/https://www.government.nl/binaries/government/documenten/parliamentary-documents/2022/12/23/government-wide-strategy-for-effectively-tackling-disinformation/nl-government-wide-strategy-against-disinformation.pdf>. See also, Ó FATHAIGH (2023).

⁷ Government-wide strategy for effectively tackling disinformation, 1.

⁸ Ibid., 9.

⁹ Ibid., 1.

¹⁰ Ibid., 1.

¹¹ Ibid., 6.

the existence of disinformation”.¹² While the State Secretary for Culture and Media is working with the Dutch Media Literacy Network (a network consisting of several organizations committed to media literacy) on an awareness-raising process to increase knowledge and skills about the value of journalism in society, and making citizens media literate so as to allow them to recognize disinformation themselves.

Second, on reducing the influence of disinformation, there are a number of new strategies, which are based on the view that every government ministry and State authority must be able to “respond effectively and appropriately” when disinformation affects their policy area.¹³ For example, in the run-up to municipal elections and European Parliament elections, there are exercises organized by the Ministry of the Interior and Kingdom Relations with other ministries, and the National Coordinator for Counterterrorism and Security (Nationaal Coördinator Terrorismebestrijding en Veiligheid) (NCTV), on how to respond quickly and proportionately to disinformation campaigns. Further, the NCTV is tasked with developing expertise in the field of communication in the event of disinformation when it affects national security. If state actors produce or disseminate disinformation, the Ministry of Justice and Security, the Ministry of Foreign Affairs and, in the event of military actors, the Ministry of Defence are also involved. Finally, the NCTV has also set up the Authority for Online Terrorist and Child Pornography Material (Autoriteit Online Terroristisch en Kinderpornografisch Materiaal), which will send removal requests to online platforms where terrorist material is found on their platforms, which “may also cover disinformation containing terrorist material”.¹⁴

Based on the foregoing, it is clear that the twin pillars of the Dutch government’s disinformation strategy are as follows: first, strengthening public debate (where promoting a media pluralism and media literacy take center stage); and second, reducing the influence of disinformation, through proportionate responses to disinformation campaigns by various government ministries and agencies, in particular, where national security and elections are concerned. Crucially, when responding to disinformation, government ministries and agencies must “keep fundamental rights in mind” and must be “cautious” about qualifying disinformation as such.¹⁵

2.2. Ministry of the Interior as a “trusted flagger” for reporting disinformation

As mentioned above, the Dutch government’s Disinformation Strategy emphasized that the overarching principles of its policy were that “qualifying disinformation as such” and “fact-checking” are *not* primary duties for the government.¹⁶ However, it must be mentioned that the Dutch Ministry of the Interior made important disclosures in late 2022, in response to

¹² Ibid., 7.

¹³ Ibid., 2.

¹⁴ See further, *Arda Gerken bestuursvoorzitter Autoriteit online Terroristisch en Kinderpornografisch Materiaal* [Press release]. Algemene Bestuursdienst, 24 August 2023. <https://www.algemenebestuursdienst.nl/actueel/nieuws/2023/08/24/arda-gerkens-bestuursvoorzitter-autoriteit-online-terroristisch-en-kinderpornografisch-materiaal>

¹⁵ Government-wide strategy for effectively tackling disinformation, 2.

¹⁶ Ibid., 1.

a number of parliamentary questions, which revealed how the Ministry has had “trusted flagger” status with the largest online platforms for a number of years. Content flagged by trusted flaggers are treated by platforms as a priority for action, including removal. Indeed, the Ministry has used its trusted flagger status on a number of occasions, especially around elections, flagging content as disinformation.

In this regard, in December 2022, in response to parliamentary questions, the Ministry of the Interior disclosed in a letter to Parliament that it has the status of “trusted flagger” with a number of online platforms, including Meta, TikTok, X (formerly Twitter), Google and Snapchat.¹⁷ The Ministry disclosed that it had been a trusted flagger with Meta and X since May 2019 in the run-up to the 2019 European Parliament elections; and a trusted flagger with Google, TikTok and Snapchat since 2021 in the run-up to the 2021 Dutch parliamentary elections. Notably, the Ministry emphasized that it uses this tool with “great restraint” and “only around elections”, and only when there is a suspicion that the spread of disinformation has the consequence that the “integrity of the electoral process” could be hampered.¹⁸ The Ministry also stressed that the platforms “make their own independent assessment” of whether there has been a violation of their terms of services, and whether “removal, labelling or other action” is justified, and the Ministry has “no authority” to have content removed via the trusted flagger mechanism.¹⁹ Finally, the Ministry also disclosed that it has used its trusted flagger status on a number of occasions, flagging content with Meta and X in the run up to the 2021 Dutch preliminary elections, including in relation to a video on Facebook falsely stating that by voting you would immediately give permission for a COVID-19 vaccination; and Facebook posts falsely stating voters could color two boxes red on the ballot paper.

Notably, in September 2023, the Ministry of Interior published a public report on the deployment of its trusted flagger status during the 2023 provincial elections in the Netherlands to target election disinformation.²⁰ The Report emphasizes that it is not a “primary task of governments to label what disinformation is,” and crucially, notes the Ministry of the Interior has “no legal basis for processing personal data” by monitoring disinformation on online platforms.²¹ The Ministry also has “no authority” to have certain content “removed” via the trusted flagger status; however, the Ministry can use its status as a trusted flagger to report content that has been “identified” by fact-checking organizations, media or other government agencies, and where the content may have an impact on the “integrity of the election process”.²² Further, the Ministry stated again it was “important to emphasize” that it uses the trusted flagger status “very cautiously”, and only when there is a “risk to the organization, implementation and integrity of the election

¹⁷ Document no. 2022-0000626170, Beantwoording Kamervragen van het lid Van Houwelingen (FvD) over de ‘escalatie kanalen’ en het ‘desinformatieteam BZK’ naar aanleiding van een Wob-verzoek. The Hague, Ministry of the Interior and Kingdom Relations, 2022. <https://open.overheid.nl/documenten/ronl-cd2964599bddf03d59dea9dcc82b95b93a21daf1/pdf>

¹⁸ Ibid., 2.

¹⁹ Ibid., 2.

²⁰ *Inzet Trusted Flagger Status BZK Provinciale Staten en Waterschapsverkiezingen 2023*. The Hague, Ministry of the Interior and Kingdom Relations, 2023. <https://www.rijksoverheid.nl/documenten/rapporten/2023/09/25/15-bijlage-bij-brief-rapportage-inzet-trusted-flagger-status-bzk>

²¹ Ibid., 2.

²² Ibid., 2.

process”.²³ The Ministry does not examine any statements about other topics related to the elections, such as “election promises”.²⁴ The platforms to which a report is made make their own independent assessment of whether there has been a violation of the user conditions and therefore whether removal, labelling or other action is justified.

The Report details how the Ministry used its trusted flagger status during the 2023 provincial elections. The first followed various media reports about Twitter (now X) posts where voters were encouraged to cast an invalid vote, by writing the name of party leaders on ballots, or checking two boxes. The Ministry considered this “could lead to an invalid vote”, and flagged the posts with Twitter to “assess them based on Twitter’s own policy” on election disinformation.²⁵ The second instance concerned posts on Twitter, with false voting advice for voters of a certain political party, instructing voters to write the name of the party leader on a ballot and to put a check mark next to it.²⁶ Again, the Ministry considered these posts “could lead to an invalid vote”, and flagged the posts with Twitter to “assess them based on Twitter’s own policy” on election disinformation.²⁷

Most recently in December 2023, the Minister for the Interior sent a report to Parliament, detailing the Ministry’s measures against disinformation during the 2023 Dutch parliamentary elections in November 2023.²⁸ First, prior to the elections, the Ministry of the Interior and Kingdom Relations had discussions at an “official level” with representatives of Google, Meta, Microsoft, Snapchat, TikTok to gain insight into the “mitigating measures that the platforms take and their effectiveness”.²⁹ During these sessions, platforms policies were discussed, as laid down in their own terms of use, to combat disinformation surrounding the elections. Second, in terms of the trusted flagger mechanism, the Ministry disclosed that it had contacted Meta during the election period concerning a message Facebook users had received, incorrectly stating “today are the elections” one week before the election; while also flagging posts on X calling for the coloring of two boxes on the ballot paper, which would lead to invalid votes.

2.3. Global Declaration on Information Integrity Online

A further development to mention in relation to the Dutch government’s policy on disinformation is the Dutch Ministry of Foreign Affairs’ central involvement in the recently adopted

²³ Ibid. 2.

²⁴ Ibid., 2.

²⁵ Ibid., 3.

²⁶ Ibid., 3.

²⁷ Ibid., 3.

²⁸ Antwoorden op Kamervragen over beïnvloeding van en door overheid op social media m.b.t. Tweede Kamerverkiezingen, November 22, 2023. The Hague, Ministry of the Interior and Kingdom Relations, 2023. <https://www.rijksoverheid.nl/documenten/kamerstukken/2023/12/22/antwoorden-op-kamervragen-over-beïnvloeding-van-en-door-overheid-op-social-media-mbt-tweede-kamerverkiezingen-22-november-2023>; See also, Sjoerd HARTHOLT: Ministerie grijpt tweemaal in tegen desinformatie. Binnenlands Bestuur, 2 January 2024. <https://www.binnenlandsbestuur.nl/digitaal/twee-keer-contact-ministerie-met-social-media-platforms>

²⁹ Ibid., 2.

Global Declaration on Information Integrity Online.³⁰ In particular, in 2023, the Dutch and Canadian governments jointly launched the Global Declaration, which has been endorsed by 32 countries worldwide.³¹ The purpose of the Declaration is to “address the erosion of information integrity online, which includes the spread of misinformation and disinformation campaigns by state and non-state actors”.³² The term “information integrity” is defined as an “information ecosystem that produces accurate, trustworthy and reliable information, meaning that people can rely on the accuracy of the information they access while being exposed to a variety of ideas”.³³ The Declaration’s use of information integrity, rather than a focus on disinformation, is designed to “offer a positive vision of a broader information ecosystem that respects human rights”.³⁴

Crucially, under the Declaration, the participating States, including the Netherlands, commit to (a) promote and respect “pluralistic media and journalism”, and protect access to media content as a “measure to counter disinformation”; take “active steps” to address misinformation and disinformation targeted at “women, LGBTQ+ persons, persons with disabilities and Indigenous Peoples”, acknowledging that women, adolescents and girls, in all their diversity, persons belonging to minorities and persons in vulnerable situations are “specifically targeted and impacted”.³⁵ Notably, the Netherlands committed to refrain from “unduly restricting” human rights online, especially freedom of expression, “under the guise” of countering disinformation, including blocking or restricting access to the Internet, eroding privacy, intimidating, harassing or abusing journalists, researchers and human rights defenders, interfering with their ability to operate freely, or “criminalising or otherwise punishing the exercise of the right to freedom of expression online”.³⁶

2.4. Dutch Code of Conduct Transparency Online Political Advertisements

In addition to government policy and trusted flagger measures, an important self-regulatory instrument to mention in relation to disinformation is the Dutch Code of Conduct Transparency Online Political Advertisements (Code of Conduct).³⁷ This is a voluntary Code of Conduct that was agreed in the run-up to the 2021 Dutch parliamentary elections, between the Ministry of the Interior and Kingdom Relations, a number of online platforms (including Facebook, Google, Snapchat and TikTok), most Dutch political parties and the Interna-

³⁰ *Canada and the Netherlands launch the Global Declaration on Information Integrity Online* [Diplomatic statement]. Government of the Netherlands, 20 September 2023. <http://www.government.nl/documents/diplomatic-statements/2023/09/20/global-declaration-on-information-integrity-online>

³¹ Ibid.

³² Ibid.

³³ Ibid.

³⁴ Ibid.

³⁵ Ibid.

³⁶ Ibid.

³⁷ Dutch Code of Conduct Transparency Online Political Advertisements. The Hague, International Institute for Democracy and Electoral Assistance, 2021. www.idea.int/sites/default/files/news/news-pdfs/Dutch-Code-of-Conduct-transparency-online-political-advertisements-EN.pdf

tional Institute for Democracy and Electoral Assistance (an intergovernmental organization intended to support and strengthen democratic institutions). Notably, the Code of Conduct contains a number of provisions which are applicable to disinformation during elections.

First, the signatories to the Code of Conduct, including political parties and on-line platforms, jointly committed to the need to “avoid” the dissemination of “misleading content”.³⁸ Indeed, all the political parties committed to refraining from “unethical online behaviour”, such as disseminating online content in view of “voter dissuasion”, “disinformation about the voting process”, using fake accounts or automatic bots to manipulate voter opinions, trolling, unfounded accusations or cyber-bullying.³⁹ Furthermore, political parties committed to “refrain from deliberately posting, disseminating, or promoting ‘misleading content’, and refrain from using ‘distorted’ audio-visual messages, including ‘deep fakes’”.⁴⁰ Notably, platforms committed to considering “countering” any “inaccurate information” on the electoral process, such as inaccurate voter or voting information, and inaccurate polling station information, in case such information is identified.⁴¹ Crucially, platforms also committed to putting in place “response mechanisms” to answer questions or address issues related to the Dutch elections; having clear and transparent rules for placing and “removing messages”; and responding “quickly” and accurately to enquiries about “removal” of messages from “competent authorities and political parties”.⁴² Additionally, platforms also committed to considering conducting a post-election review on the Dutch electoral campaign and compliance with the Code of Conduct; considering the possibility of publishing a report on any relevant incidents in relation to the Dutch elections and corresponding platform action. Finally, the Code of Conduct states that it is a voluntary code of conduct, and in “no way intends to create a legal framework”. However, while well-intentioned, there has been some critique of the Code of Conduct, as it lacks transparency in terms of actual implementation.⁴³

2.5. Regulation on state-owned media outlets engaged in disinformation and propaganda

In terms of legislation applicable to disinformation, while the Netherlands does not have a specific provision prohibiting disinformation as such, an important legislative instrument to mention is a regulation banning Russia Today and Sputnik for distributing disinformation and propaganda, including blocking websites in the Netherlands. In March 2022, the Minister of Foreign Affairs adopted a Regulation,⁴⁴ implementing the Council of the EU’s 2022 Decision and Regulation, which banned Russia State media outlets Sputnik and Russia Today, in response to Russia’s invasion of Ukraine, its ‘disinformation campaign’ and Russian

³⁸ Ibid., section 3.

³⁹ Ibid., section 3.2.9.

⁴⁰ section 3.2.10.

⁴¹ Ibid., section 23.

⁴² Ibid., section 21.

⁴³ Ó FATHAIGH et al. (2021): 871.

⁴⁴ Regeling van de Minister van Buitenlandse Zaken, nr. Min-BuZa.2022.11520-15 (2022) [Regulation of the Minister for Foreign Affairs].

media outlets engaging in ‘propaganda actions’.⁴⁵ This prohibited “operators from broadcasting or enabling, facilitating or otherwise contributing to the broadcasting of any content by [Russia Today and Sputnik], including through transmission or distribution by any means such as cable, satellite, IP-TV, internet service providers, internet video-sharing platforms or applications, whether new or pre-installed”.⁴⁶

Notably, in response to the Regulation, the largest Internet service providers in the Netherlands announced that the websites of the Russian state-owned media outlets Russia Today and Sputnik would be blocked in the Netherlands.⁴⁷ Crucially, the Netherlands Authority for Consumers and Markets (Autoriteit Consument & Markt) (ACM) also published an important statement announcing that it had informed Dutch telecom operators that the EU Open Internet Regulation was “not an obstacle to the implementation of the European sanctions regarding the suspension of distribution of Russian media channels RT and Sputnik”, and it meant that Dutch ISPs are “allowed to block the websites of RT and Sputnik as long as the thereto-related European sanctions are in place”.⁴⁸ Further, the ACM stated it would not take “any enforcement action as long as the European sanctions are in place”.⁴⁹

It should be noted that the blocking of RT and Sputnik have proved quite controversial in the Netherlands. The trade association for the Dutch broadband industry (NLConnect) issued a statement, advising its members to block the RT and Sputnik websites. However, the association also stated that the EU Council Regulation was “extremely unclear”, and was advising its members “under protest”, because “we are for a free and open Internet”.⁵⁰ Further, a Dutch coalition of ISPs and media organizations made an application to the Court of Justice of the European Union, seeking annulment of the Council of the EU’s bans on Russia Today and Sputnik. The applicants in the case are supported by the Dutch Freedom of Information Coalition, which includes the Dutch Association of Journalists and the Dutch Press Freedom Fund, and the application argues, among other grounds, that the bans violate the right to freedom of expression under Article 11 of the Charter of Fundamental Rights of the EU.⁵¹

⁴⁵ See Council of the European Union: EU imposes sanctions on state-owned outlets RT/Russia Today and Sputnik’s broadcasting in the EU [Press release]. Council of the European Union, 2 March 2022. <https://www.consilium.europa.eu/en/press/press-releases/2022/03/02/eu-imposes-sanctions-on-state-owned-outlets-russia-today-and-sputnik-s-broadcasting-in-the-eu/>; see also Council Decision (CFSP) 2022/351 of 1 March 2022 amending Decision 2014/512/CFSP concerning restrictive measures in view of Russia’s actions destabilising the situation in Ukraine. <http://data.europa.eu/eli/dec/2022/351/oj>; and Council Regulation (EU) 2022/350 of 1 March 2022 amending Regulation (EU) No 833/2014 concerning restrictive measures in view of Russia’s actions destabilising the situation in Ukraine. <http://data.europa.eu/eli/reg/2022/350/oj>

⁴⁶ Council Regulation (EU) 2022/350, Art. 1.

⁴⁷ NOS Nieuws: Sites RT en Sputnik geblokkeerd door grootste internetproviders. NOS.nl, 8 March 2022. <https://nos.nl/collectie/13888/artikel/2420302-sites-rt-en-sputnik-geblokkeerd-door-grootste-internetproviders>

⁴⁸ Netherlands Authority for Consumers and Markets: Open Internet Regulation is not an obstacle to blocking RT and Sputnik because of EU sanctions. ACM.nl, 8 March 2022. <https://www.acm.nl/en/publications/open-internet-regulation-not-obstacle-blocking-rt-and-sputnik-because-eu-sanctions>

⁴⁹ Ibid.

⁵⁰ NLconnect: Statement NLconnect over blokkeren Russia Today. NLConnect.org, 8 March 2022. <https://www.nlconnect.org/nieuws/statement-nlconnect-over-blokkeren-russia-today>

⁵¹ A2B Connect and Others v Council (Case T-307/22) (2022/C 276/24)

2.6. Digital Services Act Implementation Bill

Finally, in terms of legislation and authorities involved in the tackling of disinformation, the EU's Digital Services Act is of course a major piece of EU legislation relating to disinformation, and is designed to address the dissemination of illegal content and the "societal risks that the dissemination of disinformation" may generate.⁵² In this regard, the Digital Services Act Implementation Bill has been published by the Dutch government, and (at the time of writing) is making its way through Parliament.⁵³ Notably, the Bill designates the Netherlands Authority for Consumers and Markets (ACM) as the national Digital Services Coordinator (DSC) in the Netherlands under the DSA. The Bill also provides the ACM with various powers necessary under the DSA, including powers to impose administrative fines, requisition information, conduct inspections of business premises, and investigate cases. Notably, this includes designating "trusted flaggers" under Article 22 of this DSA.⁵⁴ While the ACM will also award "vetted researcher" status to researchers for the purpose of conducting research that contributes to the detection, identification and understanding of systemic risks under Article 34 DSA and Article 35 DSA in relation to Very Large Online Platforms, which would include disinformation.⁵⁵ Notably, while the implementing Bill was making its way through Parliament, a Ministerial Decision was adopted in February 2024, "provisionally" designating the ACM as the national DSC in the Netherlands.⁵⁶ However, this provisional designation only gives limited provisional powers to the ACM under the DSA, such as functioning as a contact point for DSCs of other Member States and the European Commission in the context of mutual assistance. Crucially, the provision designation does not extend to granting powers under the substantive provisions of the DSA, such as enforcement powers, or awarding the status of "trusted flagger" or "vetted researcher". The ACM will only have these powers once the Implementing law has entered into force. Thus, the ACM has an important role going forward as the national DSC in the Netherlands in relation to enforcement of the DSA.

⁵² Regulation (EU) 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), Recital 9. <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>

⁵³ Wetsvoorstel tot uitvoering van de digitale dienstenverordening [Digital Services Act Implementation Bill] (2022).

⁵⁴ Regulation (EU) 2022/2065, Article 22.

⁵⁵ Wetsvoorstel tot uitvoering van de digitale dienstenverordening [Digital Services Act Implementation Bill] (2022).

⁵⁶ Besluit van de Minister van Economische Zaken en Klimaat van 11 februari 2024, nr. WJZ/ 45119378, tot voorlopige aanwijzing van de Autoriteit Consument en Markt als bevoegde autoriteit en digitale dienstencoördinator in de zin van Verordening (EU) 2022/2065 van het Europees Parlement en de Raad van 19 oktober 2022 betreffende een eengemaakte markt voor digitale diensten en tot wijziging van Richtlijn 2000/31/EG (digitale dienstenverordening) (PbEU 2022, L 277) (Besluit voorlopige aanwijzing ACM als bevoegde autoriteit en digitale dienstencoördinator digitale dienstenverordening) [Decision of the Minister of Economic Affairs and Climate of February 11 2024, no. WJZ/ 45119378, provisionally designating the Authority for Consumers and Markets as competent authority and digital services coordinator within the meaning of Regulation (EU) 2022/2065 of the European Parliament and the Council of 19 October 2022 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act)] <https://zoek.officielebekendmakingen.nl/stcrt-2024-3993.html>

2.7. Dutch Advertising Code and Code for Social Media & Influencer Marketing

The self-regulatory framework in the Netherlands on advertising is also relevant in relation to the dissemination of disinformation, in particular the rules relating to social media influencers. The Dutch Advertising Code Foundation is the self-regulatory body for advertising in the Netherlands, and maintains the Dutch Advertising Code. Consumers, companies and other organizations that have a complaint about an advertisement can submit it to the independent Advertising Code Committee for a decision. Notably, the Dutch Advertising Code is based on a very broad definition of advertising, which also includes political advertising, and is defined as “any public and/or systematic direct or indirect commendation of goods, services and/or ideas by an advertiser or in whole or in part for their benefit, whether or not with the help of third parties”.⁵⁷ The Code contains numerous rules in relation to truth and accuracy in advertising, including that advertising must be in accordance with “truth, good taste and decency”;⁵⁸ advertising “shall not be dishonest”, “misleading”, and not include “incorrect information”.⁵⁹ Further, there is the Advertising Code for Social Media & Influencer Marketing, which applies to online content creators such as bloggers, vloggers and influencers.⁶⁰ All advertising and influencer marketing on social media platforms is subject to the Code, and it also applies to advertisers, having been drawn up in consultation with advertisers, media, social media professionals, communication consultancies and consumer representatives. The Code contains a ban on manipulation, including that “[m]odifying posts or other communications on social media in such a manner that the average consumer may be misled is prohibited”.⁶¹ The Code also prohibits “systematically creating and/or using false or non-existent identities in bulk to communicate about a product and/or service through social media”, and the use of “fake likes and fake followers” is also prohibited.⁶²

2.8. Guidelines of the Press Council

As mentioned above, the Dutch Disinformation Strategy views a pluralistic media environment as “key” to limiting the influence of disinformation, meaning that it is crucial for citizens to have access to high-quality and reliable news.⁶³ In this regard, the self-regulatory framework applicable to journalism in the Netherlands is also highly relevant. Notably, the Netherlands Press Council is an independent self-regulatory body for the media in the Netherlands, where interested parties can submit complaints about journalistic activities. The Council assesses whether a publication has exceeded the boundaries of journalistic ethics under the Guidelines of the Press Council.⁶⁴ If the Council upholds a complaint, the medium concerned is requested to publish the ruling in full or in summary, but the Council cannot

⁵⁷ Dutch Advertising Code, Article 1.

⁵⁸ *Ibid.*, Article 2.

⁵⁹ *Ibid.*, Article 7 and Article 8.

⁶⁰ Advertising Code for Social Media & Influencer Marketing.

⁶¹ *Ibid.*, Article 4(a) and Article 4(b).

⁶² *Ibid.*, Article 4(d).

⁶³ Ministry of the Interior and Kingdom Relations (2022): 6.

⁶⁴ Guidelines of the Netherlands Press Council, 2021. <https://www.rvdj.nl/english/guidelines>

impose sanctions. Notably, a Chairman of the Press Council has stated that the “spread of disinformation” makes it “necessary” for journalism to distinguish itself for its professionalism.⁶⁵ Crucially, the Guidelines have a number of provisions that help distinguish responsible journalism from disinformation, including the principle that proper journalism is “truthful and meticulous, impartial and fair, verifiable and honest and ethical”.⁶⁶ In particular, journalists must report “truthfully, verifiably and as completely as possible”, and avoid “one-sided and biased” reporting.⁶⁷ Additionally, journalists must make a “clear distinction between facts, allegations and opinions”,⁶⁸ and image manipulations may not be misleading.⁶⁹

3. Practice

Having set out the regulatory framework above, this section introduces some of the regulatory practice in relation to online disinformation. Notably, the Netherlands has been the scene of a number of important court cases in relation to online disinformation on online platforms, and this section will set out some of the most important judgments in relation to online disinformation.

3.1. Facebook accounts shut down over COVID-19 disinformation

A first notable judgment was delivered in late 2021, where a Dutch District Court considered Facebook’s removal of posts under its disinformation rules related to COVID-19.⁷⁰ Crucially, the case concerned whether individuals could invoke the right to freedom of expression under Article 10 of the European Convention on Human Rights⁷¹ against online platforms, when they remove content based on a platform’s terms of service on disinformation. The case involved Willem Engel, a well-known Dutch activist who campaigned against COVID-19 measures implemented by the Dutch government and was director of a high-profile campaign group “Stichting Viruswaarheid” (“Virus Truth Foundation”). Engel operated a Facebook account, and in summer 2021, Facebook removed a number of posts by Engel concerning COVID-19, for violating Facebook’s terms of service on disinformation. The posts included the statements (a) that certain substances in the Pfizer vaccine are “very, very toxic”, (b) in relation to the Ivermectin drug, “Make sure you bring it home”, and (c) “I hope people read this and see that it is a nuanced story and certainly not a killer virus that threatens the population”.⁷² In addition, Facebook suspended Engel’s access to his account for 30 days.⁷³

⁶⁵ Frits VAN EXTER: Frits van Exter over journalistieke dilemma’s op jaarvergadering Raad. *Villamedia*, 6 June 2018. <https://www.villamedia.nl/artikel/frits-van-exters-over-journalistieke-dilemmas-op-jaarvergadering-raad>

⁶⁶ Guidelines of the Netherlands Press Council, Preface.

⁶⁷ *Ibid.*, Section A.

⁶⁸ *Ibid.*, Section C.

⁶⁹ *Ibid.*, Section C.

⁷⁰ ECLI:NL:RBROT:2021:10459, Rechtbank Rotterdam, 29 October 2021, <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBROT:2021:10459>; see Ó FATHAIGH (2022).

⁷¹ European Convention on Human Rights (1950), Article 10.

⁷² ECLI:NL:RBROT:2021:10459, Rechtbank Rotterdam, 29 October 2021, para. 4.20.

⁷³ *Ibid.*, para. 2.8.

Following the removal of the posts and his account suspension, Engel initiated legal proceedings against Facebook, seeking reinstatement of the posts. Engle argued that Facebook acted “unlawfully” by violating his right to freedom of expression under Article 10 ECHR, as Facebook had a “duty of care” to protect users’ freedom of expression due to its “enormous reach”.⁷⁴ At the outset, the Court first held that while Article 10 ECHR does not have a “direct horizontal effect”, it does have “indirect effect” in the relationship between the parties.⁷⁵ In this regard, the Court noted that Engel’s right to freedom of expression must be balanced against Facebook’s right to property, which incorporates Facebook being “in principle” free to “determine the rules for the use of its platform”.⁷⁶

The Court then examined whether Facebook violated Engel’s freedom of expression through the application of its COVID-19 disinformation policy. In this regard, the Court held that Facebook had legitimate reasons for restricting Engel’s freedom of expression, especially as Facebook had issued its COVID-19 policies “in response to a call from the European Commission and international health organizations to combat disinformation about COVID-19, with the aim of protecting the safety and public health of the public”.⁷⁷ Further, Engel’s right to freedom of expression does not include a right to use the Facebook platform to express all his views on COVID-19, and the mere fact that Facebook has a ‘very large reach’ is irrelevant.⁷⁸ In particular, it was not the case that Engel has no other avenues to express his views on COVID19 policy. Notably, the Court admitted that in “specific cases, application of the policy may lead to correct messages being deleted”, but nonetheless, Facebook may “reasonably rely” on the expertise of experts from various health organizations and governments in determining its COVID-19 disinformation policy.⁷⁹

Finally, the Court considered whether the posts had been “correctly” removed by Facebook under its terms of service.⁸⁰ The Court examined the posts, and held that Facebook was entitled to consider that Engel’s statements violated Facebook’s rule on downplaying the seriousness of COVID-19.⁸¹ Therefore, the Court concluded that Facebook had not acted unlawfully, was entitled to remove the posts, and that there had been no violation of Engel’s freedom of expression. Thus, the District Court granted platforms a wide margin to remove disinformation under its own policies.

3.2. Parliamentary videos removed from YouTube

A second judgment to note involved YouTube’s removal of a Dutch Member of Parliament’s video in the run-up to the Dutch parliamentary elections in March 2021. Crucially, the Court held that YouTube had not violated the politician’s freedom of expression, and although YouTube has a “vast reach”, it does not have an obligation to allow content that vio-

⁷⁴ Ibid., para. 4.10.

⁷⁵ Ibid., para. 4.11.

⁷⁶ Ibid., para. 4.12.

⁷⁷ Ibid., para. 4.12.

⁷⁸ Ibid., para. 4.12.

⁷⁹ Ibid., para. 4.14.

⁸⁰ Ibid, para. 4.16.

⁸¹ Ibid, para. 4.22.

lated its disinformation policy.⁸² The case demonstrates how far online platforms can go in removing disinformation, even where it involves a politician's remarks during a parliamentary committee session. The case involved Wybren van Haga, a Dutch politician and Member of the House of Representatives, who was critical of the Dutch government's COVID-19 measures. In early 2021, van Haga uploaded a video to YouTube, and featured a discussion between the MP and the Director of the National Centre for Infectious Decision Control, during a parliamentary committee hearing on the government's COVID-19 measures.⁸³ During the hearing, the MP made comments that the infection-mortality ratio (IFR) of COVID was comparable to that of a severe flu.⁸⁴ Two days later, YouTube removed the MP's video for violating YouTube's terms of service and COVID-19 misinformation policy, which prohibit "Content claiming that the symptoms, deaths, or contagiousness of COVID-19 are less severe than or as severe as those of a cold or the flu".⁸⁵

On 17 March 2021, the final polling day of the Dutch parliamentary elections in 2021, the online news channel Bkbbx.tv posted a video on its YouTube channel, of an interview with the MP, discussing YouTube's removal of the MP's video featuring the parliamentary committee clip. During the discussion, the MP stated that "With those mouth caps, it is of course the biggest farce we have experienced".⁸⁶ YouTube also removed this interview video on the same day for violating its COVID-19 misinformation policy.

Following the video removals, van Haga and Bkbbx initiated legal proceedings against Google, seeking orders to have the videos reinstated on YouTube. At the outset, the Court examined the videos, and first held that Google was entitled to remove the videos on the basis of its terms of service, as the MP's statements comparing COVID-19 to the flu violated YouTube's COVID-19 misinformation policy. However, the Court then went on to examine whether this decision had been "unlawful" in light of the claimants' right to freedom of expression under Article 10 ECHR.⁸⁷ First, the Court reiterated that under certain circumstances, the state may have a "positive obligation" under Article 10 ECHR to actively protect freedom of expression in legal relationships between private parties.⁸⁸ However, the right to freedom of expression does not imply a right to the "forum" of one's choice.⁸⁹ The Court admitted that it was "understandable" that the claimants prefer YouTube because of its "enormous reach", but this is not enough to "oblige" Google to tolerate claims that violate its COVID-19 policy.⁹⁰ The Court noted that Google's COVID-19 misinformation policy was partly based on the EU Code of Practice on disinformation, and as such was "responding to the call from governments to assist them in the fight against the spread of incorrect information about COVID-19".⁹¹ Furthermore, the Court held that Google's

⁸² ECLI:NL:RBAMS:2021:4308, Amsterdam Rechtbank, 18 August 2021, <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBAMS:2021:4308>; see also, Ó FATHAIGH (2021a).

⁸³ ECLI:NL:RBAMS:2021:4308, Amsterdam Rechtbank, 18 August 2021, para. 2.5.

⁸⁴ *Ibid.*, para. 4.2.

⁸⁵ *Ibid.*, para. 2.4.

⁸⁶ *Ibid.*, para. 4.2.

⁸⁷ *Ibid.*, para. 4.8.

⁸⁸ *Ibid.*, para. 4.9.

⁸⁹ *Ibid.*, para. 4.10.

⁹⁰ *Ibid.*, para. 4.10.

⁹¹ *Ibid.*, para. 4.10.

COVID-19 policy was also an elaboration of Google's right to property: in principle, it can set the rules that apply to its platform, including the rule that content that conflicts with its COVID-19 policy is to be deleted. As such, the right to property can also serve as a "legitimate restriction" on the freedom of expression of others.⁹²

Importantly, the Court emphasized that courts can only intervene when obstacles prevent "any effective exercise of freedom of expression", or if the "essence of the right has been destroyed".⁹³ However, the Court held that this was not the case, as the claimants could also bring their opinions to the attention of the general public in other ways, with numerous "other channels" at their disposal.⁹⁴ As such, YouTube does not have an obligation to allow the statements of claimants which are in violation of its COVID-19 policy, nor to facilitate the claimants' contribution to the public debate. The Court concluded that there had been no violation of the claimants' freedom of expression.

3.3. Politicians social media accounts shut down for disinformation

Following the previous judgments, a different District Court delivered a landmark judgment concerning a politician's LinkedIn account being shut down, where the Court ordered LinkedIn to restore the account of a Dutch MP that had been closed under its disinformation policy.⁹⁵ However, the Court refused to order that LinkedIn reinstate specific posts published by the politician which had been removed. The case again involved van Haga, a Dutch MP, who was critical of the government's COVID-19 measures. Van Haga had maintained a LinkedIn account for over a decade, and had posted many messages via his account, including messages critical of COVID-19 measures. In mid-2021, LinkedIn informed the MP that his account had been permanently restricted due to a series of posts concerning COVID-19 which had been deleted for violating LinkedIn's rules on disinformation. The posts included: "The IFR (Infection Fatality Rate) of Corona is slightly higher, but comparable to the #IFR of flu", "Kids don't get sick from #COVID19 and asymptomatic contamination is close to zero", and "It remains strange that #Ivermectin does work in other countries, but this drug may not be used in the Netherlands".⁹⁶

Following the closing of his account, the MP initiated legal proceedings against LinkedIn, seeking to have his account restored and his posts re-uploaded. The Court first dealt with the issue of the closing of the MP's account. At the outset, the Court made a significant finding, holding that while LinkedIn was a private party, its parent company (Microsoft) had "responded" to the European Commission's call for online platforms to "prevent disinformation about COVID-19".⁹⁷ As such, the Court held that the case concerned a restriction on freedom of expression at the "instigation of the government, of a

⁹² Ibid., para. 4.10.

⁹³ Ibid., para. 4.11.

⁹⁴ Ibid., para. 4.11.

⁹⁵ ECLI:NL:RBNHO:2021:8539, Noord-Holland Rechtbank, 6 October 2021, <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBNHO:2021:8539>; see also, Ó FATHAIGH (2021b).

⁹⁶ ECLI:NL:RBNHO:2021:8539, Noord-Holland Rechtbank, 6 October 2021, para. 4.31.

⁹⁷ Ibid., para. 4.5.

type of information considered undesirable by the government (harmful disinformation about COVID-19), via a certain type of channel (social media platforms)”; and that had to be taken into account when determining the “freedom to be left to the platform” in that regard.⁹⁸ Further, the Court noted that the user agreement between LinkedIn and the MP was a “continuing performance agreement”, and under the Dutch Civil Code, such agreements had to include requirements of “reasonableness and fairness”, including in relation to termination of the agreement.⁹⁹ Crucially, the Court strongly criticized LinkedIn’s communication with the MP about the closing of his account based on his posts, holding that its communication had been “substantively inadequate”, “insufficiently informative” and had contained “no motivation” apart from a “single reference” to the user agreement on disinformation.¹⁰⁰ Therefore, the Court held that the termination of the user agreement had occurred “without due care”, and ordered LinkedIn to reinstate the MP’s account.¹⁰¹ However, the Court did emphasize that the MP “must adhere” to the conditions that LinkedIn imposed on the use of its platform.¹⁰²

Second, the Court considered the content of the MP’s deleted posts and referred to the case law of the European Court of Human Rights; in particular, the court judged that where damage might be caused by statements of facts, there had to be a “sufficient factual basis”.¹⁰³ Crucially, the Court held that LinkedIn had had “good grounds” to find that the MP’s posts had contained “harmful disinformation”, and to find that dissemination of that information could “diminish the willingness of readers to follow well-founded advice and adhere to prescribed measures”.¹⁰⁴ As such, the Court refused to order that LinkedIn reinstate the MP’s deleted posts concerning COVID-19 measures. Finally, the Court ordered LinkedIn to restore the MP’s account within three days of the judgment’s publication.

Thus, the District Court, while not restoring the deleted content, established a significant principle: platforms must guarantee users due process when closing accounts based on disinformation rules. Furthermore, it was established that where platforms fail to afford users such due process, the courts can step in and order platforms to restore a user’s account.

4. Conclusion

This chapter has sought to provide an overview of the current regulation of online disinformation in the Netherlands, and a number of concluding points can be made. First, the regulatory framework is heavily influenced by the principle that freedom of expression and freedom of the press “must take center stage” when adopting policy on disinformation and “qualifying disinformation as such”, and factchecking are “*not* primary government duties”. Crucially, disinformation is considered “harmful, but often legal”; expression and disinfor-

⁹⁸ Ibid., para. 4.8.

⁹⁹ Ibid., para. 4.15.

¹⁰⁰ Ibid., para. 4.22.

¹⁰¹ Ibid., para. 4.24.

¹⁰² Ibid., para. 4.24.

¹⁰³ Ibid., para. 4.29.

¹⁰⁴ Ibid., para. 4.35.

mation is “part of the freedom of speech”; and as such, the Netherlands does not have a specific legislative provision prohibiting the dissemination of disinformation. Second, the twin pillars of the government’s disinformation strategy are (a) strengthening public debate (in particular, promoting media pluralism and media literacy); and (b) reducing the influence of disinformation, through proportionate responses to disinformation campaigns. Third, a significant additional pillar for tackling disinformation is the Ministry of the Interior’s trusted flaggers with the largest online platforms, who facilitate flagging content as disinformation, especially around elections. Fourth, the regulatory framework applicable to information is mainly that of self-regulation, including the Code of Conduct Transparency Online Political Advertisements, Dutch Advertising Code, and Guidelines of the Press Council. However, there has been significant case law from Dutch courts relating to disinformation and online platforms. The case law demonstrates that platforms have considerable freedom to remove content based on disinformation rules contained in their terms of service, including removing content from elected politicians during parliamentary sessions. However, Dutch courts have stepped in, and where platforms fail to afford users such due process, the courts may go so far as to order platforms to restore a user’s account.

References

- European Regulators Group for Audiovisual Media Services (2021) = European Regulators Group for Audiovisual Media Services: *Notions and concepts of disinformation*. Strasbourg, ERGA, 2021. https://media-board.europa.eu/document/download/00eeb7c2-1abc-443f-9ead-f103b6b577f0_en?filename=ERGA-SG2-Report-2020-Notions-of-disinformation-and-related-concepts-final.pdf
- Ó FATHAIGH (2021a) = Ronan Ó FATHAIGH: Google not required to reinstate Dutch MP’s YouTube video on Covid-19. *IRIS*, 2021-9:1/29.
- Ó FATHAIGH (2021b) = Ronan Ó FATHAIGH: LinkedIn ordered to restore Dutch politician’s account closed over Covid-19 disinformation. *IRIS*, 2021-10:1/21.
- Ó FATHAIGH (2022) = Ronan Ó FATHAIGH: Facebook not required to reinstate posts by prominent public figure. *IRIS*, 2022-1:1/11.
- Ó FATHAIGH (2023) = Ronan Ó FATHAIGH: *New government-wide strategy on disinformation*, *IRIS*, 2023-3:1/9.
- Ó FATHAIGH et al. (2021) = Ronan Ó FATHAIGH – Tom DOBBER – Frederik ZUIDERVEEN BORGESIU – James SHIRES: Microtargeted propaganda by foreign actors: An interdisciplinary exploration. *Maastricht Journal of European and Comparative Law*, Vol. 28. No. 6. (2021) 856–877. <https://doi.org/10.1177/1023263X211042471>
- VAN HOBOKEN et al. (2019) = Joris VAN HOBOKEN – Naomi APPELMAN – Ronan Ó FATHAIGH – Paddy LEERSSEN – Tarlach MCGONAGLE – Nico VAN EIJK – Natali HELBERGER: *The legal framework on the dissemination of disinformation through Internet services and the regulation of political advertising*. Amsterdam, Institute for Information Law, 2019.

Case study: Slovakia

ANDREJ ŠKOLKAY

1. Introduction

This chapter explores how Slovakia tackled online disinformation from regulatory perspective and analyses the operating practices in place during the years 2018–2023. It also covers regulatory changes (following Digital Services Act) impacting social media platforms, video sharing platforms, and search engines, resulting from the media regulator's expanding competences from the summer of 2024 onward (some of which have been in effect only since June 2025).

It should be noted that the strategic and tactical efforts for combating disinformation were rather diverse and became increasingly intensive. However, some of them were also arguably unconstitutional. Moreover, it remains questionable whether authorities can or cannot actually block websites as of 2024. The legislation, amended by Parliament in 2022, is rather confusing in that regard. The fact is that the state has not blocked any website on the basis of broadly understood political arguments, since 2022 (as of July 2024). The exception is a pan-European decision taken to block certain Russian websites, as will be discussed later.

This chapter offers an overview of the key regulations, key actors and key regulatory outputs (including drafts and temporary regulations which were only in force for a brief time).

Why was tackling online disinformation seen as a key issue to address in Slovakia? The blocking of some local online news and current affairs outlets was seen as controversial from a constitutional perspective (freedom of speech and of the press) by some lawyers and NGOs. Within this context, it could also be said that the causes of relative popularity and relative ease of penetration in society of disinformation and fake news/hoaxes, selection of regulatory approaches how to tackle this issue, were by and large based on incorrect assumptions. Specifically, it was assumed that online disinformation is capable of having a significant impact on the population at large. However, such a claim – that propaganda can always have immediate and direct impact on humans without any pre-conditions – is not supported by almost a century of research into propaganda.¹ Of course, this does not contradict the idea that “small effects suddenly matter a great deal”.² This issue is illustrated further on an example of fake video used in the campaign before the 2023 general elections.

It should be acknowledged that fake news and hoaxes are a political problem, especially for a society that is inclined to believe in conspiracy theories. In fact, in the region, Slovak

¹ HAMID (2022); ANDERSON (2021); URSPRUNG (1994).

² SANDER VAN DER LINDEN: Psychological weapons of mass persuasion. *Scientific American*, 10 April 2018. <https://www.scientificamerican.com/article/psychological-weapons-of-mass-persuasion/>

respondents are the most likely to believe in conspiracy theories. Sixty percent of respondents believed that world events are decided by secret groups and associations that strive to create a totalitarian world government. Half believe that the Jews have too much power and covertly control the world. More than half the Slovak population believes that its values are threatened by the USA. Almost two thirds of respondents (also the highest figure in the region, expressed their readiness to give up their freedoms in exchange for other benefits. Respondents who were prone to believe conspiracy and disinformation narratives also tended to prefer autocratic leaders and were more willing to give up their democratic freedoms in exchange for other benefits. Such as improving the financial situation, security, or preserving traditional values.³

It has already been established in scientific literature that political discussions on social media often lead to (greater) polarisation (echo chambers) and, especially on Facebook, to (even more) anger.⁴ It is also the case that ‘alternative media’⁵ are most often accessed via Facebook.⁶

Considering these points, it should not be surprising that trust in the traditional media and authorities is low overall, among a large part of the public. In the years 2017–2023, overall trust in the media oscillated below 27%. Respondents expressed higher trust (37%) in the media that they followed. Only two monitored media – public service media *RTVS* (*Radio and Television of Slovakia*, now *Slovak Television and Radio* or *STVR*), and news television *TA3*, reached over 50% trust.⁷ However, other research suggested that these two media channels attained significantly higher levels of trust (absolute trust and more likely trusted) – with trust levels of 74% and 71% respectively.⁸ In a similar vein, two thirds of both online users and offline users noticed fake news in the sources they followed.⁹

Some other media, like online only news portal *aktuality.sk* (*Hot News*), business newspaper *Hospodárske noviny* (*Business Newspaper*), radio *Rádio Express*, commercial *TV Joj*, and regional and local newspapers, reached trust levels of almost 50%, according to one study.¹⁰ In another study, they reached over 50% trust level (absolute trust and more likely trusted).¹¹ However, it is doubtful that “Very low trust (27%) in news in general reflects decades of interference by business and political leaders”,¹² as the authors of the report claimed. In particular, there was no direct business or political interference in public service media or the majority of other major media outlets over the last decade. In fact, there are other more plausible explanations available for explaining low trust in the media. In addition to those already mentioned, this include the fact that trust in political institutions is

³ HAJDU–KLINGOVÁ (2020)

⁴ OSMUNDSEN – BANG PETERSEN – BOR (2021), SPOHR (2017), ANSTEAD (2021).

⁵ There are many definitions of alternative media. However, in local discourse these are understood mostly normatively negatively, as those outlets (almost exclusively online) that do not follow journalistic standards. Furthermore, Facebook bans most controversial alternative media outlets.

⁶ ŠTEPANOVÍČ (2021)

⁷ CHLEBCOVÁ HEČKOVÁ – SMITH (2023)

⁸ 2MUSE (2021)

⁹ 2MUSE (2021)

¹⁰ CHLEBCOVÁ HEČKOVÁ – (2023)

¹¹ 2MUSE (2021)

¹² CHLEBCOVÁ HEČKOVÁ – SMITH (2023)

low in general, with maximum trust found in case of local governments (trusted by 36%).¹³ For the media, less than 20% of respondents believed that media activities are somehow useful for society.¹⁴ Moreover, 61% of respondents thought that it was complicated for them to check whether they follow trustworthy news.¹⁵ There is also an issue with the liberal attitude of the key mainstream and agenda-setting media, given that the majority of population is ideologically and culturally more conservative in many respects.¹⁶ This is related to a human need to share politically congruent information. It could be that a key turning point is when the media by and large failed to cover large-scale corruption – which could be seen as an early phase of state capture.¹⁷

There is also overall low trust in society. About 10% of Slovaks – the lowest results among OECD members – believed they can trust others in 2014. The average for positive answers for OECD member countries was 36%.¹⁸ However, more recent data has suggested that general interpersonal trust has been stagnant for a long time – it has been oscillating around 23% for the last three decades. Moreover, trust in (some) institutions was much higher.¹⁹ Pro-Russian attitude (e.g. delivery of weapons to Ukraine is more often seen as provocation towards Russia, rather than help for Ukraine)²⁰ can also be explained by deeper and long-term cultural factors, as opposed to successful Russian propaganda.²¹

As a result of the above-mentioned deeper and long-established social-cultural factors and trends, ‘alternative’ media or ‘alternative sources’ in general, are popular among certain relatively limited segments of population. Additionally – and perhaps more importantly – dissemination of disinformation is widespread via social media. As mentioned above, messages produced by ‘alternative media’ are most often accessed via Facebook.²² Moreover, with the exception of *Zem & vek* (The Earth & Age) monthly journal, alternative media are not distributed in printed format (except for occasionally published special issues, for example, *Hlavné správy* – The Headlines or the Main News). There is a ratio about 1:2 of those who prefer ‘alternative’ media (such as *Hlavné správy*, *Zem a Vek*, *Slobodný vysielač*, *Infovojna*) and ‘standard’ media (such as *SME [We Are]*, *Denník N [Newspaper N]*, *Aktuality [Hot News]*).²³ However, this means that about 13% of the population follows alternative media a few times a week; while almost 28% of the population prefers standard media. Yet, at the

¹³ Ako sa máte, Slovensko? Dôvera v inštitúcie 2. [Trust in institutions]. (Survey), *Ako sa máte, Slovensko?* <https://www.akosamateslovensko.sk/tema/dovera-v-institutcie/>

¹⁴ Ibid. It could be said that the loss of trust in ‘mainstream’ media is a global trend, however, these figures are still higher than those in the USA. Megan BRENAN: Media Confidence in U.S. Matches 2016 Record Low. Gallup, 19 October 2023. <https://news.gallup.com/poll/512861/media-confidence-matches-2016-record-low.aspx>

¹⁵ 2MUSE (2021)

¹⁶ ŠKOLKAY (2022)

¹⁷ ŠKOLKAY (2018)

¹⁸ Spectator staff: Slovaks are among the least trusting people. The Slovak Spectator, 2 March 2017. <https://spectator.sme.sk/c/20473017/slovaks-are-among-the-least-trusting-people.html>

¹⁹ DEKK Institute: *Trends of [dis]trust 2023*. Bratislava, DEKK Institute, 2023. <https://www.dekk.institute/wp-content/uploads/2023/10/en-dekk-report-trendy-nedovery-2023-web.pdf>

²⁰ HAJDU et al. (2023)

²¹ Dušan FÁBIK: Prečo Slovensko podlieha ruskej propagande? DenníkN, 22 August 2023. <https://dennikn.sk/3532347/preco-slovensko-podlieha-ruskej-propagande/?ref=tit>

²² ŠTEPANOVÍČ (2021)

²³ Standard media are locally informally defined as those that follow standard professional-ethical journalistic norms.

same time, almost 70% of followers of alternative media also follow standard media.²⁴ This data may help to explain why authorities decided to approach tackling online disinformation by banning some of these alternative media outlets, as we shall discuss further.

Therefore, probably all initial efforts and policies aimed at countering disinformation were based on false premises – ie. that popularity of disinformation is primarily the result of effective foreign (especially Russian) subversive propaganda. We shall see that such erroneous arguments and followup policies were often used to justify (or could be detected in background discussion and in normative justifications) regulatory – and other counter-measures. It is such measures that we will be discussing next.

First, we shall discuss the regulatory background. -Then, we shall discuss selected institutions and stakeholders who were involved in handling disinformation; we shall discuss their powers, obligations, tools and the procedures they used. Finally, a few negative examples of excessive scope in fighting misinformation shall be presented.

2. Regulatory overview

There are numerous strategic and conceptual documents (including draft-laws that were never adopted) produced by the authorities since 2018. There was temporary, severe and opaque regulation of local online news sources that were seen as ‘harmful’ to national security; this could actually be invoked ex officio by the National Security Authority (NSA or NBU) ex officio.²⁵ It is still not clear, or at least it has not been unambiguously interpreted by legal professionals, whether NSA can actually still block some websites (in addition to those related to cybersecurity incidents) upon the request of other parties. There is still hard-law regulation in place against some external (Russian) news sources following the invasion of Ukraine. This latter regulation followed pan-European *ad hoc* regulation.²⁶ We will discuss these issues later on.

Co-regulation of disinformation in online media – in particular social media – is by and large tackled at a supranational (European Union) level. However, there are some attempts which focus at regulating fake news and hoaxes in online media. We will discuss these later on, especially those that seem to be operating on longer and more professional basis.

There are quite a number of state institutions which tackle – or are supposed to tackle – disinformation in general, as a part of their activities. Some of them (e.g. the police) specialise (or rather specialised, following the change in governmental policies after the 2023 early general elections) in the fight against online disinformation in particular, as will be discussed further.

²⁴ Piško (2020)

²⁵ Local: Zoznam subjektov blokovanych EÚ. <https://www.nbu.gov.sk/zoznam-subjektov-blokovanych-eu/>

²⁶ EU: Council Decision (CFSP) 2022/2478 of 16 December 2022 amending Decision 2014/512/CFSP concerning <https://www.nbu.gov.sk/urad/o-urade/hybridne-hrozby-a-dezinformacie/19163-2/index.html> restrictive measures in view of Russia’s actions destabilising the situation in Ukraine. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.LI.2022.322.01.0614.01.ENG&toc=OJ%3AL%3A2022%3A322I%3ATOC>; Council Decision 2014/512/CFSP of 31 July 2014 concerning restrictive measures in view of Russia’s actions destabilising the situation in Ukraine [consolidated text]. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02014D0512-20221204>

It should be noted that before the 2023 early general elections, state authorities (e.g. NSA), together with the European Commission and a US Senator, warned against possible online public disinformation campaigns, and put pressure on major platforms to prevent possible dissemination of the most problematic fake news and hoaxes (especially those based on AI – deep fake videos) (wire agency TASR [2023]²⁷, national security authority NBU [2023]²⁸). Indeed, such deepfake videos did appear towards the end of the electoral campaign (in particular, during the electoral blackout period).²⁹ One of these deepfake videos was critically reflected upon in post-elections period by Věra Jourová, the vice-chairperson of the Commission and the Commissioner for transparency and values.³⁰ From a legal regulatory point of view, deepfake videos highlighted a problematic enforcement of the blackout period.³¹ Interestingly, the police investigation concluded that “no sane person could ever believe its content”.³² In fact, this fake video was massively disseminated via Telegram, Facebook, TikTok, Instagram, YouTube, but also via WhatsApp, before the general election. Wider dissemination of this fake video was limited as a result of intervention of specialised governmental units entrusted with the fight against disinformation.³³ This case has not been closed at the time of writing – the prosecutor has re-opened the investigation into this case, and Reporters without Borders has even suggested that such messages should be criminalised.³⁴

Finally, it should be recalled that a new government lead by Robert Fico (since October 2023) has radically modified or strengthened the coordination of state policies towards disinformation, which we will be discussing in the next section. This included a new Concept of Strategic Communication of the Slovak Republic, approved by the Government in

²⁷ TASR: V USA žiadajú Metu, aby pred voľbami v SR nešírila ruské dezinformácie. *Teraz.sk*, 13 September 2023. <https://www.teraz.sk/spravy/americky-senator-vyzval-metu-aby-pred/740397-clanok.html>

²⁸ Varovanie v súvislosti s voľbami do NR SR 2023, <https://www.nbu.gov.sk/varovanie-v-suvvislosti-s-volbami-do-nr-sr-2023/>

²⁹ Filip STRUHÁRIK: Šíria sa falošné nahrávky politikov. Niekoľko môžu pobaviť, podstatné je, že majú ovplyvniť voľby. *DenníkN*, 29 September 2023. <https://dennikn.sk/3595174/siria-sa-falosne-nahravky-politikov-niekoľko-mozu-pobavit-podstatne-je-ze-maju-ovplyvnit-volby/>

³⁰ Robert BARCA: Údajná nahrávka telefonátu predsedu PS a novinárky Denníka N vykazuje podľa expertov početné známky manipulácie. *AFP Slovensko*, 28 September 2023. <https://fakty.afp.com/doc.afp.com.33WY9LF>; Česká Tisková Kancelár: Slovenské voľby provázala bezprecedentní dezinformační masáž, myslí si Jourová. *Aktuálně.cz*, 1 October 2023. <https://zpravy.aktualne.cz/zahranici/slovenske-volby-provazela-bezprecedentni-dezinformacni-masaz/r-59a28b50605f11eea25a0cc47ab5f122/>

³¹ See Martina HILBERTOVÁ – Michal PÍŠKO – Ľuboš KOSTELANSKÝ: Moratórium dezinformácie nezastavilo, mohli tak ovplyvňovať voličov, *DenníkN Blog*, 7 December 2023. <https://dennikn.sk/blog/3625432/moratorium-dezinformacie-nezastavilo-mohli-tak-ovplyvnovat-volicov/>

³² Vladimír ŠNÍDL: Polícia odmietla vyšetrovať predvolebný deepfake. Nikto súdny mu vraj nemohol veriť. *DenníkN*, 14 January 2024. <https://dennikn.sk/3777044/policia-odmietla-vysetrovat-predvolebny-deepfake-nikto-sudny-mu-vraj-nemohol-verit/?ref=list>

³³ Vladimír ŠNÍDL: Podvrh rozhovoru predsedu ps pred voľbami ukázal, že deepfake slovenčinu nezvláda. onedlho to umelá inteligencia zvládne. *DenníkN*, 6 October 2023. <https://dennikn.sk/3608268/predvolebny-deepfake-nebol-kvalitny-proti-proruskym-kanalom-zakrocil-stat-coskoro-to-moze-byt-horsie/>

³⁴ RSF: As the investigation into a Slovak journalist Monika Tódová's 'deepfake' is reopened, RSF is calling for this type of attack to be criminalised. *RSF*, 6 March 2024. <https://rsf.org/en/investigation-slovak-journalist-monika-t%C3%B3dov%C3%A1-s-deepfake-reopened-rsf-calling-type-attack-be>, see also interview with journalist Tódová: #RSFTalks: Journalist Monika Tódová, victim of a deepfake to influence the election in Slovakia. *RSF YouTube*, 6 March 2024. https://m.youtube.com/watch?si=FJfhiGa4iqGjzXjc&fbclid=IwAR26CptfoRYSveKJpy3yHeBlhirsL_u1tf0Hyp-k5QQGWaXeUqKsws1MjE&cv=LXhX3Elvpo&feature=youtu.be

March 2024 (although the document itself is dated 2023). In this document, it was explicitly stated by the Government that the goal was to prevent a situation in which “NGOs would perform strategic communication of the State on their own, without accepting responsibility for the impact on Slovak society”.³⁵ Instead, the key role was taken over by the ‘Coordination Organ for Strategic Communication at the Office of the Government of Slovak Republic’.

2.1. State regulation

A temporary regulation on online disinformation was passed immediately after the Russian invasion of Ukraine, but this was short lived and was only in force during spring and summer of 2022.³⁶ As mentioned, there is legal ambiguity over whether the NSA can still block some websites with controversial content (with allegedly harmful content of political nature), at the request of other state authorities. The NSA provided only a general answer on that point, citing law, but not clearly answering this question when it was sent via e-mail.³⁷ (A pan-European ban on selected Russian media is still in force.) It should be noted that “taking action against media companies who broadcast infringing content has so far been the exclusive competence of Member States. War, as a special situation, has been exempted from this rule under the Regulation and under other exceptional circumstances.”³⁸

By contrast since 2023, the NSA has not been able to block online disinformation *ex officio*. As mentioned, this is in fact a contested issue. Some lawyers, e.g. Mesarčík³⁹ agree that although the NSA cannot block websites *ex officio*, specified third parties can actually ask the NSA to block websites. It should be clarified that there is no question of whether the NSA can fully or partially ban or limit usage of certain products (e.g. some online applications, processes, or services related to e.g. cybersecurity). We shall discuss this former controversial issue later on in section 2.1.2.

The Manifesto of the Government (2021–2024)⁴⁰ already included a plan to tackle dissemination of disinformation and hoaxes through criminal law.⁴¹ The plan was described as follows: “The Government of the Slovak Republic will examine the increase in the protection of interests of the Slovak Republic through criminal law standards against the risk of hybrids threats, especially related to the spread of propaganda and disinformation, support for extremists groupings and cyber attacks.”⁴² There is no legal definition of

³⁵ Koncepcia strategickej komunikácie Slovenskej republiky. UV SR (2024). https://www.hybridnehrozby.sk/wp-content/uploads/2023/08/koncepcia_strategickej-komunikacie_sr.pdf

³⁶ Then valid § 27b and § 27c Act 69/2018 on Cybersecurity.

³⁷ E-mail letter No. 04940/2024/SRD/ORM-002, dated 05.03.2024, signed by plk. Ing. JUDr. Alexandra Kaľavská Dianišková.

³⁸ KOLTAY (2023) 52.

³⁹ MESARČÍK (2023)

⁴⁰ As mentioned, there were early general elections in September 2023.

⁴¹ Programové vyhlásenie vlády Slovenskej republiky na obdobie rokov 2021–2024 (PVVSR 2021–2024). <https://www.nrsr.sk/web/Dynamic/DocumentPreview.aspx?DocID=494677>

⁴² Ibid., 11.

disinformation. Legal expert Mesarčík⁴³ believes that this lack of legal definition of disinformation is actually a good thing. In his view, “[i]t could happen that the definition would be too broad or too narrow and would not capture the cases we want to regulate”. In any case, previous temporary regulation, as well as various drafts usually did not use the term ‘disinformation’ specifically or exclusively, but, approached this issue inconsistently. For example, they used the terms ‘online threats’ or ‘hybrid threats’ or ‘harmful content’ or used other linguistic terms as described further on in this paper. However, it should be noted that the Constitutional Court (III. ÚS 288/2017/112) has found that there is a difference between ‘disinformation’ versus ‘untruth’. The Constitutional Court has also differentiated between ‘disinformation’ versus ‘information for the purpose/targeted information’ (účelové informácie) (PL. ÚS 26/2019).⁴⁴

The Manifesto of the Government reflected “concern of the Government with the increase of disinformation in the media and in particular on social networks, which is for the purpose of evoking fear and distrust in state and liberal democracy. Therefore, the Government [...] will support initiatives fighting hoaxes, fake news and other disinformation, as well as creating mechanisms to fight hoaxes, fake news and other disinformation.”⁴⁵ The Manifesto of the Government included many mentions which reflected concerns about disinformation as well as suggesting measures against disinformation, in various sectors of the government.

Indeed, there had been discussion about at least three drafts (including one submitted with substantial changes at a later stage) that were intended to regulate controversial online content, especially – in effect – on social media. However, neither of these proposals (draft laws) materialised during previous governments lead by Igor Matovič (2020–2021) or by Eduard Heger (2021–2023), or during the provisional government lead by Ľudovít Ódor (2023). We will discuss these draft proposals on regulation of disinformation later on in section 2.1.3. since these are interesting case studies. They both illustrate the efforts and arguments leading up to statutory regulation. As mentioned, the Cybersecurity Act (69/2018 Z.z., 2022 updated versions) actually regulated fake news and hoaxes under a temporary regulatory approach in 2022. This temporary legal regulation, which is probably no longer in force (at least not used since mid-2022) will also be discussed here. There was a plan to continue regulating disinformation further (including maintaining the option of banning some alternative media) but the Parliament did not approve a draft proposal to that effect of the Cabinet in 2022 (UV-42058/2022).⁴⁶ Among the suggested changes was that the Supreme Administrative Court should be involved as guarantor of legality, and NSA should publish all decisions on its website. The current version [§ 27b sec (7)] includes the right to appeal to a court. It also mentions that this appeal (to the administrative court) has no immediate/postponing effect on the decision in question. However, no blocking order have been issued by the NSA since March 2022.

⁴³ Veci verejné: Prepis diskusie Pozrime sa na to na TA3: Môžeme hovoriť a písať čo chceme? aj s Romanom Michelkom. Veci verejné, 22 November 2023. <https://web.archive.org/web/20231122145154/https://veci-verejne.sk/prepis-diskusie-pozrime-sa-na-to-na-ta3-mozeme-hovorit-a-pisat-co- chceme-aj-s-romanom-michelkom/>

⁴⁴ MESARČÍK (2023): 30

⁴⁵ PVVSR 2021–2024: 130.

⁴⁶ Návrh zákona, ktorým sa mení a dopĺňa zákon č. 69/2018 Z. z. o kybernetickej bezpečnosti a o zmene a doplnení niektorých zákonov v znení neskorších predpisov – nové znenie, UV-42058/2022, 26.10.2022. <https://rokovania.gov.sk/RVL/Material/27764/1>

Not well known within Slovakia and therefore also worth discussing here is a regulation of specific content in the Press Act (Act on Publications), passed after the Russian invasion of Ukraine (and mentioned with reference in footnote 6 above). This is actually transposition of pan-EU blocking legislation into domestic legislation.

2.1.1. Temporary regulation of fake news and hoaxes

The National Security Authority (NSA or NBU) was able to block either ‘harmful content’ or ‘harmful activity’ throughout half of the 2022 and (still may) fully or partially ban or limit usage of certain products (e.g. some online applications, processes, or services, due to cybersecurity concerns) (Act 69/2018 Z.z., updated versions). Following the Russian invasion of Ukraine, emergency regulation was swiftly passed [including Act 55/2022 Z.z. on Certain Measures Concerning Situation in Ukraine, as an update to Act 69/2018 Z.z., § 5, sec (1) letter *ag*), and in particular § 27b, § 27c] that allowed blocking of certain websites for a few months. In particular, blocking was targeted at [Act 69/2018 Z.z., § 27b, sec (3)] malicious/harmful content or malicious/harmful activity.

“Malicious content means a program resource or data that causes or may cause a cyber security incident. Malicious activity means any activity that causes or may cause a cyber security incident, fraudulent activity, theft of personal or sensitive data, serious misinformation and other forms of hybrid threats.”

However, due to heavy criticism originating among some lawyers, the opposition and NGOs, this regulation did not continue in force beyond September 30, 2022 [Act 231/2022 Z.z., an update to Act 69/2018 Z.z., § 27b, sec (9)].

The NSA was able to decide whether to block “harmful/malicious content or harmful/malicious activity that can cause serious disinformation or other forms of hybrid threats” [§ 27b sec (3)]. However, neither ‘disinformation’, nor ‘other forms of hybrid threats’ were defined in § 3 – ‘Clarification of Key Terms’. Interestingly, the NSA published an informal ‘Short Dictionary of Hybrid Threats’ on its web.⁴⁷ This online dictionary defines and explains ‘disinformation’ as well as ‘hybrid threats’.

The NSA was able to execute such decisions or block certain websites, following a request by a third person [§ 27c sec (1)], [further specified as a public authority, in effect meaning state authority, see also § 27b sec (8)], or ex officio decision [§ 27b sec (1)]. Interestingly, this activity could be carried out not only in cooperation with state authorities [§ 27b sec (8)], but also in cooperation with natural and legal persons – e.g. internet service providers [§ 27b sec (5), (6)]. The NSA should decide about blocking (according to the law) in such a way that blocking would be effective, expedient, and adequate to possible risks related to blocking [§ 27b sec (4)]. As mentioned, NSA was also able to decide about blocking ex officio, initially up until June 30; this was later extended until September 30, 2022 (Act 231/2022, an update to Act 69/2018 Z.z.). This regulation was de iure based on emergency Act 55/2022 (passed on February 26, 2022, just two days after the invasion began), which allowed and specified in article III, resulting in changes to Act 69/2018 Z.z. (arguably

⁴⁷ Krátky slovník hybridných hrozieb, <https://www.nbu.gov.sk/kratky-slovník-hybridnych-hrozieb/>

unconstitutional activity, as will be discussed later on) the blocking of certain websites by the NSA, initially effective until June 30, 2022.

The decision to block had to include [§ 27b sec (2)] identification of the authority, identification of the person running the infrastructure to be used for blocking, identification of harmful content or harmful activity, the reason for blocking, the method of blocking, terms for execution of blocking, duration of blocking, and options for ‘de-blocking’; finally, the decision had to include legal-administrative instructions (such as the possibility for appeal, if relevant). This should have prevented arbitrariness on the part of the NSA. If a third legal person were to ask the NSA for blocking [§ 27c sec (1)] of a certain subject, the administrative procedure conditions would be almost identical, and it would be specified that “the reason for blocking which should be an executable decision of requested authority” were to be included [§ 27c sec (2) letter e)].

Similarly, costs associated with blocking – if such blocking was requested by a third person – were split. While damage caused by *execution* of blocking were charged to the NSA, costs associated with the act of blocking as well as *the cost of any damage caused by blocking* were charged to the applicant [(§ 27c sec (7))]. This suggests that immediate and much lower costs were to be covered by the NSA, while long-term costs of such decision were to be covered – if such long-term costs arose – by the applicant.

The measure of blocking ‘harmful content’ passed in the spring and then extended in the summer of 2022 appears to no longer be in force (except those blockings related to ‘cybersecurity incidents’). However, as mentioned, this is legally controversial or at least a confusing issue. In § 27b, sec (9), it is stated that “Making decision on blocking of harmful content or harmful activity is possible only until 30 September 2022”. In § 27c sec (1) it is stated that “The NSA may also carry out blocking on the request of a subject according to a special document”. In a footnote, examples of these ‘special documents’ are given – the Act 46/1993 Z.z. on Slovak Intelligence Services and the Act 30/2019 Z.z. on gambling. This later Act 30/2019 Z.z. on gambling refers to Act 30/2019 Z.z. on electronic communications that [§ 41, sec (6)] stipulates that “an enterprise is obliged to limit access to website whose address is mentioned in court order”. However, this type of website blocking is carried out under the auspices of the Gambling Regulation Authority.

For execution of blocking based on the Act on cybersecurity [§ 27c sec (9)] “instructions of § 27b sec. (5)–(7) will be used appropriately”. In turn, these instructions are mostly of a technical nature. However, § 27b sec (7) also states that the decision to block is subject to appeal before the court, and administrative lawsuits do not have a delaying effect (*nemá odkladný účinok*). Thus, on the one hand, the NSA *may block at the request of another subject*. On the other hand, it *can not decide* about blocking. Ultimately, it is § 27b sec (2) which clearly sets out the formal requirements for issuing a blocking decision. Yet this § 27b sec (2) is in contradiction with § 27b sec (9), since the latter does not allow the NSA to act at all. So does this mean that the NSA may – in accordance with § 27c sec (1) – perform blocking without actually issuing a decision on a given case? This seems strange, but at the same time, § 27c sec (9) states explicitly that from among the provision of § 27b, only § 27b sec (5)–(7) are to be applied for performing blocking.

In any event, there has been a rather fierce discussion about the legality and constitutionality of these measures, as we shall discuss later on.⁴⁸ It should be noted that § 27a still allows “placing limits on the use of products, processes of services of a third party”. This is different level of regulation (in contrast to ‘blocking’ as mentioned in § 27b and in § 27c). There is also a justification for ‘banning or limiting’ when something “threatens the security interests of the Slovak Republic” [§ 27a, sec (1), letter b)]. In effect, theoretically, the NSA can still ban or limit certain use e.g. the use of news applications.

2.1.2. Regulatory efforts that did not materialise (draft legislation)

In addition to temporary regulation directly implemented by the NSA as discussed above (Act 55/2022 Z.z. as part of the Act on Cybersecurity) and actual blocking of selected news and current affairs websites, there were regulatory efforts that never made it into law. Due to space limitations, we shall not discuss *ad hoc* legal regulatory initiatives by MPs here, such as an initiative by deputies from OĽANO, SaS and Za ľudí parties in April 2021.⁴⁹

The first publicly available governmental draft text (an update to the CriminalCriminal Code) aimed at regulating fake news and hoaxes was a draft proposal prepared by the Ministry of Justice under Mária Kolíková (*Za ľudí – For the People Party*). It was a part of wider updates in criminalCriminal legislation. The draft was first submitted for public discussion in late 2021, and again (in a new version) in 2022. The draft proposal was justified primarily by COVID-19 pandemic controversies about the safety and efficiency of vaccination. However, this seemed to be just the reaction to the culmination of a longer trend of spreading conspiracy theories, fake news and hoaxes which had started some time earlier.

The specific section of the proposal (§ 361, part of larger update to Criminal Code) was called ‘The offence of Spreading Untrue Information’. Specifically, this crime was supposed to include “production and dissemination of knowingly untruthful information which is capable of causing a risk of serious concern/alarm (*znepokojenia*) at least among part of the population in a certain place, threatening the life or health of inhabitants or influencing inhabitants in their decision-making on serious issues of societal importance (*celospoločenského významu*)”. The suggested maximum criminalty for such a crime was up to 10 years imprisonment. For example, even repeated cases of negligence dissemination of such ‘untrue information’ should be punishable up to a year imprisonment.

It should be noted that the CriminalCriminal Code (Act 300/2005 Z.z.) already includes ‘the offence of fearmongering’ (§ 361). The text of this paragraph is as follows: “Whoever willfully causes a risk of serious disturbance to at least part of the population of any place by spreading an alarming report which is false, or who commits other similar conduct

⁴⁸ It should be mentioned that at least one these previously blocked websites changed its territorial internet-based domicile as to avoid or complicate further attempts at blocking.

⁴⁹ For example, TASR: Koaliční poslanci chcú predĺžiť zákaz webov, ktoré považujú za dezinformačné. *Pravda*, 8. April 2022. <https://spravy.pravda.sk/domace/clanok/622943-koalichni-poslanci-chcu-predlzit-zakaz-webov-ktore-povazuju-za-dezinformacne/>; Ivana ŠALGOVÁ: Boj štátu s hybridnými hrozbami Nevypíname neviniatka, ale štváčov, vraví Krúpa. Weby bude možné blokovať trvalo. *Postoj*, 25 April 2022. <https://www.postoj.sk/104458/nevypiname-neviniatka-ale-stvacov-vravi-krupa-blokovanie-webov-moze-byt-trvale>

capable of causing such a risk shall be punished by imprisonment of up to two years.” Depending on further conditions, the criminality may be of up to eight years imprisonment. Moreover, § 362 that states “Whoever, in a crisis situation of the state, causes a serious risk, even through negligence, anxiety, small-mindedness or defeatist mood in at least part of the population of some place, by spreading an alarming message, shall be punished by imprisonment from six months to three years.” Thus, this is even more strict regulation, as it is sufficient to be negligent in this case.

In any event, a draft proposal intended to introduce the ‘offence of spreading untrue information’ was criticised by many lawyers, NGOs as well as by both special and general prosecutors. For example, Maroš Žilinka, prosecutor general, called this draft “absurd, ambiguous and empirically impossibility to prove”. Therefore, it was seen as inappropriate proposal according to legal standards, as well as an inadequate regulatory reaction to the current situation in a society.⁵⁰

The second draft proposal with the strange working title ‘Act About Measures to Prevent Spreading Not-Allowed Content’ was aimed at “measures to prevent dissemination of unallowed content”.⁵¹ It was submitted for public discussion (PI/2022/83) by Veronika Remišová (Za ľudí – For the People Party), the then vice-prime minister and minister of investments, regional development and digitalisation in April 2022. The inspiration for ‘technical’ aspects of regulation was taken from Germany. The justification for such a proposal was found in two cases. In the first case, a student attacked fellow students with an axe. Moreover, in another hate crime, a student randomly killed LGBT+ persons, while expressing hatred also towards some local politicians and Jews. Both youngsters were supposed to have become radicalised on social media with conspiratorial content.⁵² However, a few months earlier justification for such a regulation was more standard – it included negative examples of disinformation which appeared during campaigns before the US elections in 2016, Brexit, the Russian annexation of Crimea, and campaigns by well-organised fascist movements, as well as allegedly many people who have been misled, including the most vulnerable groups, such as youngsters and seniors.⁵³ It appears that these regulatory efforts included previous consultations or initiatives at the EU level, as well as bilateral talks with representatives of Meta Company.⁵⁴

⁵⁰ TREND: The Criminal Act of Spreading False Information is Absurd and Inappropriate, Says Žilinka to Kolíková. *TREND*, 3 December 2023. <https://www.trend.sk/spravy/m-zilinka-novy-trestny-cin-sirenieniepravdivej-informacie-je-nenalezity>

⁵¹ Vicepremiérka Remišová: Nenávisť na sociálnych sieťach rozkladá demokratickú spoločnosť – nový zákon zavedie pravidlá pre platformy, ktoré takto zarábajú milióny eur. *MIRRI*, 20 April 2022. <https://mirri.gov.sk/aktuality/digitalna-agenda/vicepremierka-remisova-nenavist-na-socialnych-sietach-rozklada-demokraticku-spolocnost-novy-zakon-zavedie-pravidla-pre-platformy-ktore-takto-zarabaju-miliony-eur/>; PI/2022/83 Predbežná informácia k návrhu zákona o opatreniach proti šíreniu nedovoleného obsahu (012942/2022/OL-3), (PI/2022/83), 19.04.2022; <https://www.slov-lex.sk/legislativne-procesy/SK/PI/2022/83>

⁵² Branislav CABAN: Ministerka Remišová chce zákonom regulovať sociálne siete, rozdeľujú podľa nej spoločnosť. *Touchit*, 9 November 2022. <https://touchit.sk/remisova-chce-regulovat-socialne-siete/450791>

⁵³ Vicepremiérka Remišová na Globsecu 2022: Sociálne platformy musia prebrať väčšiu zodpovednosť za škodlivý obsah, ktorý šíria. *MIRRI*, 2 July 2021. <https://mirri.gov.sk/aktuality/podpredsednicka-vlady/vicepremierka-remisova-na-globsecu-2022-socialne-platformy-musia-prebrat-vacsu-zodpovednost-za-skodlivy-obsah-ktory-siria/>

⁵⁴ Adriána Belej MAJERČINOVÁ: Remišová sa stretla s predstaviteľmi Facebooku. Rezort chystá právnu reguláciu mesiace. *Pravda*, 21 October 2022. <https://spravy.pravda.sk/domace/clanok/644463-remisova-sa-stretla-s-predstavitelmi-facebooku-nas-rezort-chysta-pravnu-regulaciu-digitalnych-platforiem/>

With regard to the type of content which was intended to be regulated by this draft law, there was a rather broad focus. It was supposed to have included almost all existing criminal offences such as sexual abuse, blackmail, making threats, (electronic) harassment, spreading alarming news, defamation of nation, race, and belief or incitement to national, race, or ethnic hatred, treason, collaboration with enemy or espionage. Thus, this was supposed to be *lex generalis*. It was planned to be used when there was a case of spreading forbidden content via electronic services used for sharing content with other users. The draft law focused on ‘services’ with higher number of users. The responsible entity was supposed to be a service provider and it was the service provider who was supposed to check controversial content, once it was reported to them. If the content was found to be against the rules, the service provider was supposed to prevent further dissemination.

In case of violation of this obligation, the state would have been able to impose monetary sanctions on service operators, impose the obligation to remove illegal content and, in extreme cases, even block the service provider nationwide, as it has already done, for example, with the disinformation portal *Hlavné správy* (translated as Headlines or Main News).

The regulation was intended to introduce an obligation for electronic service providers to allow users to verify their identity and mark their profiles as official, as is the case today, e.g. on social networks (previously Twitter, now X, or Instagram). Such profiles should be clearly marked as verified, so that it was clear that it was not a fake profile impersonating another person or company.

For the purposes of preparing the draft law, the Ministry also suggested a definition of the terms ‘disinformation’ and ‘hoax’.

“The law should distinguish between them according to the following definitions: disinformation:

1. demonstrably (or obviously) false, misleading or manipulatively presented information that is intentionally created, presented and disseminated with the clear aim of lying or misleading, capable of causing some harm or securing some benefit;
2. often contains an element that is obviously true, which gives it credibility and can thus complicate its detection;
3. misinformation does not include inadvertent errors in reporting (factual error), satire and parodies, or biased reports and comments that are clearly marked as such;

hoax:

1. a virally spread fake news or deliberate lie;
2. usually has three typical features – (i) urgency, (ii) reference to an illusory authority and (iii) request for further dissemination;
3. often the intention is to induce alarm or fear.”⁵⁵

These definitions were criticised during the public legislative process as “having included elements which are obviously true, which give it credibility and can thus complicate its detection; misinformation”. Such a definition of disinformation can be applied to any manipula-

⁵⁵ Stanislav VINČ: Štát chce regulovať sociálne siete. Ak nezmažú hoax, môže ich pokutovať. *Techbox*, 20 April 2022. <https://www.techbox.sk/stat-chce-regulovat-socialne-siete-sirenje-hoaxov-by-mal-pokutovat>

tive information which could lead to confusion in terminology, argued the Office of the Government and Chancellery of the Security Council.⁵⁶ Furthermore,

“at the same time, the need to define the term ‘hoax’ is redundant, as it basically meets the definition of disinformation (‘deliberate lying’). Also, the typical features of a hoax mentioned in the text (‘urgency, reference to illusory authority, request to spread’) are features applicable to the concept of disinformation.”⁵⁷

The Office of the Prosecutor General demanded, as a ‘fundamental request’, to delete the words

“dissemination of untruthful information according to § 361a and 362”. In contrast, the Ministry of Foreign and European Affairs suggested adding “in addition to the term information operations also to use a term disinformation, alternatively (also include) disinformation activity (pôsobenie) or deliberate dissemination of untruthful and misleading information.”

Additional criticism of the draft law included its redundancy in view of existing measure or measures that were at that time in process of introduction, such as more active communication by state authorities towards citizens. The Slovak Chamber of Attorneys argued that § 361 of the Criminal Code already covers such crime (dissemination of untruthful information). The Ministry of Culture as well as (then) Broadcasting and Retransmission Board (now Media Services Board) pointed out that there was already a draft law on Media Services (transposed EU Directive 2010/13/EU) undergoing a second reading in the Parliament. The draft law on media services introduced proceedings to prevent the spread of illegal content on a content-sharing platform (in particular, § 112, § 151, § 152 and § 153). Moreover, the Ministry of Culture argued that it had exclusive competence to regulate content in general. The Ministry of Interior (Hybrid Threats Centre) suggested to re-consider the use of the term ‘illegal content’ itself, as it falls exclusively under the jurisdiction of the courts to assess whether the elements of the criminal offense are fulfilled, as well as to take a decision on guilt and punishment. Finally, it was also seen as essential to re-assess whether the proposed legislation limits the constitutional right to freedom of speech and the free dissemination of information.

The third proposal was submitted for public and professional discussion in March 2023, under the working title ‘LP/2023/129 Act on Measures to Increase Security and Trustworthiness of Platforms in Online Environments’. The author of this draft proposal was once again the Ministry for Investments, Regional Development and Information lead by Veronika Remišová. There were 184 comments raised in the public legislative process.⁵⁸

⁵⁶ Predbežná informácia k návrhu zákona o opatreniach proti šíreniu nedovoleného obsahu. *Slov-Lex*, 2022. https://www.slov-lex.sk/legislativne-procesy?p_p_lifecycle=0&p_p_col_id=column-2&_processDetail_WAR_portletsel_action=pplpZobrazPripomienky&p_p_id=processDetail_WAR_portletsel&_processDetail_WAR_portletsel_cisloLP=PI%2F2022%2F83&p_p_col_count=1

⁵⁷ Ibid.

⁵⁸ LP/2023/129, Zákon o opatreniach na zvýšenie bezpečnosti a dôveryhodnosti platforiem v on-line prostredí a o zmene a doplnení niektorých zákonov. <https://www.slov-lex.sk/legislativne-procesy/SK/LP/2023/129/pripomienky/zobraz?page=2>

The draft included a definition of disinformation (§ 2) as “unambiguously untruthful information, created, presented and disseminated with the aim of deceiving society (klamať verejnosť) or a certain group of persons, the consequence of which is or may be damage, or detriment or contribute to benefit (zabezpečenie prospechu)”.

Clearly, there were many formal and specific objections against these draft laws, so let us return to broader and currently valid regulatory options.

2.1.3. Currently valid legal regulations on fake news and hoaxes

There is no specific local hard law regulation tackling fake news and hoaxes directly or generally in the online environment in Slovakia. However, there have always been legal regulations that tackle what are broadly understood as being fake news and hoaxes, either online or offline. These legal tools include public personal or institutional libel (civil or criminal law), or defamation of race, nation, or religion (criminal law), or publishing ‘a false or incomplete factual claim’ (the press law and the Act on Media Services), or broadcasting ‘illegal content’ (formerly audiovisual regulation, now the Act on Media Services) to name a few typical examples.

However following the Russian invasion of Ukraine, the Act on Publications (‘the Press Law’) includes a specific ban on the dissemination of certain information, as defined by state authorities or supranational bodies (the EU). This recent regulation should also of interest to us. It covers specifically online versions of offline publishing activities. In effect, it introduces, or transfers supranational censorship into the national domain. Let us first discuss a more general law which covers video-sharing platforms.

The Act on Media Services (264/2022 Z.z.)

This act applies to a broadcaster of a radio programme service, retransmission services and multiplex providers, if they have a registered office or place of business or residence in Slovakia. There are more detailed conditions for television broadcasters and (§ 7) video-sharing platforms (including § 9 ‘content-sharing platforms’, if there is the aim to make profit, e.g. works protected by the Copyright Act). The regulator (MSB) is also entitled (§ 110) to assess the appropriateness of measures taken by video sharing platform provider to protect the public.

A 2024 update (Act passed on 27 June 2024, without numbering) gives regulatory rights over content sharing platforms as well as engine search providers [§ 108 sec (1)] to the MSB.⁵⁹ Moreover, § 108 sec (6) specifies that central state authorities and public administration in general have to cooperate with the MSB. The MSB is referred to as the ‘coordinator of digital services’, which includes audiovisual services on demand, content sharing platforms and internet search engines – in addition to those legal subjects mentioned earlier. In

⁵⁹ Zákon z 27. júna 2024, ktorým sa mení a dopĺňa zákon č. 264/2022 Z. z. o mediálnych službách a o zmene a doplnení niektorých zákonov (zákon o mediálnych službách) v znení neskorších predpisov a o zmene a doplnení niektorých zákonov, https://www.slov-lex.sk/ezbierky/pravne-predpisy/SK/ZZ/2024/203/vyhlasene_znenie.html.

[§ 109 sec (2)], it is stipulated that the regulator performs supervision of whether legal documents that tackle content sharing platforms are followed. Perhaps redundantly, this is further highlighted in § 109 sec (3), letter *c*), where state regulation of content sharing platforms is mentioned. However, both these legal rights and duties of the MSB [§ 109 sec (2) and § 109 sec (3), letter *c*)] have only been in force since 28 June 2025. Finally [§ 110 sec (3), letters *w*)–*y*)] entrust the MSB to fulfill the tasks of coordinator of digital services, to participate at work in European Committee for Digital Services, and, among other tasks, to entrust, temporarily postpone or remove status of trustworthy reporter (understood as concerning problematic content online).

The Act on Media Services is obviously of more general purpose than just regulating fake news or hoaxes. Yet it should be noted that even before 2024 (2025 respectively) there was a regulatory sea-change §110 sec (3) letter *q*), which allows cooperation with platforms with respect to their content moderation practices and in an effort to enforce their respective ‘community’ standards. This cooperation should aim at the efficient, proportional and non-discriminatory application of services’ standards. In fact, until 2024 or 2025 respectively, the MSB could only take action if potentially illegal content was being spread online. These cases of problematic content include child pornography, extremist materials, posts inciting terrorism or national, racial and ethnic hatred, posts denying or approving the Holocaust and crimes against humanity, or posts defaming a nation, race or belief. Obviously, some may argue that, for example, some extremist material is at the same time disinformation (or fake news and hoaxes). In such case, however, the argument should be made (if one turns to this legislation with respect to tackling disinformation) that it is – for example – in the first instance, output which is ‘extremist’ or ‘defaming a nation, race or belief’. Before users turn to the regulatory authority with a request for intervention, the operators of the website on which the illegal content is being spread must be notified (§ 152). This is relevant where potentially illegal content is found on a content-sharing platform, or in particular for electronic periodic publications, news webportal, or wire agency portal [§ 152 (2)].

Whether the content breaches the law will be decided by the board’s three-member panels. If the Board comes to the conclusion that the content in question is illegal and at the same time its dissemination threatens the public interest or represents a significant interference with individual rights citizens, it will issue a decision to prevent its spread. If the platform operators do not remove the illegal content and do not prevent it from spreading further, they may be fined between 2,500 and 100,000 euro by the Board.

During the period 1 August 2022 – 10 November 2023, the Media Services Board dealt with 53 cases of possible online illegal content. However, only in 2 cases were official legal-administrative procedures initiated. The first case concerned a blog post published on the *hlavnespravdy.sk* portal, which questioned the number of victims of the Holocaust and at the same time contained hateful comments interwoven with conspiracy theories about members of the Jewish ethnicity and their role on a global scale (AO-RPMS/2022/00254). The second case (AO-RPMS/2023/00143) tackled comments under a video posted by a municipality on the YouTube video sharing platform to news channels. The video concerned the exclusion of members of the Roma ethnic group from a public transport vehicle. A comment containing hate speech appeared under this video, calling for the extermination of this ethnic minority.

However, as could have been expected, disinformation cases did not fall under the competences of the Media Services Board. For example, in the case AO-RPMS/2023/01084, the Board dismissed the case (presenting pro-Russian propaganda on social network) due to a lack of any illegal component. The possibility that disinformation was included was insufficient.⁶⁰

As previously mentioned, some forms of fake news and hoaxes have been dealt in general terms in § 212. “If a false or incomplete statement of fact was mentioned in the broadcast program, which affects the honor, dignity or privacy of a natural person or the good reputation of a legal entity, on the basis of which the person can be precisely determined, this person has the right to reply (strictly linguistically, in Slovakia it is called ‘the right to expression’).”

The Press Act (The Act on Publications 265/2022 Z.z.)

This act (officially it is called the Act on Publications) regulates the obligations of the publisher of a periodical publication, the operator of a news web portal and a news agency when disseminating information, and obligations of the publisher of a non-periodical publication when distributing publications. It also tackles self-regulatory mechanisms. As mentioned, § 8 defines the Right to Expression. Specifically,

“If a false or incomplete factual claim has been mentioned in a periodical publication or agency service which interferes with the honour, dignity or privacy of a natural person or with the reputation of a legal person, on the basis of which the person can be accurately identified, that person has the right to have the statement published. The publisher of a periodical publication, the operator of a news web portal and a news agency are obliged to publish a statement at the request of this person.”

Furthermore, there is § 9 and the Right to Supplementary Notice:

“If a periodical publication or agency service containing communiqués of a journalistic nature or communiqués of a journalistic nature contains a factual claim of proceedings before a public authority against a person on the basis of which the person can be precisely identified, and this proceeding has been terminated by a final decision, that person shall have the right to request the publication of a supplementary notice of the final outcome of those proceedings.”

Interestingly, a new § 4 section 2 defines that:

“The publisher of a periodical, the operator of a news web portal and a news agency are obliged to ensure that they do not allow the end user or the public access to content the provision of which is prohibited to persons, entities, bodies or their representatives or intermediaries under a special regulation directly or through a periodical publication published or made available to the public.”

⁶⁰ MESARČÍK (2023): 112–117.

In a footnote, this content is defined as “for example, Articles 2f and 12 of Council Regulation (EU) No 833/2014 of 31 July 2014 concerning restrictive measures in view of Russia’s actions destabilising the situation in Ukraine (OJ L 229, 31.7.2014), as amended.” In other words, there is an obligation for censorship of certain types of information as requested by the state authorities or by supranational authorities. There are sanctions, too, in case of breach – notification of a violation of the law or a fine of up to € 20,000. One might wonder whether this is in line with the Constitution that ‘forbids censorship’ (Article 26, sentence 3). It is true that the next sentence (4) of the Constitution allows for exceptions to ban on censorship: “Freedom of speech and the right to seek and disseminate information can be limited by law if the measures in a democratic society are necessary to protect the rights and freedoms of others, the security of the state, public order, and the protection of public health and morals.” Moreover, it is stated in Article 13 (2) of the Constitution that: “The limits of fundamental rights and freedoms can be adjusted under the conditions established by this Constitution only by law.” Thus, although this censorship was established by the law, does it really fulfill all other conditions as defined in Article 26(3) of the Constitution? Slovakia was not at war with Russia, and it is highly questionable whether a few Russian disinformation webportals can be seen as a threat to “the security of the state, public order, and the protection of public health and morals.” Unlike in the case of right to expression, or right to additional announcement, the purpose of this institute is not the protection of the personal rights of individual persons (or good reputation of a legal entity), but rather the protection of the public interest. From a constitutional perspective, in each case it must be determined which particular ‘public interest’ is to be protected and if measures taken are appropriate. At a first glance, one may argue that this provision does not impose any new restrictions on publishers or news agencies, as this is already given by the applicable legislation defining or determining ‘prohibited content’.

The Cybersecurity Act (Act 69/2018 Z.z.)

As mentioned, the Cybersecurity Act is the only act which has tackled fake news and hoaxes explicitly and relatively systematically in Slovakia – albeit for a temporary period in 2022 – but it did so in what was arguably an unconstitutional manner. Currently, cooperation and providing information “important for securing cybersecurity and solving cybersecurity incident” for fulfilling tasks due to a cybersecurity event is still compulsory for public authorities, as well as for providers of ‘basic services’ and ‘legal entities’ as defined in Appendix 1 of the Act on Cybersecurity. There is a caveat: it is only obligatory to provide such information which is ‘necessary’ for resolving the cybersecurity incident. However, Appendix 1 lists not only essential ‘digital services infrastructure’ but practically all industrial sectors (such as transport, energy sector, etc).

Furthermore, as previously mentioned, the NSA can ban or limit the use of certain products (e.g. some applications), processes, or services, if it finds that they do not enable, or in a fundamentally hamper cybersecurity (§ 27a). It can be assumed that this ban may include social media applications. For this regulation to apply, there must be a risk to life and health of persons, the economy of the state, public order, or property .

Furthermore, these potential threats to cybersecurity include “the security interests of Slovakia”. In such situation, the NSA (§ 27a) can act *ex officio*, or following a request from another public authority (here the term used is ‘public’, not ‘state’). Information on initiating such regulatory initiative must be published at least 30 days prior totaking action, in a single information system for cybersecurity, as well as on the NSA website. Interestingly, during this ‘waiting’ period the NSA can not issue a decision. Any decision has to be preceded by analysis of risks (in cooperation with the Ministry of Foreign Affairs, Ministry of Economy, Ministry of Interior, Slovak Intelligence Services, and Ministry of Defense. A draft proposal is then submitted to the Security Council and the Government [Cabinet of Ministers]). The opinion of the Government is binding for making any decision.

There is another condition – the decision of the NSA is conditioned by activity or lack of thereof of the third party in question. In other words, the decision will be issued only if a provider of basic services or a third party, do not make the requested corrections within a reasonable deadline as set by the NSA. Finally, if there is a decision by the NSA, this must be published in the Collection of Laws. There is time limitation for such total or partial ban. The NSA has to stipulate a reasonable period for banning or limiting usage of certain product, process, or service. This period cannot be longer than for two years. However, it is possible to issue such a full or partial ban repeatedly. There are also sanctions – these are between 300 euro up to 1% of overall turnover, with a maximum in value of 300,000 euro.

With respect to blocking controversial content, it is instructive to mention that the NSA just a year earlier (i.e. in 2021) explicitly stated that it does not have a legal right to collect content of communications. Moreover, the NSA stated that all blocking has to follow administrative law procedures. The NSA argued that it regularly informs the Special Supervising Committee of the Parliament in annual reports, as well as providing any other requested information. The NSA argued that all its decisions are subject to review by the Constitutional Court. Finally, blocking on the request of another legal subject was originally requested by the Ministry of Economy. It was meant to be a supplementary tool in cases concerning consumer protection – with reference to a European regulation. Blocking on request from a third party was accepted by other state authorities such as the Ministry of Defense, Ministry of Interior and Consumer Protection Authority.⁶¹

2.2. Conceptual regulatory documents

The authorities have enacted mostly cross-sectoral documents which partly or entirely tackle disinformation (as, for example, one of the identified it as ‘hybrid threats’). However, it should be mentioned that some of them are no longer valid:

⁶¹ NBÚ: Reakcia NBÚ na medializované informácie k novele zákona o kybernetickej bezpečnosti. *NBÚ*, 31 March 2021. <https://www.nbu.gov.sk/reakcia-nbu-na-medializovane-informacie-k-novele-zakona-o-kybernetickej-bezpecnosti/>

*2.2.1. The Concept of the Slovak Republic's fight against hybrid threats (2018)*⁶²

This concise (six pages) conceptual document summarised an emerging institutional framework for tackling the issue of hybrids threats. The main purpose of the document was to set up a mechanism for information exchange and evaluation. It also defined basic indicators for hybrid threats, of which online disinformation was merely one of many. The document stated that on the one hand, not a single department possesses material, technical and human resources for full coverage of this issue. On the other hand, it argued, hybrid threats demonstrate cross-sectoral and inter-institutional features. The document summarised the role of the two most important overarching (coordinating) bodies: the national contact point for hybrid threats – Situational Center (SITCEN) – and the national cooperation centre for hybrid threats, the National Security and Analytical Centre (NBAC). Which institution actually drew up this document and when is not clearly stated.

2.2.2. The Manifesto of the Government (2020–2024) (PVVSR 2021–2024)

The most important task in this field was “to prepare an action plan to coordinate the fight against hybrid threats and the spread of disinformation and build adequate central capacities for its implementation”. As previously mentioned, the Government announced its aims to “examine the increase in the protection of interests of the Slovak Republic through criminal law standards against the risk of hybrids threats, especially related to the spread of propaganda and disinformation, support for extremists groupings and cyber attacks”. Furthermore, the Manifesto stated that the Government “will support initiatives aiming at fighting hoaxes, fake news and other disinformation”. The Ministry of Defense was asked “to take into account the changing security environment in the context of cyber and hybrid threats, misinformation, climate change and challenges resulting from technological development, including AI”. There was also a plan to update the content of education and training towards critical thinking, with the aim of creating better resistance to conspiracy theories and misinformation.

*2.2.3. The Action plan for coordinating the fight against hybrid threats for the period from 2022 to 2024 (NBU [2022])*⁶³

This 22 page document was drawn up by the Ministry of Defence. It includes the following sections: General tasks, Systematic measures, Strategic communication and disinformation, Building the resilience of society, Protection of electoral processes and Foreign influence. The Action Plan represents a reaction to “the security risks associated with foreign investments in critical infrastructure, media, or the academic sector, influencing election process-

⁶² Konceptcia pre boj Slovenskej republiky proti hybridným hrozbám. NBU, 2018. <https://www.nbu.gov.sk/data/att/1126.pdf>

⁶³ <https://www.nbu.gov.sk/akcny-plan-koordinacie-boja-proti-hybridnym-hrozbam/>

es, strategic corruption, or the action of disloyal paramilitary groups, with increased emphasis on subversive influence activities of actors, using propaganda and disinformation or spreading hate speech or extremist content”.

2.2.4. Security Strategy 2021

This strategy is the key strategic document for state security. It explicitly names the readiness to effectively respond to hybrid threats as one of its strategic interests and contains a commitment to increase Slovakia’s resilience.⁶⁴ Specifically, it defines among its ‘strategic security interests’ resistance of the state and society to security threats as well as citizens’ trust in a democratic state and independent public institutions and social cohesion. This is in response to “the increase in importance and scope of non-military threats and procedures”. The Security Strategy highlights that

“The number and scope of subversive and coercive activities by various actors is increasing using disinformation and propaganda to disable or manipulate decision-making mechanisms in the state, to influence public opinion in their favor and destabilise the political situation... The public is exposed to the increasing spread of misinformation and conspiracy theories that can threaten human health, subvert cohesion society or provoke public violence and social unrest.”

2.3. Self and co-regulation of online disinformation

In the light of the lack of an overarching effective hard law regulation for disinformation in online environment, a grass-root process has developed some attempts to create self or co-regulation. However, the overview of the annual reports of the main or potentially the most important self-regulatory body – the Press and Digital Board of the Slovak Republic (TRSR) – suggests that it has not yet got involved in any activity that would tackle disinformation in online news and current affairs webportals. This may be partly explained by the very essence of self-regulation, in as much as that very much depends on the good will of those involved. In addition, self-regulation of digital (and print) media in Slovakia can be seen as having a confusing design. Although according to Section I, subsection 1, letter *a)* of the Code of Journalistic Ethics, this Code is binding “for natural persons and legal entities that issue periodical or non-periodical print in print or digital version intended for distribution in the Slovak Republic”, at the same time, those natural persons and legal entities have to first publicly declare their willingness to follow this Code of Ethics.⁶⁵ In contrast, and paradoxically, the scope of the Statute of TRSR encompasses virtually all natural persons and legal entities active in this field (without mentioning their prior declaration of willingness to be covered by this self-regulation).⁶⁶

⁶⁴ Bezpečnostná stratégia Slovenskej republiky. *MOSR*, 2021. https://www.mosr.sk/data/files/4263_210128-bezpecnostna-strategia-sr-2021.pdf

⁶⁵ Etický kódex novinára. https://www.aonesr.sk/wp-content/uploads/2025/06/Etický_kodex.pdf

⁶⁶ Štatút tlačovo-digitálnej rady Slovenskej republiky. https://web.archive.org/web/20240629122646/http://trsr.sk/wp-content/uploads/2016/05/Statut_TR_SR-1.pdf

The Code of Ethics of Electronic Media addresses advertising and marketing communication only. It has been in effect since 2010.⁶⁷ Similarly, the Code of Use of Content from the Internet deals only with copyright-related issues.⁶⁸ It should be noted that there is a new self-regulatory initiative: the Ethical Code for Marketing Influencers.⁶⁹ This Ethical Code has been drafted by the Interactive Advertising Bureau Slovakia (IAB Slovakia) and Association of Digital Marketing Agencies (ADMA) in 2022. Potentially, it can be assumed that influencers may disseminate fake news and hoaxes. However, no such case has as yet been identified. Therefore, this initiative shall not be discussed in any further detail here.

By virtue of its nature, the co-regulatory body, the Advertising Council, can only tackle online disinformation on an ad hoc basis, if at all.

Then there remains enforced self-regulation, as represented by the webportal and associated NGO Konšpiratori.sk. This involves trying to indirectly impose some moral and monetary pressure on controversial online news and current affairs portals, as well as many other dubious online resources. Konšpiratori.sk is one of many NGOs and private initiatives involved in enforced regulation or simply in monitoring disinformation in the online environment. Next, we shall discuss only those which were operating in 2022. For example, an anti-hoax initiative by online versions of the quality liberal newspaper *Sme* (hoax.sme.sk), which ended its activities in 2021.

2.3.1. Webportal Konšpiratori.sk (*Conspirators*)

This webportal is a cross-border or two-nation (covering both Czech Republic and Slovakia) attempt to disclose the monetary income from ads in online advertising flowing towards controversial quasi-journalistic online sources, which are seen by some actors as being involved in production and dissemination of fake news and hoaxes; it is operated by an NGO. This is an approach which involves disclosing online disinformation indirectly. The webportal (NGO) provides a list of ‘controversial’ websites and more recently (since September 2023), ‘controversial’ YouTube channels, or, those of a ‘conspiratorial’ nature, as the original name (translated as ‘conspirators’) suggests. In other words, it provides a list of those involved in ‘conspiratory’ style reporting and commenting online. Specifically, it claims that it is “a public database of websites with non-serious, deceptive, fraudulent, conspiratorial or propaganda content”.⁷⁰ Next, the initiative specifies that “we compile a list of websites and YouTube channels where members of our committee have doubts about their credibility and quality of content”.⁷¹ The list includes approximately 300 items. These are categorised from one to 10 (worst). There is also ‘a grey zone’. This grey zone includes websites with an index in the range of 6–4. These domains (about 30) currently allegedly do not

⁶⁷ IABSlovakia: Etický kódex elektronických médií, *IABSlovakia*, 2009. <https://www.iabslovakia.sk/wp-content/uploads/2023/04/IABsk-Eticky-kodex-elektronickych-medii.pdf>

⁶⁸ IABSlovakia: Kódex preberania obsahu na internete, *IABSlovakia*, 2022. https://www.iabslovakia.sk/wp-content/uploads/2023/02/KODEX_preberania_obsahu_24_06_2022.pdf

⁶⁹ The Ethical Code for Marketing Influencers. <https://kodexinfluencermarketingu.sk>

⁷⁰ <https://konspiratori.sk/>

⁷¹ <https://konspiratori.sk/zoznam-stranok>

represent a significant risk for brands, but the Evaluation (expert) Commission is still paying attention to their activities, it continuously observes and reevaluates the content on these webdomains.

Four interesting legal case studies emerged as a result of this initiative. Two of them involved major controversial news and current affairs webportals (*Hlavný denník* and *Hlavnespravy.sk*) and two of these webportals (*Parlamentnilisty.cz* and *AC24*) were actually originally from the Czech Republic.⁷² All four websites sued the initiative based on business law court cases or injunctions. This was in itself an unusual, and initially quite effective self-defense. The final verdicts have been handed down, except in the case of *Parlamentnilisty.cz*. Thus, for *Parlamentnilisty.cz*, there is (at the time of writing) an ‘urgent injunction’ in place. Activity regarding further court proceedings has been slow.

Hlavný denník request that the court issue an “urgent injunction against the NGO for unfair interference into its business activities”. More specifically, it claimed that being included on the list of controversial sources may cause a negative and immediate business impact and could be seen as unfair business practice – in effect, giving advantage to its competitors. Therefore, it asked the court to order an ‘urgent injunction’ that would prohibit including *Hlavný denník* on the list. Lower courts accepted this request. However, *Hlavný denník* did not proceed further with the regular court case for two years, so the Supreme Court considered that this ‘urgent injunction’ had lasted too long. Moreover, the Supreme Court did not see any unfair business practice in issuing such a ‘recommendation’ by the NGO.⁷³ There were no further legal steps taken from the side of *Hlavný denník*.

In a similar case, *Hlavnespravy.sk*, was again placed on the list of controversial websites, after years (2018–2022) of court battles. Interestingly, the court considered the decision and assessment of the National Security Authority (NBU) in its judgement. The NBU argued that this online outlet represents a ‘security threat’.⁷⁴

The court case initiated by the Czech webportal *AC24* attempted to follow identical legal argumentation as it was used in previous cases (business law – urgent injunction). However, in this case both county and regional court in Slovakia dismissed this motion in 2022.

It should be mentioned that the first verdict in which webportal *Hlavnespravy.sk* used the argument of ‘reputation damage’ by Konšpirátori.sk NGO was handed down in October 2023. The lower court dismissed the case as unfounded.⁷⁵ because it accepted monitor-

⁷² Konšpirátori.sk: Konšpirátorom dočasne stiahnuť web hlavnespravy.sk zo svojej databázy. *DenníkN Blog*, 3 March 2020. <https://dennikn.sk/blog/1775813/sud-prikazal-konspiratorom-docasne-stiahnut-web-hlavnespravy-sk-zo-svojej-databazy/>; P Konšpirátori.sk: arlamentnilisty.cz sa súdia s Konšpirátormi. Je to ako žalovať Google za zlé recenzie, odkazujú im autori projektu. *DenníkN Blog*, 4 December 2020. <https://dennikn.sk/blog/2170626/parlamentnilisty-cz-sa-sudia-s-konspiratormi-je-to-ako-zalovat-google-za-zle-recenzie-odkazuju-im-autori-projektu/>; Konšpirátori.sk: Web, ktorý straší peklom a tvrdí, že očkovanie zabíja, musí ísť dole zo zoznamu Konšpirátorov. *DenníkN Blog*, 12 August 2021. <https://dennikn.sk/blog/2501424/web-ktory-strasi-peklom-a-tvrdi-ze-ockovanie-zabija-musi-ist-dole-zo-zoznamu-konspiratorov/>

⁷³ Konšpirátori.sk: Hlavný denník neuspel ani na Najvyššom súde, putuje späť na zoznam dezinfowebov. *DenníkN Blog*, 19 May 2022. <https://dennikn.sk/blog/3378382/hlavny-dennik-neuspel-ani-na-najvyssom-sude-putuje-spat-na-zoznam-dezinfowebov/>

⁷⁴ Konšpirátori.sk: Hlavné správy putujú späť na zoznam dezinfowebov. *DenníkN Blog*, 22 December 2023. <https://dennikn.sk/blog/3161411/hlavne-spravy-putuju-spat-na-zoznam-dezinfowebov/>

⁷⁵ Konšpirátori.sk: Hlavné správy prehrali súd s projektom Konšpirátori.sk. *DenníkN Blog*, 30 October 2023. <https://dennikn.sk/blog/3652929/hlavne-spravy-prehrali-sud-s-projektom-konspiratori-sk/>

ing activity as lying within the limits of freedom of expression. However, the court also used a less important argument: the fact that the NBU had also blocked the website, so there were reasons for monitoring this webportal.⁷⁶ Interestingly, the verdict about the constitutionality and legality of blocking by the NBU (which is part of a separate court case) has not yet been issued. There is an interesting legal analysis of this case study written before the court's verdict.⁷⁷

2.3.2. *The Advertising Council*

The Advertising Council has existed for a much longer time. However, the Advertising Council only signed up to the self-regulatory mechanism in October 2023. This is in line with § 130 of the Act on Media Services which set up a co-regulatory alternative in advertising regulation. This development is potentially relevant here, there are two major TV networks among members of the Advertising Council: the Association of Radio Stations of Slovakia and the oldest news TV TA3, and public media RTVS (now STVR). They all have their own websites, and in most cases presence on social media such as Facebook. Yet it remains questionable whether and how the Advertising Council will become involved in regulating online disinformation effectively, if at all.

2.3.3. *The Press and Digital Board of the Slovak Republic*

In the print sector (including online print) there is the Press and Digital Board of the Slovak Republic (TRSR, in old terms, the Press Council). AONE, the Association for Protection of Journalistic Ethics (the founding body of the Press and Digital Board) has incorporated among its members the Interactive Advertising Bureau Slovakia (IAB Slovakia) in 2017. Since then, the Press and Digital Board, as defined in the founding charter of AONE, seeks: “to supervise the maintenance of ethical principles of journalistic work in print and digital media created in Slovakia (...)”.⁷⁸ The scope of its mission is also specified in § 2 of the Statute of the Press and Digital Board as including among its legitimate targets: “natural and legal persons who run websites (webpages) on which there are published products of journalistic activities of natural persons or editorial offices of periodical and non-periodical press, intended for dissemination in the Slovak Republic”.⁷⁹

The Press and Digital Board discussed ethical issues that appeared in online versions of (primarily) offline publications in three of all six cases which were submitted to it, exclu-

⁷⁶ 30.10.2023, Dnes sudkyňa na prvostupňovom súde rozhodla v náš neprospech v spore s Konšpirátormi, <https://www.hlavnespravdy.sk/dnes-sudkyna-na-prvostupnovom-sude-rozhodla-v-nas-neprospech-v-spore-s-konspiratormi/3304090>

⁷⁷ LANGER (2023)

⁷⁸ Asociácia na Ochranu Novinárskej Etiky: Stanovy Asociácie na ochranu novinárskej etiky v Slovenskej republike. AONE, https://www.aonesr.sk/wp-content/uploads/2016/05/Stanovy_AONE_09.2021.pdf

⁷⁹ Štatút tlačovo-digitálnej rady Slovenskej republiky. https://web.archive.org/web/20240629122646/http://trsr.sk/wp-content/uploads/2016/05/Statut_TR_SR-1.pdf

sively or partially, in 2022.⁸⁰ In 2021, the number of online-only ethical cases reached about the same number – four out of nine cases (of which two were originally submitted in 2020).⁸¹ Finally, in 2020, it was possible to identify four out of 10 cases (three from 2019) as originating exclusively from online publishing activity.⁸² However, none of the above online cases dealt with typical fake news or hoaxes. They merely involved biased, incorrect or unfair reporting.

3. Institutions/stakeholders involved in tackling disinformation: issues, powers, obligations, tools and procedures

In general, it is likely easier to mention authorities and institutions which are *not* involved in handling disinformation in Slovakia.

3.1. State authorities

The list of institutional members of the Committee for Hybrid Threats of the Security Council of Slovakia defines all key state actors that are seen as relevant stakeholders in this field. These are *a)* Ministry of Foreign Affairs and European Affairs, *b)* Ministry of Defense, *c)* Ministry of the Interior, *d)* Ministry of Economy, *e)* Ministry of Finance, *f)* Ministry of Investments, Regional Development and Information, *g)* Ministry of Transport, *h)* Ministry of Health, *i)* Ministry of Justice, *j)* Ministry of Culture, *k)* Ministry of Education, Science, Research and Sports, *l)* Ministry of Labour, Social Affairs and Family, *m)* Office of the Government, *n)* the Authority of Nuclear Supervision, *o)* National Security Authority, *p)* Slovak Intelligence Service, including additional representative of Department for security threats, *r)* Military intelligence, and finally *s)* the Police. Since 2024, the Office of the Government is has been the key coordinating body here, as we shall explain later on.

In general, state authorities must, according to the Action Plan (AP), put “emphasis on cooperation with alliances (ie NATO) partners, centers of excellence, etc. when performing tasks through entrusted state organs”. Furthermore, there has to be an “emphasis on cooperation with the academic sector, i.e. domestic as well as foreign universities, research organisations and professional public in the performance of tasks through entrusted state organs”.

⁸⁰ Výročná správa 2022. https://web.archive.org/web/20240719053756/https://trsr.sk/wp-content/uploads/2016/05/TR_SR_VS_2022.pdf

⁸¹ Výročná správa 2021. https://web.archive.org/web/20240712144058/http://trsr.sk/wp-content/uploads/2016/05/TR_SR_VS_2021.pdf

⁸² Výročná správa 2020. https://web.archive.org/web/20221006045334/https://trsr.sk/wp-content/uploads/2016/05/TR_SR_VS_2020.pdf

Among specific tasks, there was a plan to include strategic communication under the organisational statutes of all ministries and the Office of the Government, by a deadline of the first quarter of 2023.⁸³

Furthermore, state authorities are obliged, as stated in the Concept of the Slovak Republic's fight against hybrid threats, to initially evaluate the reported hybrid threats incidents falling under their competence, and according to the definitions of indicators set out in the Concept of the Slovak Republic's fight against hybrid threats, they then have to forward information about an incident meeting any of the hybrid threat indicators) for further processing, assessment or evaluation by the National Security Analytical Center (NBAC). State authorities also have to provide information additionally to the reported information, which will be evaluated by the NBAC to decide whether it is a hybrid threat. When doing so, state authorities shall follow their own internal regulations or legal regulations for security incidents.

In addition to generally mentioned and specifically defined state authorities, we present selected state authorities which have been specifically designated to tackle disinformation. The following order of appearance of selected institutions does not represent any subordination or indicate the level of importance of those institutions: in fact, the relationships between them are quite complex and intertwined.

3.1.1. The National Security Authority (NSA or NBU) and in particular, its National Security Analytical Center (NBAC)

The NBAC is the national cooperation center for hybrid threats. Its main functions are to collect information where there is a suspicion of a hybrid threat and classify under the evaluation process by creating an NBAC file. After a statement from the state authorities represented in NBAC, in cooperation with SITCEN, it should add 'context' to the relevant NBAC file from the point of view of EU and NATO security interests and further decide whether there is a need to process the output for the recipients.

3.1.2. Center Against Hybrid Threats (CBHH)

The Ministry of Interior (MV) has established a Center Against Hybrid Threats (CBHH) within the Slovakian Police. The task of the CBHH is to increase the resilience of society through activities such as monitoring and processing analytical outputs, especially in relation to the MV. With regard to 'Hoaxes and Scams', the Facebook page of the Police claims to be the most followed page in Slovakia which focusses on misinformation, with almost 150,000 followers in 2023. It has been online since 2018. The selection criteria for hoaxes and frauds (many of which are suggested by Facebook users) are: mass sharing on Facebook (or potential for such mass impact), topic seemed to be interesting and have a clear message addressed to the national public. One of the most common disinformation narratives that it

⁸³ Akčný plán koordinácie boja proti hybridným hrozbám 2022–2024. NBÚ, 2022. <https://www.nbu.gov.sk/data/att/1127.pdf>

debunked was related to foreign affairs: the pandemic is a secret plan by the powerful to rule over or destroy humanity. Significantly, the 2021 Report raised the issue of foreign actor involvement: “It is possible that their actions were trying to support the foreign policy interests of state powers abroad. Foreign state powers tried to spread their narratives through their own media or allied media, or via fictitious independent activists, often communicating in different ways within their own state.”⁸⁴ The Police is also present on Instagram. However, these debunking activities are very controversial and often slip into subjective statements and comments, which are beyond the statutory competence of the Police. Since late 2023, there has been change in focus of the CBHH.

3.1.3. *The Ministry of Justice*

The Ministry of Justice drafted the first governmental draft aimed at regulating fake news and hoaxes *in 2021*.

3.1.4. *The Ministry of Defence*

As mentioned above, the Ministry of Defence has been involved in drafting some strategic documents. In addition, it was assigned to produce an “In-depth vulnerability analysis of selected bodies of the state administration against hybrid threats”, with a deadline of December 2022 (AP). It has established the Department for Hybrid Threats and Strategic Communication in 2022. The main activities of the department include the creation of a situation overview, analysis of the impact of information operations and implementation of reactive and preventive measures intended for mitigating their consequences.⁸⁵ The Ministry of Defence produced analytical material focused on the income of ‘controversial’ websites from commercial activities (ads).⁸⁶

3.1.5. *The Ministry of Foreign Affairs and European Integration*

The Ministry of Foreign Affairs and European Integration is responsible for efficiency, ability to act as well development of tools and initiatives of NATO, the EU and other organisations in the fight against hybrids threats (Action Plan – AP). Moreover, it was the first state ministerial body which established a special unit for this agenda, in the form of the Department for Strengthening Resilience and Hybrid Threats in May 2020.

⁸⁴ See *Police Force Report on Disinformation of the Slovak Republic in 2021*. Communication and Prevention Department of the Presidium of the Police Force, 2022. 5. https://www.minv.sk/swift_data/source/images/slovak-republic-report-dezinfo-2021.pdf

⁸⁵ 2022 Ročenka Ministerstva Obrany SR. MOSR, 2022. https://www.mosr.sk/data/files/5032_rocenka_2022_print.pdf

⁸⁶ ŠTEPANOVÍČ (2021)

3.1.6. The Security Council

The Security Council is a permanent advisory, initiative and coordination body of the Government in the area of state security. The key task of the Security Council is to participate in the creation and implementation of a reliable security system for Slovakia, and to ensure the coordination of measures on the practical implementation of tasks in the field of security and the state's international obligations. The Security Council is expected to prepare proposals for maintaining the security of the state and to prevent crisis situations (including proposals for solutions to such crisis situations). For example, it is tasked with producing an annual report on the fulfillment of the Action plan for coordinating the fight against hybrid threats from 2022–2024. Day to day work is performed by the Office of the Security Council. The Office ensures the fulfillment of coordination, professional and consultative tasks in the field of security for Slovakia.⁸⁷

3.1.7. The Committee for Hybrid Threats of the Security Council of Slovakia

The Committee for Hybrid Threats of the Security Council of Slovakia (as one of the permanent committees of the Security Council) was established in 2023 (Act 97/2023 Z.z. as supplement to the Act 110/2004). The main mission of the Committee for Hybrid Threats is to coordinate activities intended to evaluate the security situation at home and abroad, while taking measures to deal with hybrid action against the interests of the country, and drafting relevant policies in the field of combating hybrid threats. The law defines the tasks of the Committee quite extensively. The Committee for Hybrid Threats is tasked to report on the results of the evaluation to the Office of the Security Council, at least once a year. On behalf of the Security Council, it has to draw up proposals for measures to increase the resistance of the state and society to the risks of hybrid action. It participates in shaping policy and drafting conceptual documents in the field of hybrid threats, as well as in the coordination of interdepartmental and international cooperation in the field of hybrid threats. It also has to make proposals to the Security Council for measures to increase society-wide awareness of hybrid threats. It is expected to discuss proposals submitted to the Security Council, which are related to the performance of tasks in the field of hybrid threats. Finally, it has to prepare expert analyses related to hybrid threats and submit them to the Security Council, and assess drafts of binding legal regulations and international treaties related to hybrid threats, which are submitted for discussion by the Security Council.

3.1.8. The Situational Center of the Slovak Republic (SITCEN)

The Situational Center of the Slovak Republic (SITCEN) is no longer defined as the national contact point for hybrid threats, but since 2024, it has been defined as a “government information analytical workplace with nationwide scope for the field of evaluating changes in

⁸⁷ Štatút Bezpečnostnej rady Slovenskej republiky. https://www.vlada.gov.sk/site/assets/files/1765/s_tatu_t_u_lohy_kbr.pdf

the security environment”⁸⁸ – SITCEN’s roles are therefore wider than previously defined. Hybrid threats are just one of seven broadly defined threats.

3.1.9. *The Office of the Government*

The Office of the Government is also active in this field. For example, the Department for Strategic Communication of the Government Office of the Slovak Republic and the Slovak Debating Association jointly launched a project to support the development of critical thinking and digital literacy for students from five randomly selected secondary schools in various regions of Slovakia in 2023. The topics included digital citizenship, misinformation and hybrid threats, as mentioned earlier, since 2024, SITCON has been in existence with a new mission under the Chancellery of the Security Council. However, employees are paid by the Office of the Government. Similarly, as mentioned, a new Statute of the Office of the Government [Art. 5, letter *m*)] specifies that the Office oversees the activities of the Chancellery of the Security Council. Most importantly, according to the Concept of Strategic Communication of the Slovak Republic, the Office of the Government has been the key coordinating body at a national level since early 2024.⁸⁹

3.2. Non-governmental and private entities

There is a surprising variety of non-governmental and private entities involved in the fight against disinformation. This selection is perhaps representative of this diversity of funding, focus, scope, approaches used and intensity, as well as the timescale for their involvement.

3.2.1. *AFP*⁹⁰

Since 2020, the national agency TASR has been cooperating with AFP’s (Agence France-Press) fact-checking programme (since 2023 there are two fact-checkers). The French agency provides this programme in cooperation with Facebook (Meta) and other relevant partners. The project produces photographs with descriptions, online status, videos and other evidence documenting when a hoax is identified. Once a post is rated

⁸⁸ Štatút Situačného centra Slovenskej republiky. https://www.vlada.gov.sk/site/assets/files/1765/statut_situacneho_centra_slovenskej_republiky.pdf

⁸⁹ Konceptia strategickkej komunikácie Slovenskej republiky. https://www.hybridnehrozby.sk/wp-content/uploads/2023/08/koncepcia_strategickej-komunikacie_sr.pdf

⁹⁰ See Martin HODÁŠ: Overuje správy pre Facebook. Hoax vymyslíte za 10 minút. Vyvraciame ho celé dni [It verifies messages for Facebook. You can make up a hoax in 10 minutes. We spend days refuting it.]. *ZIVE*, 9 May 2020. <https://zive.aktuality.sk/clanok/146760/overuje-spravy-pre-facebook-hoax-vymyslite-za-10-minut-vyvraciame-ho-cele-dni/>

as ‘false’ by a fact-checker, it will appear lower in the News Feed and be displayed with a warning on Facebook.⁹¹

3.2.2. ‘Elfs’

‘Elfs’ or Elfovia.sk (elv.ai) is a start-up focused on monitoring of online comments. It is based on human and machine (AI) monitoring and intervention. The project employs a team of people tasked to remove misinformation, hate speech and block trolls from social media discussions. It cooperates (offers such services) for over 100 websites (for both Czech Republic and Slovakia), mostly media outlets such as private *TV Markíza* and private *Radio Express*. However, moderation also covers some public institutions and non-governmental organisations, including, for example, the Representation of the European Commission. Content moderation takes place in their own software (previously used Zoomsphere), which allows the semi-automatic processing of all comments on the profiles of the media websites concerned. Elf only relies on posts of a few sources (initially it was just a single medium). About 20 percent of online comments require being hidden or deleted. During one or maximum two hours shifts (initially it was a four-hour shift), an elf manages to go through 1,400 to 1,700 posts. In total, about 40,000 comments can be checked daily.

3.2.3. Demagog

Demagog.sk and *Demagog.cz* are fact-checking portals. They primarily check the statements of politicians and other publicly active persons which are made in political discussions or in the public sphere. However, occasionally, they focus on fake news and hoaxes.⁹² In particular, *Demagog.cz* in cooperation with Meta (Facebook) has focused on fake news and hoaxes in Slovakia.⁹³ Since 2010, *Demagog.sk* has analysed more than 700 political debates and more than 15,000 factual statements.

3.2.4. Infosecurity.sk

The information and watchdog platform *Infosecurity.sk* publishes a biweekly report on emerging disinformation trends. The portal also produces a regular podcast series, Disinfo report, which is published on a weekly basis and is unique in that it is the only podcast series focused on disinformation and other forms of non-factual content. The project also strives to raise public awareness and build social resilience by publishing web content (authored articles), a regular weekly newsletter on disinformation, preparing fact-checks in co-

⁹¹ Check permanently updated list of identified fake news and hoaxes at: <https://fakty.afp.com/list>

⁹² See Pavol VALÍK: Umelo vytvorený hlas Michala Šimečku vyzýva na zdrazenie piva. *Demagog.sk*, 27 September 2023. <https://demagog.sk/umelo-vytvoreny-hlas-michala-simecku-vyzyva-na-zdrazenie-piva>

⁹³ See <https://demagog.cz/tag/slovensko>

operation with CEDMO, or publishing reports and research studies with different thematic focus within projects with domestic and foreign partners.

3.2.5. Hoax.sk

Webportal *Hoax.sk* was founded in 2012. It publishes a few examples of hoaxes/fake news annually. In addition, it documents internet scams such identity theft. The project is run by volunteers and it is unique in that it has been operating without any subsidies for over a decade.

3.2.6. Seesame

PR company Seesame created what it claims is the first Facebook chatbot ‘Checkbot’ to help detect fake news on the web. It is based on verification of a link to the article in Facebook chat. Checkbot should recognise the words most commonly used by fake news media and disinformation sites. It is also capable of performing a reverse image search of images used in the article and will report on how many manipulations it found. It is claimed that there were more than 6,000 users, including 50 media companies of this chatbot’s services. Facebook chat users with Checkbot can be warned when a hoax is being spread on the Internet and the bot can be managed by a live admin team as needed. It should be noted that Seesame was behind a fictitious public conflict between two local YouTubers based on a false report in 2017. On the sixth day, a joint video was released where they explained that the purpose of this campaign was to highlight the problem of misinformation and fake news. It was a part of a video series that was targeting 500,000 12–18 years olds.

3.2.7. Globsec

Globsec is a global think-tank with local office in Bratislava. It has a Centre for Democracy & Resilience which conducts research in the field of information operations and hybrid threats. In December 2022, this Centre for Democracy & Resilience became a part of the European Digital Media Observatory (EDMO), which specialises in the fight against disinformation. In addition, Globsec became a participant in the VIGILANT project in late 2022. That project was intended to equip Police Authorities and individual officers with disinformation identification and analysis tools, technology, and knowledge that would make their fight against criminal online content easier and more effective.

3.2.9. CEDMO

The Central European Center for Digital Media Observatory (CEDMO) is a project – a multidisciplinary hub funded by the European Commission. It aims to identify, research and prioritise the most critical sources and causes of information issues in Slovakia (also in Czechia and Poland).

3.2.10. *KInIT*

KInIT is an independent, non-profit institute dedicated to intelligent technology research. The members of KInIT have produced some studies on fake news and disinformation.⁹⁴

3.2.11. *Blbec.online*

Blbec.online is another online project which does not receive any funding. Volunteers create a list of Facebook pages and profiles (over 1,800), for which the contributors have doubts about their trustworthiness and the quality of content. Moreover, it provides the ability to manage one's own discussions on Facebook. There is a unique function such as Autopilot, which automatically hides comments from unwanted people. Another function allows users to set up individualised Facebook monitoring.

Additionally, there have been many ad hoc (short-term) campaigns, studies or projects. For example, there were campaigns run by O2 telephone operator (2021), MEMO98.sk NGO (2022) and hackathon Hacknime.to (2022). A comparative overview report suggested as policy recommendations that the key issue is to increase media literacy and critical thinking.⁹⁵ Therefore, the weakest link among institutions was identified as being the Ministry of Education. In addition, continuous cooperation between non-governmental and governmental sectors was highlighted as very beneficial to all parties.

4. Practice

In this section, we discuss five interesting case studies which reflect the legal-constitutional, political-diplomatic and methodological challenges when in fighting disinformation, as well as the challenges faced by activists.

4.1. Case study: Constitutional challenges with the ban on 'alternative' news and current affairs portals in early 2022

The temporary ban on selected 'alternative' online news and current affairs portals was widely seen as unconstitutional. It should be explained that all these webportals have been placed on the list of 'controversial' websites by konspiratori.sk NGO. On the one hand it is true that these were not typical news and current affairs portals but rather hastily produced collection of usually unverified claims and gossip. On the other hand, it was also true that these webportals tended to publish stories that have often been ignored by the major traditional media (online or offline). This blocking was widely seen as controversial from a legal point of view; specifically because the justification given was too vague and it was in breach of ECtHR case law — *OOO Flavus and others against Russia*. This intervention was taken

⁹⁴ SIMKO et al. (2022); HRCKOVA et al. (2022).

⁹⁵ PISKLOVÁ-SÝKORA (2021)

by the NSA and justified on the vague grounds of “blocking harmful activity”. More specifically, it was stated that the NSA “has identified harmful activity that may produce serious disinformation”. No further specific evidence or arguments were mentioned or made available. These were classified as ‘sensitive’ (dôverné) and ‘secret’. The law did not define ‘serious disinformation’. The director of the NSA further justified the blocking and its scope, arguing that “blocking should be effective, with purpose and able to address possible risks associated with blocking”. Legal experts criticise this measure.⁹⁶

4.2. Case study: Globsec and its methodological problem with the ‘Vulnerability index’

It is argued here that the Vulnerability Index, which was developed as an analytical tool, is misleading. The Vulnerability Index (2021) analyzed the ‘vulnerabilities’ of selected countries to foreign malign influence in five dimensions: public attitudes, political landscape, public administration, information landscape and civic and academic space. There is also a Kremlin Influence Index produced by another international team.

On a scale of 1–100 (0 is the most resilient and 100 the most vulnerable) the Vulnerability Index for 2021 evaluated vulnerability to Russian and Chinese influence in Czechia (at 29 points), Slovakia (at 32) and Hungary (at 44), while data for Poland were unavailable. An earlier Vulnerability Index (2017) identified Hungary (at 57 points) as the most vulnerable country, closely followed by Slovakia (51), then followed at some distance by both Czechia (38) and Poland (30).

Similarly, based on a different methodology, the Kremlin Influence Index (2017) produced the following scores: Hungary 61, Czechia 48, Georgia 54, and Ukraine 49, identifying Hungary as the country most vulnerable to Russian influence over processes in the information space, ie. the production, exchange and consumption of information.

In contrast to the basic assumptions behind the Vulnerability Index and the Kremlin Influence Index, in both cases one sees more of a self-induced vulnerability or congruence of chosen foreign policies with selected foreign actors, rather than the impact of some foreign propaganda/public diplomacy or misinformation and disinformation. This differing analytical perspective is not just a minor challenge to either index. Rather, it has fundamental analytical, and in effect, policy consequences. In that sense, it is also instructive for searching for the proper approach to fighting or tackling disinformation. In short, among analysts there is analytical confusion or an unacknowledged convergence between ‘vulnerability’ and ‘congruence’. Congruence suggests a more active approach and in effect, a policy choice. It also suggests the limited impact of propaganda (or fake news and disinformation). In con-

⁹⁶ Peter ŠAMKO: Blokovanie webových stránok a jeho možný rozpor s judikatúrou Európskeho súdu pre ľudské práva. *Právne listy*, 26 February 2022. <http://www.pravnelisty.sk/clanky/a1062-blokovanie-webovych-stranok-a-jeho-mozny-rozpor-s-judikaturou-europskeho-sudu-pre-ludske-prava>; Martin HUŠOVEC: Súčasný blokovanie dezinformačných stránok je ústavne problematické. Čo s tým? [The Current Blocking of Disinformation Sites is Constitutionally Problematic. What About That?] *Denník N*, 222 April 2022. <https://dennikn.sk/2818631/sucasne-blokovanie-dezinformacnych-stranok-je-ustavne-problematicke-co-s-tym/>

trast, vulnerability paints rather passive actors – possibly because of the huge impact on propaganda and limited foreign policy choices.⁹⁷

4.3. Case study: Educational activism which turns into disseminating disinformation

An NGO, MEMO'98 and Globsec produced two short brochures with the aim to educate people in general, and elderly people in particular about disinformation. The first booklet is called 'War in Ukraine and energy. Questions and answers'. This brochure aims to "raise awareness of the energy crisis and its solutions" and was published by Globsec. The second booklet was published under the title 'Seniors in the light of information'. This brochure was "intended as a basis for conversations on the topic (...) within the framework of discussions with seniors", and was published as part of the project Increasing the resistance of Central European seniors to misinformation by MEMO'98.

A fundamental problem with the text of the first booklet (War in Ukraine and energy. Questions and Answers) is that it claims

"Russia has long used the supply of its mineral resources, especially oil and gas, to exert geopolitical pressure on various countries. Such activities include artificially raising the prices of oil and gas, reducing supplies or completely shutting off the flow of these raw materials, or building new gas pipelines that bypass transit countries of the former Soviet Union with which Moscow had disagreements" (p. 12).

In fact, this basic statement is not true. In the long run, i.e. for decades, Russia (the former Soviet Union) has fulfilled its oil and gas supply obligations. Finally, if Russia were to proceed in a long-term and systematic way in the case of supplies of such raw materials (oil and gas), transported mainly by oil and gas pipelines, as described in the brochure, no one (or only a gambler) would extend existing long-term agreements with it, or conclude new trade agreements. No new gas or oil pipelines would ever be built. Russia would have to pay repeatedly large sanctions for violating trade agreements. Nothing like this is known. Such a procedure could not even be in the interest of Russia, which would needlessly spend huge sums on unnecessary, producing loss making rather than profit making gas and oil pipelines. The examples given by the brochure (two-day and 12-day supply interruptions through some gas pipelines, and further unspecified 'interrupted supplies' over the last 20 years), basically do not confirm, but refute the basic thesis of the brochure (p. 12).

The second booklet is aimed at seniors. What will seniors learn? Among other things, two or three controversial, or partially or completely false statements:

- a) Questioning (by the authors of the publication) the statement that: "Measures to restrict movement and other anti-pandemic measures are a threat to democracy and human rights" (p. 16). In fact, the Constitutional Court declared some measures

⁹⁷ Andrej ŠKOLKAY: Vulnerable 'Vulnerability Index' – or what we got wrong with analysis of Hungary's foreign and domestic policies. *DenníkN Blog*, 13 April 2023. <https://dennikn.sk/blog/3324338/vulnerable-vulnerability-index-or-what-we-got-wrong-with-analysis-of-hungarys-foreign-and-domestic-policies/>

restricting movement during the pandemic to be unconstitutional – parts of the law on the protection of public health, especially those related to mandatory state quarantine elsewhere than in the home environment. Since the decision of the ÚS SR zn. PL. ÚS 4/2021 was issued in 2021, and the cited publication was issued in 2022, it is not true that these claims about a threat to democracy and human rights are “dubious claims” and that “all of them can be unequivocally refuted by facts” (p.16).

- b) Questioning (by the authors of the publication) the statement that: “Kyiv bans the Russian language and culture”. (p. 17). If we were writing this before 2019, the authors of the publication would be more or less right. In 2023, it was only partially true. The Russian language was not banned in common usage but Russian culture certainly was.
- c) Questioning (by the authors of the publication) the statement that: “The Russian army only attacks military targets, never civilians” (p. 17). This statement is, of course, true as it is formulated. But no army can absolutely rule what is called collateral damage during the war. Moreover, some objects have, or can have, both a civilian and a military character (e.g. energy infrastructure or airports).

4.4. State authorities and dissemination of fake news online

The Slovak Embassy in France claimed on Facebook in August 2022 that it was Russia that had attacked Georgia in 2008. Following the same reasoning, the Slovak Ambassador in the UK claimed on Facebook that it was Russia that attacked Georgia in 2008.⁹⁸ In fact, although the official position of the Ministry of Foreign Affairs and European Integration of Slovakia was identical,⁹⁹ an Independent International Fact-Finding Mission on the Conflict in Georgia found out that it was Georgia that initiated that conflict.¹⁰⁰ There are many cited studies that – sometimes reluctantly – accept that this time Russia was not an aggressor, or at least not the side which fired first.

4.5. NGO Konšpirátori.sk and controversial assessment of online news webportal *Standard.sk*

As it was discussed in section 2.3.1, self-regulatory activities in the civil society has raised some interesting legal case studies. However, there is a particularly controversial case study within this civil society activism that may or may not end in court, but nonetheless deserves a mention here. In summary, one (*Standard.sk*)¹⁰¹ of two relatively recently (2020, 2015) es-

⁹⁸ Exact wording of public statement by Igor Slobodník, ambassador of Slovakia to France, was: “That one anniversary is 7 August 2008, when Georgia was attacked by neighbouring Russia” (8 August 2022). Robert Ondrejcsák, ambassador to the UK, wrote similar statement the same day on Facebook: “Open aggression of Russia against its neighbours started already in 2008, by attack on Georgia.”

⁹⁹ E-mail from Michal Slivovic, Director of Department of States of Eastern Europe, Southern Caucasus and Central Asia (8 November 2022) (on file with author).

¹⁰⁰ Eur. Ct. H.R., O.J. (L 323/66), https://www.echr.coe.int/Documents/HUDOC_38263_08_Annexes_ENG.pdf

¹⁰¹ It was on 20th–30th place in readership of online media in 2023. Jaroslav DANIŠKA: Dobrá správa: čítanosť a vplyv Štandardu rastú. *Standard.sk*, 19 June 2023. <https://standard.sk/373242/dobra-sprava-citanost-a-vplyv-standardu-rastie>

established moderate conservative-national online newsportals (the second one is *Postoj.sk* which means both ‘Attitude’ and ‘Wait a moment’) was rated at *Konspiratori.sk* portal among highly problematic outlets, with a negative rating 6.9 out of 10 points. This was followed by an extensive (detailed) public exchange of opinions among editors at *Standard.sk*,¹⁰² supported by editors of *Postoj.sk*,¹⁰³ with the author of this negative assessment.¹⁰⁴ Based on this exchange of opinions and arguments, it is highly questionable whether NGO Konšpiratori.sk had developed and applied really fair and persuasive enough methodology for classifying online outlets as ‘controversial’.

5. Conclusions

As can be seen, the fight against disinformation is rather complex and controversial from legal-constitutional or other perspectives. Enormous human and financial resources have been invested in this field. In three years, at least three different draft laws which were intended to tackle disinformation with a more multi-faceted approach were produced by two ministries (not to mention individual or group initiatives by MPs). Yet, none of draft laws was passed in the end.

The new government has cancelled the Action Plan, and approved a new Concept of Strategic Communication in March 2024. In effect, there is shift from a very loose approach, represented by the important role of the NGOs, to a highly centralised approach, represented by the key role played by the Office of the Government. Moreover, an update to Act on Media Services, following the Digital Services Act, has granted new and important rights to the MSB with respect to content sharing platforms. However, some of these rights have only be in effect since June 2025.

The first real emergency intervention (justified by alleged national security issues) was to regulate controversial webportals which had allegedly disseminated online disinformation posing a national security risk. This was immediately heavily criticized and ended in fierce public legal debates about the constitutionality of such measures. As a result, the extension of legal option for the NSA to ban some controversial NSA ex officio was not approved beyond September 2022 by the Parliament in the summer of 2022. Therefore, the first ban of some controversial websites lasted only a few months during the spring of 2022. As of 2024, or since June 2022, no websites have been banned (except those regulated by EU-wide *ad hoc* emergency regulation of selected Russian websites). However, whether the NSA may actually still block some websites with controversial content on request by third parties remains untested. Some lawyers argue that such legal option does exist., but the NSA has not produced an unambiguous statement in that regard: it merely copied the text of the law, law which remains blurred on this point.

¹⁰² Peter ŠTEVKOV: Analýza Štandardu. Prečo Konšpiratori.sk nehovoria pravdu. *Standard*, 6 December 2023. <https://standard.sk/511125/analiza-standardu-preco-konspiratori-sk-nehovoria-pravdu>

¹⁰³ Martin HANUS: Týždenný výber šéfredaktora. Ako prvé ruší Fico IV. najkonzervatívnejšiu reformu bývalej vlády. *POSTOJ*, 13 November 2023. <https://www.postoj.sk/141573/ako-prve-rusi-fico-iv-najkonzervativnejšiu-reformu-byvalej-vlady>

¹⁰⁴ Pavol HARDOŠ: Posudok – standard.sk. 2023. <https://konspiratori.sk/stranka/1250>

For the NSA to act, it would need to follow all administrative-legal steps in the procedures, which seems to be impossible at the moment due to confusing and contradictory legal regulations. Even the public media discourse seems to suggest that the NSA is not allowed to block any controversial content at all.¹⁰⁵ The new government appears to be unlikely to welcome or to support such ban being issued or tested by the NSA on request by a third party. For example, Matúš Šutaj Eštok, the minister of interior, participated in an online discussion lead by Daniel Bombic *aka* Danny Kollar in March 2024. At that time, Bombic was subject to three international warrants due to dissemination of extremist materials, cyberbullying and making public private data. He was living in the UK and faced extradition requests the same month. Moreover, this debate was co-moderated by Judita Laššáková, then candidate for Smer-SSD for elections to the European Parliament (elected). The British lower court approved extradition of Bombic in April 2024, subject to appeal.¹⁰⁶

Some educational and analytical outputs produced by analysts and activists contribute to spreading disinformation rather than to rational debate or to fighting disinformation. Some state authorities, and some diplomats also disseminate or support public dissemination of misinformation. The Police have also become involved in the controversial fight against misinformation and disinformation. Ironically, some new MPs, including new minister of culture, Martina Šimkovičová, gained their popularity (which translated into success in general elections) from being active on or appearing on some webportals listed as controversial and on alternative media such as online TV Slovan.

On the other hand, 'alternative' webportals (or social media for that matter) do not represent quality journalism or an ideal space for rational debate about public affairs. As mentioned, the efforts of some controversial webportals to fight back legally against being publicly listed as 'conspiratory-like' sources has not succeeded. Although, it should be mentioned that they have attempted to use rather innovative argumentation legally speaking. Their argumentation was based – in addition to standard legal arguments such as being under some kind of libel attack – on business law arguments (i.e. that the media-business competition was trying indirectly to destroy/damage their reputation, using unfair methods). Some of these cases are still awaiting a final judicial verdicts

In summary, although misinformation and disinformation is indeed a problem in Slovakia, no solution has been found that would tackle this issue effectively and efficiently, either from a legal perspective or from another (e.g. educational, analytical, monitoring) perspectives. Perhaps the only partial exception was the relative success in stopping deepfake video disseminated shortly before the 2023 general elections. It appears that approaches to fighting disinformation are based on dubious assumptions – that propaganda can succeed on its own under normal conditions within the population at large, without having met any necessary pre-conditions (such as

¹⁰⁵ See TASR: NBÚ od októbra nemôže blokovať škodlivý obsah, zákon nestihli prijať. *Teraz.sk*, 30 September 2022. <https://www.teraz.sk/slovensko/nbu-od-oktobra-nemoze-blokovat-sko/663946-clanok.html>; TASR: NBÚ od októbra nemôže blokovať škodlivý obsah, doteraz blokoval štyri weby. *Postoj*, 30 September 2022. <https://www.postoj.sk/115351/nbu-od-oktobra-nemoze-blokovat-skodlivy-obsah-doteraz-blokoval-styri-weby>

¹⁰⁶ Šimon KÉRY: Londýnsky súd rozhodol o osude extrémistu Bombica. „Som vydaný, prehral som,” oznámil. *StartitUp*, 24 April 2024. <https://www.startitup.sk/londynsky-sud-rozhodol-o-osude-extremistu-bombica-som-vydany-prehral-som-oznamil/>

a very low mutual trust in society in general and in particular, towards institutions,¹⁰⁷ poor media literacy and general low knowledge levels among segments of the population, combined with real or imagined threats, with all this facilitated by occasional biased reporting from major media, occasional biased communication by ‘standard’ politicians and diplomats, and social media algorithms). Mesarčík identified sociological factors that either cause or facilitate disinformation (in addition to technical and economic factors) including: decrease of social capital (i.e. trust), increase in social inequality, increase in social polarisation, distrust in science, and evolution of media environment. Indeed, ‘alternative’ media and social media are part of the problem rather than a solution to it.¹⁰⁸

References

- 2MUSE (2021) = 2MUSE: *Dezinformácie a regulácia sociálnych sietí. Prieskum realizovaný pre europoslancu SPOLU / EPP Vladimíra Bilčíka*. https://europske.noviny.sk/wp-content/uploads/2021/12/Prieskum_2muse_Bilcik.pdf
- ANDERSON (2021) = Christopher William ANDERSON: Propaganda, misinformation, and histories of media techniques. *Harvard Kennedy School (HKS) Misinformation Review*, Vol. 2., No. 2. (2021) <https://doi.org/10.37016/mr-2020-64>
- ANSTEAD (2021) = Nick ANSTEAD: *What do we know and what should we do about...? Fake news*. Thousand Oaks, Sage, 2021. <https://doi.org/10.4135/9781529738896>
- CHLEBCOVÁ HEČKOVÁ – SMITH (2023) = Andrea CHLEBCOVÁ HEČKOVÁ – Simon SMITH: Slovakia. In: Nic NEWMAN – Richard FLETCHER – Kirsten EDDY – Craig T. ROBERTSON – Rasmus Kleis NIELSEN: *Digital News Report 2023*. Oxford, Reuters Institute, 2023. 96–97. <https://doi.org/10.60625/risj-p6es-hb13>
- HAJDU et al. (2023) = Dominika HAJDU – Jana KAZAZ – Katarína KLINGOVÁ – Patrik SZICHERLE – Viktória MUSILOVÁ: *GLOBSEC Trends 2023. United we (still) stand*. Bratislava, Globsec, 2023. <https://www.globsec.org/what-we-do/publications/globsec-trends-2023-united-we-still-stand>
- HAJDU–KLINGOVÁ (2020) = Dominika HAJDU – Veronika KLINGOVÁ: *Voices of Central and Eastern Europe. Perceptions of democracy & governance in 10 EU countries*. Bratislava, Globsec, 2020. <https://www.globsec.org/sites/default/files/2020-06/Voices-of-Central-and-Eastern-Europe-read-version.pdf>
- HAMID (2022) = Nafees HAMID (2022): *Mass media and persuasion. Evidence-based lessons for strategic communications in CVE. Briefing Note*. London, King’s College, 2022. https://www.xcept-research.org/wp-content/uploads/2022/07/2022-07-19_FINAL_Briefing-note-Mass-media-Jul19.pdf

¹⁰⁷ A Czech study has found that only low trust in institutions correlates with conspiratory vision of the world. See <https://www.irozhlas.cz/neduvera> (the full report is not officially available).

¹⁰⁸ MESARČÍK (2023), 38.

- HRCKOVA et al. (2022) = Andrea HRCKOVA – Robert MORO – Ivan SRBA – Jakub SIMKO – Maria BIELIKOVA: Autonomation, not automation: Activities and needs in European fact-checking organizations as a basis for designing human-centered AI systems. *ACM Journal on Responsible Computing*, Vol. 2., No. 4. (2022) Art. no. 17. 1–42.
<https://doi.org/10.1145/3764592>
- HUSOVEC (2022) = Martin HUSOVEC: Súčasné blokovanie dezinformačných stránok je ústavne problematické. Čo s tým? [The Current Blocking of Disinformation Sites is Constitutionally Problematic. What About That?] *DenníkN*, 222 April 2022.
<https://dennikn.sk/2818631/sucasne-blokovanie-dezinformacnych-stranok-je-ustavne-problematicke-co-s-tym/>
- KOLTAY (2023) = András KOLTAY: Censorship as a tool against state disinformation. The freedom of expression implications of the Russian–Ukrainian war. *Journal of International Media and Entertainment Law*, Vol. 10., No. 1. (2023) 1–53.
<https://www.swlaw.edu/sites/default/files/2024-01/JIMEL%2010.1%20Koltay231215.pdf>
- LANGER (2023) = Tomáš LANGER: *Case study slapp in Slovakia*. Bratislava, MEMO 98, 2023.
https://memo98.sk/uploads/content_galleries/source/memo/patfox_2022/slovakia-case-study-en.pdf
- MESARČÍK (2023) = Matúš MESARČÍK: *Nástroje verejného práva pre boj s dezinformáciami v online prostredí*. Bratislava, Právnická fakulta Univerzity Komenského v Bratislave [Comenius University Faculty of Law], 2023. https://www.flaw.uniba.sk/fileadmin/praf/Pracoviska/Ustavy/UPITPDV/E-KNIHY/Monografia_mesarcik_01.pdf
- OSMUNDSEN – BANG PETERSEN – BOR (2021) = Mathias OSMUNDSEN – Michael BANG PETERSEN – Alexander BOR: How partisan polarization drives the spread of fake news. *Brookings*, 13 May 2021. <https://www.brookings.edu/articles/how-partisan-polarization-drives-the-spread-of-fake-news/>
- PISKLOVÁ–SÝKORA (2021) = Miroslava PISKLOVÁ – Juraj SÝKORA: Disinformation and the Slovak Republic. In: Miroslava PISKLOVÁ – Juraj SÝKORA (Eds.): *Disinformation and the V4*. Bratislava, Slovak Foreign Policy Association, 2021. 3–10.
<https://www.sfpa.sk/wp-content/uploads/2021/08/Disinformation-and-the-V4.pdf>
- PIŠKO (2020) = Michal PIŠKO: Dezinformačné weby pravidelne číta každý siedmy. *Transparency International Slovensko*, 20 February 2020.
<https://transparency.sk/sk/dezinformacne-weby-pravidelne-cita-kazdy-siedmy/>
- SIMKO et al. (2022) = Jakub SIMKO – Patrik RACSKO – Matus TOMLEIN – Martin HANAKOVA – Robert MORO – Maria BIELIKOVA: A study of fake news reading and annotating in social media context. *arXiv:2109.12523v2* [cs.HC].
<https://doi.org/10.48550/arXiv.2109.12523>
- ŠKOLKAY (2018) = Andrej ŠKOLKAY: A case-specific instance of media capture. The Gorilla Case of Slovakia. *Środkowoeuropejskie Studia Polityczne*, No. 4. (2018) 65–94.
<https://doi.org/10.14746/ssp.2018.4.4>

- ŠKOLKAY (2022) = Andrej ŠKOLKAY: *Journalism and Populism in Slovakia, National Riport. DEMOS Project. Work Package 7, Task 7.3, Version 2*. Bratislava, School of Communication and Media, 2022. http://www.skamba.info/wp-content/uploads/2022/03/Task_7_3_National_reportsLOVAKIAmediaPopulism-10.11-4.pdf
- SPOHR (2017) = Dominic SPOHR: Fake news and ideological polarization. Filter bubbles and selective exposure on social media. *Business Information Review*, Vol. 34., No. 3. (2017) 150–160. <https://doi.org/10.1177/0266382117722446>
- ŠTEPANOVIČ (2021) = Dominik ŠTEPANOVIČ: Cena za (dez)informácie. *Výnosy z reklám webov s problematickým obsahom. Komentár č. 1-2021*. Analytical Department of Ministry of Defence, April 2021. https://www.mosr.sk/data/files/4364_2021-k-01-cena-za-dezinformacie-recenzovane.pdf
- URSPRUNG (1994) = Tobias URSPRUNG: The use and effect of political propaganda in democracies. *Public Choice*, Vol. 78. (1994) 259–282. <https://doi.org/10.1007/bf01047758>

French online disinformation regulation in practice

PÉTER NEMERE

1. Introduction

Two major current challenges facing French society and France's economy are digitalization and Internet regulation. France endeavors to be a leading example for Europe in that field and in accordance with these efforts, digital regulation and tech sovereignty were the priorities of the French EU Council presidency during the first half of 2022.¹ Thus, the French presidency of the European Council made further steps toward implementing the Digital Markets Act² and the Digital Services Act (DSA)³ at an EU level.

France is not only calling for EU action but has also attempted to regulate the “secteur du numérique” (digital sector) within its own borders. Therefore, it is essential to study the French example when discussing online disinformation and potential regulation in Europe. The main objective of this paper is to present the French regulation on online disinformation, which entered into force in 2019, and to summarize the experience of its application in practice by analyzing the transparency reports of online platforms and Arcom's (*Autorité de régulation de la communication audiovisuelle et numérique*, Regulatory Authority for Audiovisual and Digital Communication) recommendations and other procedures and initiatives prescribed by law.

This paper consists of three main parts. In the first section, French general disinformation regulations and provisions of the Law on the “fight against the manipulation of information”⁴ (*Loi relative à la lutte contre la manipulation de l'information*, hereinafter: the Infox Law) are described.

In the second part, Arcom's monitoring procedure, the declarations of online platforms and the Authority's recommendations are presented.

The third part shows the other measures prescribed by the Infox Law, such as the judicial procedures and the efforts taken in media literacy.

¹ <https://www.euractiv.com/section/digital/news/france-to-prioritise-digital-regulation-tech-sovereignty-during-eu-council-presidency/>

² Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828.

³ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC.

⁴ Loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information.

1.1. How do the French become informed?

Some 94% of French people say they are interested in being informed – as determined from a representative survey led by Arcom in 2024.⁵ Their main fields of interest were social, cultural and sporting issues, the environment and politics. Participants were asked to list their motivations in priority order, from which it emerges that they primarily gather information to understand the world around themselves (44%), educate themselves (42%), to create their own opinion (38%) or stay informed about major societal events (37%).⁶ However, the desire to understand other opinions (11%) or discussing or debating with others (17%) were less common motivations in French society.

The survey also tried to measure how well informed participants were by way of four multiple-choice questions on politics.⁷ It may be surprising that after the high percentage who declared themselves interested in being informed, only one person out of ten could answer all the questions correctly.

Main information sources of the French are still the television (80%) and the radio or podcasts (67%), but search engines (65%) and social media (56%) are also among the most popular choices. It is rare that they acquire information exclusively from one type of media. However, 11% of the population only consults online platforms. Even if the French prefer to consult sources that they already know when gathering information (84%), and 32% of the population finds anonymous sources as credible as journalists.

According to the survey, the French assess information critically, and most of them are aware that algorithms disseminate true and false information online. However, users' fears relate primarily to the reliability and tone of the information. Fear of being isolated in 'filter bubbles' is secondary.

The survey results suggest that while the French do strive to make conscious choices when consuming media, the dominant role of online platforms and professional content in information still means that it is appropriate to regulate online disinformation and promote solutions to support media literacy.

1.2. The regulation of fake news before the Infox Law

French law has regulations to address the dissemination of disinformation in the offline context. The Law on the Liberty of the Press of July 29, 1881 (Law on the Liberty of the Press)⁸ under Chapter IV punishes, among others acts, crimes and offences committed in the press or by any other means of publication. It lists infringements such as incitement to

⁵ A survey made of a representative sample of 3,356 French people aged 15 and over. The participants were interviewed online and by telephone between November 22 to December 20, 2023. <https://www.arcom.fr/nos-ressources/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/les-francais-et-linformation>

⁶ The survey listed their first, and their first three motivations.

⁷ The four questions were related to the following topics: 1) Who attacked Israel recently? 2) Who is the prime minister of the UK? 3) When does the mandate of the current French President end? 4) In 2023, what was the average inflation rate?

⁸ Loi du 29 juillet 1881 sur la liberté de la presse.

commit crimes or offences, the denial of acts of genocide, crimes against humanity, defamation, or forbidden publications.

Article 27 of the law prescribes that

“The publication, dissemination or reproduction, by any means whatsoever, of false news, fabricated or falsified documents or documents falsely attributed to third parties when, done in bad faith, they disturb the public peace or are likely to disturb it, shall be punishable by a fine.”⁹

The fine for such infringements is up to 45,000 euros. In more severe cases

“The same facts shall be punishable by a fine of 135,000 euros when the publication, dissemination or reproduction made in bad faith is likely to undermine the discipline or morale of the armed forces or to hinder the war effort of the Nation.”

Article 27 of the Law on Liberty of the Press penalizes the public dissemination of false news which may disturb the peace. The offense applies to any “publication, dissemination, or reproduction” of false information, emphasizing the need for publicity as a prerequisite for legal action. This means that private messages or unpublished thoughts are not punishable. The falsehood must also be objective and verifiable, pertaining to current or recent facts. Statements about past events or speculative predictions do not fall under the scope of this law. Additionally, for prosecution, bad faith must be proven – negligence or failure to verify sources is insufficient. Courts evaluate intent through the circumstances of the case, often holding professionals such as journalists to higher standards due to their responsibility to verify information. Furthermore, false news must have the potential to disturb the peace, such as by causing panic, economic disruption, or diplomatic tension.¹⁰

1.3. Disclosing false information in France

In addition to false news, the law addresses among other infringements, defamation (*diffamation*),¹¹ insult (*injure*),¹² outrage against state representatives,¹³ and offences against public officials (*délit offense*).¹⁴ Defamation is defined as the public attribution of a fact that damages an individual’s honor or reputation. The accused may avoid liability by proving the truth of the statement, provided it was in the public interest and made in good faith. Unlike defamation, insults focus on statements or actions intended to demean or offend without the need to prove that the statement’s factual accuracy. Public insults based on race, religion, or gender are particularly severely dealt with and subject to increased penalties according

⁹ Author’s translation.

¹⁰ STELLA (2019): 80–85.

¹¹ Article 29–35 quarter Loi du 29 juillet 1881 sur la liberté de la presse.

¹² Ibid.

¹³ Article 37. Loi du 29 juillet 1881 sur la liberté de la presse.

¹⁴ Articles 38–41-1 Loi du 29 juillet 1881 sur la liberté de la presse.

to Article 32 of the Law on Liberty of the Press. Offences against public officials or institutions aim to protect the dignity of figures like the President or judiciary, so ensuring trust in these entities.

New challenges have arisen with the proliferation of social media and digital platforms, given that these enable rapid information dissemination and amplify the potential for harm. Courts have adapted these traditional legal frameworks to address the complexities of digital communication, holding both individuals and platforms accountable for the spread of misinformation and offensive content.¹⁵ However, ongoing debates persist regarding the definitions of ‘false news’ and ‘public disturbance’, highlighting the need for careful application of these laws.

In conclusion, the Law on Liberty of Press represents a nuanced approach to regulating disinformation, defamation, and offensive content. It underscores the importance of balancing the rights to freedom of expression with the need to maintain public order and protect individual dignity, which remains essential even in online communication.

2. The changes introduced by the Infox Law

As the Infox Law applies the same terms for defining ‘false news’, and given that the Law on Liberty of the Press’s scope applies to public online communication services¹⁶ as well, the proposal for a new ‘fake news’ regulation provoked a heated debate. Some considered it pointless to introduce new legislation. Even the Council of the State (*Conseil d’État*) found the terminology imprecise (for example, in the case of false information and false news) when the proposal was debated in parliament.¹⁷

2.1. The Infox Law

The Infox Law focuses on five main fields: reinforced transparency obligations for online platforms during electoral periods, cooperation requirements for online platforms, new competencies of the authority in terms of the broadcasting conventions and their conclusion, the implementation of new judicial proceedings and efforts on media literacy.

The text of the Infox Law gives specific powers to the Higher Council of Broadcasting (*Conseil supérieur de l’audiovisuel*, CSA). Even if the text has not changed in its wording, it is essential to note with regard to this part of the text that these competencies had been transferred to Arcom, when the Council merged into this authority. Thus, while describing the provisions of the law, I will refer to these bodies as ‘Authority’ or ‘Arcom’ for a better understanding.

¹⁵ STELLA (2019)

¹⁶ Loi n° 2004–575 du 21 juin 2004 pour la confiance dans l’économie numérique, Chapitre V.

¹⁷ <https://www.conseil-etat.fr/avis-consultatifs/derniers-avis-rendus/a-l-assemblee-nationale-et-au-senat/lutte-contre-les-fausses-informations>

2.1.1. Obligations for online platforms during electoral periods

The Infox Law prescribes additional duties during electoral periods for online platforms whose activity exceeds a given threshold (five million individual users per month) of a number of connections on the French territory.¹⁸ The scope of this provision was modified by the SREN Law (Sécurité et Régulation de l'Espace Numérique – Security and Regulation of the Digital Space).¹⁹ Since the adoption of the SREN Law in May 21, 2024 (entered into force on February 17, 2024), these obligations are prescribed for the very large online platforms and very large search engines defined in the DSA. During the three months preceding the first day of the month of a general election and up to the date of the ballot in which the election is held, the online platform operators are required to comply with transparency provisions. General elections include nationwide elections. Therefore, these rules shall apply during the electoral period of the elections of the President of the Republic, the deputy to the Assemblée Nationale, the senators of the Sénat, the members of the European Parliament and national referendums.

The online platforms shall provide users with fair, clear and transparent information on the identity of the entity which pays the platform remuneration in return for the promotion of information content relating to a debate of general interest; they shall inform users with fair, clear and transparent information on the use of their personal data in the context of the promotion of information content relating to a debate of general interest; and make public the amount of remuneration received in return for the promotion of such information content when the amount exceeds a certain threshold.²⁰

The Infox Law also prescribes aggregating this information and making it available in an open format, and updating it regularly during the above-mentioned period.

2.1.2. New competencies of the Authority in terms of the broadcasting conventions and their conclusion

The Infox Law also modifies the Law No. 86–1067 of September 30, 1986 on the freedom of communication (Law Léotard).²¹ It gives increased control over the broadcasting of radio and television services to the Authority.

According to the prescribed conditions, the Authority can reject the request for a broadcasting agreement or even stop or suspend broadcasting. The Authority may reject the request for the conclusion of an agreement if the broadcasting of the radio or television service entails a serious risk of infringing human dignity, the freedom and property of others, the pluralistic nature of the expression of currents of thought and opinion, the protection of children and adolescents, the safeguarding of public order, the needs of national defense or

¹⁸ Décret n° 2019-297 du 10 avril 2019 relatif aux obligations d'information des opérateurs de plateforme en ligne assurant la promotion de contenus d'information se rattachant à un débat d'intérêt général.

¹⁹ Loi n° 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique.

²⁰ Article 1(2), Infox Law.

²¹ Loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication.

the fundamental interests of the Nation, including the regular functioning of its institutions. The same applies when broadcasting the said service, in view of its very nature, would constitute a violation of the laws in force.²²

When the conclusion of the agreement is requested by a legal entity controlled by a foreign state or placed under the influence of that state, the authority may, in order to assess the request, take into account the content that the applicant, its subsidiaries, the legal entity controlling it or its subsidiaries publish on other electronic public communication services.²³

During the electoral period, the authority may, if it finds that the service which has been the subject of an agreement with a legal entity controlled by a foreign state or placed under the influence of that state, deliberately broadcast false information likely to alter the fairness of the ballot, in order to prevent or stop the broadcasting of false information may, to prevent or put an end to this disturbance, order the suspension of the broadcasting of this service by any electronic communication process until the end of the voting operations.²⁴ The authority may, after formal notice, impose the sanction of unilateral termination of the concluded agreements as well.²⁵

If the Authority considers that the facts justify the initiation of the above-mentioned procedure, the Authority shall notify the complaints of the persons concerned. The latter may submit their observations within forty-eight hours of the notification.

The Authority's decision at the end of this procedure shall be substantiated and notified to the persons concerned and to the distributors or satellite operators who broadcast the service in France and must ensure that the suspension measure is carried out.

2.1.3. New judicial proceedings

The Infox Law creates new judicial proceedings. During the election period, an interim relief judge (*juge dés référes*) rules on cases in which inaccurate or misleading allegations or imputations of a fact are likely to affect the fairness of the forthcoming ballot if they are deliberately, artificially or automatically disseminated on a massive scale via an online public communication service. The procedure can be stated at the request of the Public Prosecutor (*ministère public*), any candidate, political party or grouping or any person with an interest in taking action, and without prejudice to compensation for the damage suffered, and the interim relief judge may prescribe proportionate and necessary measures to stop the dissemination. The judge shall give a ruling within forty-eight hours of the referral. The same time limit applies in the event of an appeal.

²² Art. 5., Infox Law.

²³ Ibid.

²⁴ Ibid., Art. 6.

²⁵ Ibid., Art. 8.

2.1.4. *Media literacy*

The Infox Law contains provisions relating to media and information education.²⁶ It completes existing legislation, and modifies the Education Code (*Code de l'éducation*). These modifications aim to strengthen media and information literacy education in schools and develop students' critical thinking skills related to fake news and misinformation.

The Infox Law has integrated media and information literacy into school curricula, enabling students to exercise their citizenship in the context of the information and communication society.²⁷ It supplemented the curricula to include training on the critical analysis of available information, thereby strengthening students' ability to assess the reliability of sources.²⁸ The Infox Law adjusted teacher training, and included specific modules dedicated to media and information literacy, so preparing educators to pass on these skills to students.²⁹

However, these measures were only supplementary elements to existing key provisions, and did not create completely new obligations for education.

2.1.5. *Subsequent changes in the legislation*

Since its adoption on October 22, 2018, the Infox Law has been amended by legislators several times. The two most recent amendments, on May 21 and November 13, 2024,³⁰ were primarily necessary for harmonization with EU legislation, the Digital Services Act (DSA).

These amendments modified and abrogated several provisions of the Infox Law; therefore, to facilitate understanding, the abrogated and modified provisions are presented in the following section based on a comparison of the original and the current versions.

Harmonization with EU law made it necessary to amend the articles relating to cooperation with platform operators to combat dissemination of false information. It abrogated several obligations of the platform operators prescribed by the Infox Law such as the transparency measures for content promoted on behalf of a third party and other transparency measures regarding online platforms algorithms; promoted content from press companies and agencies and audiovisual communication services; accounts that spread false information on a massive scale; the identity of the persons paying them remuneration in return for the promotion of information content relating to a debate of general interest; the nature, origin and distribution methods of content, and media and information education.

²⁶ Ibid. Art. 16–19.

²⁷ Article L312–15 Code de l'éducation.

²⁸ Article L332–5 Code de l'éducation.

²⁹ Article L721–2 Code de l'éducation.

³⁰ Loi n° 2024–449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique; Ordonnance n° 2024–1019 du 13 novembre 2024 portant extension et adaptation en outre-mer des dispositions de la loi n° 2024–449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique et du règlement européen 2022/2065.

Seemingly a step forward from the previous national legislation, the EU legislation established a detailed regulatory framework. Since the entry into force of the Infox Law, the EU has introduced legislation not only in the field of online platform regulation but on political advertising³¹ as well.

2.2. The intersections of EU legislation and the Infox Law

To ensure that the DSA is implemented consistently at national level, each EU country designates one or more competent authorities to enforce the regulation, including an independent authority to act as the digital services coordinator. In France, Arcom was designated to fulfil this duty. In order to ensure full implementation of the DSA, Arcom also cooperates with the French Data Protection Authority (*Comission nationale de l'information*, CNIL) with regard to privacy protection provisions, and the Directorate-General for Competition, Consumer Affairs and Fraud Control (*Direction générale de la concurrence, de la consommation et de la répression des frauds*, DGCCRF), with regard to provisions specific to marketplaces.³²

2.2.1. Arcom's activities under the DSA and the Infox Law

The DSA and the DMA aim to set up a common regulatory framework within the European Union for the digital space to protect fundamental rights and foster innovation, growth and competitiveness.³³ The EU realized that Member States were introducing, or considering introducing national laws on that field, and wished to impose diligence requirements for providers of intermediary services with respect to how they should tackle illegal content, online disinformation or other societal risks.³⁴ As these diverging national laws negatively affect the internal market, the EU Commission launched a regulatory procedure to harmonize the rules for the common market on December 15, 2020.³⁵ The DSA was finally adopted after a two-year-long debate on October 19, 2022, and its first provisions entered into force on August 25, 2023. Therefore, the French legislator introduced the Law to Secure and Regulate the Digital Space (*Loi n° 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique*, the SREN Law).

³¹ Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising.

³² L'Arcom se réjouit de la promulgation de la loi visant à sécuriser et à réguler l'espace numérique, qui la désigne coordinateur pour les services numériques en France [Press release]. *Arcom*, 22 May 2024. <https://www.arcom.fr/presse/larcom-se-rejouit-de-la-promulgation-de-la-loi-visant-securiser-et-reguler-l'espace-numerique-qui-la-designe-coordonateur-pour-les-services-numeriques-en-france>

³³ The Digital Services Act. <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>

³⁴ Preamble (2) DSA.

³⁵ Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final.

2.2.2. The SREN Law

The SREN Law aims to secure and regulate the digital space to protect Internet users, particularly children, and to strengthen digital sovereignty. It introduces several key measures beyond the harmonization of EU law in the field of protecting minors, combating illegal content online, and fighting scams and online hate.

The Arcom is tasked with establishing a framework requiring pornographic websites to verify users' age under penalty of sanctions.³⁶ Non-compliant websites may be blocked or removed from search engines following the procedure of an administrative judge (*juge administrative*). Hosting providers must remove child pornography content reported by authorities within 24 hours. Warnings will be mandatory for videos simulating sexual violence, and the persons shown in such content can also request the removal of pornographic content made public without their consent. The Pharos platform,³⁷ which was launched at the request of the Ministry of the Interior (*Ministre de l'Intérieur et des Outre-Mer*), and allows members of the public, Internet service providers, government watchdog and other stakeholders to report illegal content,³⁸ will gain new powers to act swiftly against content involving torture or barbarity.

The SREN Law prescribes introducing an anti-scam filter that will alert users when attempting to access malicious websites following fraudulent messages. The law toughens penalties for cyberbullying and serious offences, introducing temporary bans or suspensions from social networks. Platforms failing to enforce these bans risk fines. Harmful deep-fakes will also face stricter penalties according to the legislation.

The law regulates cloud computing services (*services d'informatique en nuage*), mandating interoperability and imposing limits to reduce dependence on American tech giants.³⁹ The SREN Law handles sensitive administrative data issues as well. It centralizes the information on tourism rentals and aims to improve oversight of this market by a platform called API meublées (*application programming interface*).⁴⁰ Emerging digital games which allow players to win redeemable digital asset issued by the game operator, (*Jonum, Jeux à Objet Numérique Monétisable*) will be subject to similar obligations as online gambling game operators.⁴¹ It also deals with money laundering and terrorism prevention measures.

Above all of these measures, the online platform-related measures that are linked to DSA prescription are the most relevant in the present case. The SREN Law harmonizes the definitions of French legislation with EU legislation. Therefore, it replaces the online plat-

³⁶ Articles 1–6 SREN.

³⁷ <https://www.internet-signalement.gouv.fr/PharosS1/>

³⁸ Specially violence, endangering people, threatening or advocating terrorism, insult or defamation, incitement to racial hatred or discrimination, offences against minors.

³⁹ Loi du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique. *Vie-publique.fr*, 17 April 2025. <https://www.vie-publique.fr/loi/289345-loi-du-21-mai-2024-securer-et-reguler-lespace-numerique-sren>

⁴⁰ <https://meubles.peren.fr/>

⁴¹ The French law on the regulation of games including NFT is passed: Place your bets. *BCLP*, 22 April 2024. <https://www.bclplaw.com/en-US/events-insights-news/the-french-law-on-the-regulation-of-games-including-nft-is-promulgated-place-your-bets.html>

form operator (*opérateurs de plateformes en ligne*) definition⁴² by EU conform terminology.⁴³ This includes both the online platform service providers, the search engines (*moteurs de recherche en ligne*) and video-sharing platforms (*plateformes de partage de vidéos*). Consequently, the scope of the provisions on online platforms such as obligations during elections periods described in point 2.1.1, are now applicable to the same services as in the DSA. The subjects of reporting activities are also changed, as discussed in the following chapter.

2.3. Platform annual reports

As discussed in point 2.1.1, the Infox Law prescribed additional reporting obligations for electoral periods for those online platforms whose activity exceeds a given threshold (five million individual users per month) for the number of connections on French territory.⁴⁴ The scope of this provision was modified by the SREN Law; therefore, since February 17, 2024, these obligations are prescribed for the very large online platforms and very large search engines defined in the DSA.

Therefore, the CSA, and later Arcom has monitored activities related to the manipulation of information during election periods since 2019. The Authority made public reports for four consecutive years related to the years from 2018 to 2022. In the first reporting period, sixteen services provided by eleven operators were included. In 2023, twelve operators⁴⁵ published their reports, which are publicly available. According to the law, these transparency reports are related to legislative (general elections of deputies and senatorial elections), presidential and European elections. National referendums are also subject to this legislation.⁴⁶ As there were no national elections in 2023, this year is not relevant from the perspective of the provisions discussed in this chapter.⁴⁷

2.3.1. Implementation and Platform Efforts

Between 2019 and 2022, platforms such as Facebook, Google, YouTube, Snapchat, and Twitter demonstrated varying levels of commitment to implementing anti-disinformation measures. Arcom's evaluations revealed a mix of progress and challenges. On the posi-

⁴² See Article L. 163–1 du code électoral and Article L. 111–7 du code de la consommation.

⁴³ “A la première phrase du cinquième alinéa de l'article 14 les mots (...) sont remplacés par les mots: fournisseurs de plateformes en ligne, au sens du paragraphe i de l'article 3 du règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/ CE (règlement sur les services numériques), les moteurs de recherche en ligne, au sens du paragraphe j du même article 3, et les plateformes de partage de vidéos, au sens des cinq derniers alinéas de l'article 2 de la présente loi.” SREN, Art. 54.

⁴⁴ Décret n° 2019–297 du 10 avril 2019 relatif aux obligations d'information des opérateurs de plateforme en ligne assurant la promotion de contenus d'information se rattachant à un débat d'intérêt général.

⁴⁵ Dailymotion, Google, LinkedIn, Meta, Microsoft, Pinterest, Snapchat, TikTok, Twitch, the Wikimedia Foundation, X and Yahoo.

⁴⁶ L'élection du Président de la République, les élections générales des députés, l'élection des sénateurs, l'élection des représentants au Parlement européen et les opérations référendaires.

⁴⁷ There were senatorial elections in 2023, but as the Senators are elected indirectly by an electoral college, the above-mentioned provisions are not applicable.

tive side, many platforms established direct communication channels with regulators and implemented tools to combat false information.⁴⁸

The reports made by the platforms are then reviewed by Arcom to assess the effectiveness of the measures taken. The reports draw from a combination of data sources including operator submissions, Arcom's own questionnaires,⁴⁹ expert consultations, and direct communication with platform representatives. This comprehensive approach allows Arcom to evaluate the performance of the platforms in complying with legal obligations.

The methodology employed by Arcom includes both qualitative and quantitative assessments. The primary evaluation criteria encompass the existence and accessibility of reporting mechanisms for false information, the transparency of algorithms used to curate content, the promotion of verified and credible sources, user education initiatives, and partnerships with third-party fact-checkers. However, Arcom notes that the level of detail in the declarations varies significantly between platforms, with many operators providing limited information on the human and financial resources allocated to these efforts.

The 2019 report published in 2020, the first in the series, shows that platforms introduced initial measures to comply with the new legal requirements. The report analyzed mechanisms for reporting false information likely to threaten public order or undermine the fairness of the ballot, the transparency of algorithms, the promotion of content from press companies and news agencies and from audiovisual services, the mechanisms to tackle accounts disseminating false information on a massive scale, users' information,⁵⁰ and media and information literacy. Arcom found that most platforms had implemented reporting tools against disinformation, although the ease of use varied. It was noted that some platforms required users to fill out lengthy forms to report false content, while others offered more user-friendly, accessible options. Arcom emphasized that simplifying these tools would make it easier for users to report misleading information and urged platforms to improve transparency in how their algorithms prioritize content.⁵¹

The 2020 report highlights the significant impact of the COVID-19 pandemic on the volume of disinformation. Platforms faced an unprecedented surge in false content related to health information. For instance, Meta reported the removal of 12 million posts on Facebook and Instagram containing false claims about the pandemic.⁵² Arcom acknowledged the efforts made by platforms to address this issue but reiterated the need for increased transparency, particularly regarding how algorithms identify and manage disinformation and how platforms handle commercial communications.⁵³ However, the report stated that specific initiatives, such as promoting content from health authorities and removing false claims about vaccines, showed promise, even if they were inconsistently applied according to Arcom.⁵⁴

⁴⁸ CSA (2020) 4.

⁴⁹ <https://www.arcom.fr/internet-et-reseaux-sociaux/lutte-contre-la-manipulation-de-linformation-declarations-des-operateurs-de-plateformes-en-ligne-et-questionnaires-de-larcom>

⁵⁰ On the nature, origin and modalities for dissemination of content, and the identity of individuals providing remuneration in return for the promotion of information content.

⁵¹ CSA (2020) 4–8.

⁵² CSA (2021) 15.

⁵³ Ibid. 61–62.

⁵⁴ Ibid. 71–75.

In the report for 2021, Arcom noted further improvements in the measures taken by platforms, but it expressed concerns about the inconsistent quality of the data provided by operators.⁵⁵ The report stressed, similarly to previous reporting periods, the importance of more precise reporting, particularly regarding the role of algorithms in moderating content and combating disinformation. Arcom also highlighted the need for platforms to establish stronger partnerships with independent fact-checking organizations to enhance the credibility and accuracy of the information shared on their services.⁵⁶

Regarding the 2022 elections,⁵⁷ Arcom published a comprehensive report on electoral campaigns, in which it identified online disinformation as a key challenge during the 2022 elections. The authority intensified cooperation with digital platforms in order to reduce the spread of false information. Platforms such as Facebook, Google, and Twitter deployed specialized moderation teams to combat disinformation who focused on flagging and removing harmful content in real time.⁵⁸ Arcom reported that platforms were increasingly transparent, providing regular updates on their content moderation efforts and engaging more proactively with public authorities.⁵⁹

One significant measure was the creation of dedicated points of contact between platforms and regulatory bodies during the campaign period. However, questions remained about the platforms' adherence to French electoral laws, especially concerning foreign influence.⁶⁰ Arcom recommended further public awareness campaigns and educational resources for political actors to mitigate disinformation risks.⁶¹

Arcom noted that while platforms had made progress, disinformation continued to pose a threat to democratic processes. The report emphasized the importance of stronger collaboration between digital platforms, authorities, and media outlets to tackle this issue effectively.⁶²

In 2024, the Authority observed European parliament elections. These occasions marked the first application of the EU's Digital Services Act, which imposed stricter obligations on online platforms to address disinformation risks. Arcom coordinated efforts at the national level to ensure compliance with the new regulations and protect electoral integrity.⁶³ Platforms increased their real-time monitoring of disinformation, implementing advanced moderation measures, sharing best practices with campaign teams, and providing educational kits to political parties to clarify platform policies and help avoid inadvertent rule violations.⁶⁴ Stress tests organized by the European Commission simulating disinformation scenarios were introduced to improve response strategies.⁶⁵ The report stressed the importance of public awareness campaigns to educate users on disinformation risks. Arcom

⁵⁵ Arcom (2022a): 3–5.

⁵⁶ Ibid. 56–60.

⁵⁷ Presidential elections (April 10–24, 2022) and legislative elections (June 12–19, 2022).

⁵⁸ Arcom (2022b) 46.

⁵⁹ Ibid. 48.

⁶⁰ Ibid. 55.

⁶¹ Ibid. 45.

⁶² Ibid. 52.

⁶³ Arcom (2024): 32–33.

⁶⁴ Ibid. 36.

⁶⁵ Ibid. 34.

partnered with media and civil society to launch initiatives aimed at promoting critical thinking among voters.⁶⁶

Both reports related to election campaigns conclude that while progress has been made, continued cooperation between platforms, regulators, and civil society is essential to strengthen resilience against online disinformation.

Overall, the reports underscore several achievements made by platforms in combating disinformation. These include the development of mechanisms for users to report false content, the promotion of verified information sources, and collaboration with fact-checkers. Despite these advancements, Arcom has identified persistent challenges, particularly concerning the transparency of algorithmic processes and the lack of detailed data. Arcom calls for platforms to improve their transparency towards both regulators and the general public, in order to build greater trust in their efforts to tackle disinformation.⁶⁷

3. Other measures under the Infox Law

3.1. Ban of russian media outlets in France

According to the conditions laid down for the new competencies of the Authority in terms of the broadcasting conventions and concluding such conventions (See point 2.1.2), the Authority may reject the request for a broadcasting agreement or even stop or suspend broadcasting in certain situations, in connection to the dissemination of disinformation or other unlawful conditions. Because of the Russian war of aggression against Ukraine, permission for several Russian media outlets to broadcast in France was rescinded. However, these sanctions were based not on the provisions prescribed by the Infox Law or the Law L  otard but on European Law.

The European Council first imposed sanctions on the state-owned media outlets Russia Today (RT) and Sputnik on March 1, 2022, less than a week after the conflict broke out. As a part of a broader sanctioning package, these companies' broadcasting activities in the EU or directed into the EU were urgently suspended in the member states.⁶⁸ RT France was also suspended by this Council decision.

The Preamble of this decision states several reasons that could have underpinned Arcom's action as well as the basis to be found in the Law L  otard, such as the "propaganda has repeatedly and consistently targeted European political parties, especially during election periods, as well as targeting civil society, asylum seekers, Russian ethnic minorities, gender minorities, and the functioning of democratic institutions in the Union and its Member States". The European Council also "called on Russia and Russia-backed armed formations to stop their disinformation campaign". However, since the European Union

⁶⁶ Ibid. 36–41.

⁶⁷ Arcom (2022c): 3–5.

⁶⁸ Council Decision (CFSP) 2022/351 of 1 March 2022.

acted first, Arcom did not launch its own procedure, and it is unclear whether it would have been possible to act against those under French jurisdiction based on the Law Léotard.⁶⁹

When the Russian-Ukrainian conflict broke out, Laurent Lafon, senator and president of the Commission of Culture, Education and Communication of the Senate,⁷⁰ issued a letter to the president of Arcom, Roch-Olivier Maistre, asking for the suspension of RT France. He based his demand on the Law Léotard which provides that the exercise of this freedom may be limited in the name of respect for human dignity, pluralism of expression of currents of thought, public order or national defense. In this letter, he pointed out forms of non-compliance with RT France's agreement with the Authority on respect for pluralism and 'discriminatory behavior'. He claimed that the broadcaster failed to give a voice to the opposition in Russia, to question the Russian people and to highlight the "disastrous consequences of this aggression on Ukrainian civilians". The President of the Authority explained in his answer that Arcom is particularly vigilant in ensuring that RT France complies with its legal and contractual obligations. If the regulator deems it justified, they may go as far as requesting the suspension of its broadcasting if there were an accumulation of evidence of breaches by RT. The President based this statement on a procedure based on Article 46-10 of Law Léotard⁷¹ which represents a different legal instrument to those provided by the Infox Law.

In the end, the European Union acted more quickly than Arcom, so there was no need to apply such sanctions. By contrast, the German regulator actually announced the prohibition of the German-speaking Russia Today channel before the armed conflict, on December 22, 2021, in reply to which Deutsche Welle was also sanctioned in the Russian Federation.⁷²

In summary, it can be found that the new provisions dedicated to tackling disinformation introduced by the Infox Law did not come into play in this serious situation in which there was a disinformation campaign. This is despite the fact that the incidents mentioned above were within the three-month pre-election period, a fact which could have justified calling for special procedures. In addition, even though serious sanctions were applied against Russian media outlets in other countries, France only acted after a common EU decision was made.

3.2. Judicial proceeding under the Infox Law

The French legislation on judicial proceedings to tackle the spread of false information online during national election periods was considered effective by several authors,⁷³ especially in conjunction with the possibility of banning foreign media outlets under the Infox Law.

⁶⁹ Amaury de ROCHEGONDE: Faut-il interdire RT France? *Stratégies.fr*, 25 February 2022. <https://www.strategies.fr/actualites/medias/LQ280308C/faut-il-interdire-rt-france.html>

⁷⁰ Commission de la culture, de l'éducation et de la communication du Sénat.

⁷¹ Marina ALCARAZ: La chaîne russe RT de plus en plus contestée en Europe occidentale. *Les Echos*, 23 February 2022. <https://www.lesechos.fr/tech-medias/medias/rt-de-plus-en-plus-contestee-en-europe-occidentale-1389222>

⁷² Marina ALCARAZ: L'Allemagne interdit la diffusion de Russia Today; la Russie ferme Deutsche Welle. *Les Echos*, 2 February 2022. <https://www.lesechos.fr/tech-medias/medias/lallemagne-interdit-la-diffusion-de-russia-today-1384111>

⁷³ TENOVE (2020) 10.

However, since it has come into force, there has been only one proceeding before the interim relief judge under Article L.162-2 of the Electoral Code modified by the Infox Law to stop the dissemination.

It was a case of Christophe Castaner, the Minister of the Interior (*ministre de l'intérieur*), who published a tweet on Twitter on May 1, 2019. The post said, that “Here, at the Pitié-Salpêtrière hospital, a hospital was attacked. Medical staff were attacked. And a police officer mobilised to protect was injured. Unwavering support for our police: they are the pride of the Republic.”⁷⁴ The post was accompanied by pictures of the Minister surrounded by a doctor or police officers while talking or shaking hands.

European Parliament member Marie-Pierre Vieu and Senator Pierre Ouzulias launched the procedure on May 10, 2019 at the High Court of Paris (Tribunal de Grande Instance de Paris, referred to as the Court) and asked for the removal of the tweet. The Court held a hearing six days later, and the decision was made on May 17, 2019.⁷⁵ The Court refused the withdrawal request and ordered the requesting parties to pay the costs.

The judgment is significant not only because it was the first time these provisions were applied, but also because the Court interpreted the provisions of the Election Code in question. The Court considered the real-life circumstances of the statement in the post. Among other things, it refers to a report in the newspaper *Le Figaro*, which indeed claims that there was an intrusion into the hospital building during a demonstration where brief acts of non-violence took place. It also relies on the investigation of the Paris public prosecutor's office, during which 32 people were arrested by the police on May 2, and the testimony of the hospital director related to people who had made violent and threatening gestures when the incident occurred. Among the testimonies, some witnesses claimed that they simply wanted to get away from the tear gas applied by police officers during the demonstration.

Firstly, from all the documents produced by the parties, the Court found that although the message written by Mr. Christophe Castaner appears exaggerated in that it refers to an attack and injuries, this exaggeration relates to facts that are real, namely the intrusion of demonstrators into the grounds of the Hospital of Pitié-Salpêtrière on May 1, 2019. Consequently, the information is not unrelated to actual facts, and the requirement of Article L.163-2 of the Electoral Code – that the allegation must be manifestly inaccurate or misleading – is not met.

Secondly, it also analyzed the other conjunctive conditions defined by law, such that the allegations are deliberately, artificially or automatically disseminated on a massive scale via an online public communication service. Among these criteria, the Court emphasized that the character of being artificial or automated was not met, according to the explanatory memorandum of the law, as it primarily refers to sponsored content – by paying third parties to artificially extend the distribution of information – and content promoted using automated tools, i.e. by using bots.

Thirdly, the interim relief judge must assess the manifest nature of the risk of altering the fairness of the ballot linked to the dissemination of this tweet. In this matter, the Court took into account not just the circumstances which motivated publication of the post, but also the public discourse that followed the actions of May 1, 2019 in the Hospital of Pitié-Salpêtrière. While the tweet may have used exaggerated language, as has just been con-

⁷⁴ <https://x.com/CCastaner/status/1123664392011304961>

⁷⁵ TGI Paris, 17 mai 2019, n° 19/53935.

sidered by the decision, the Court found that it did not alter the public debate on the facts and did not “overshadow the debate”, since it was immediately contested by numerous articles in the press and on the Internet that events had not unfolded in the way the Minister described them. The Court also stated that even if different versions had emerged, each voter could form an informed opinion without any obvious risk of manipulation.

Several observations can be made when assessing the specific case and this special procedure introduced by the Infox Law. Firstly, even if the legal instruments created by the law appear to be effective, they may not be applicable in practice or applied by the parties entitled to take the initiative. This is evidenced by the fact that this law, adopted in 2018, has only been applied in one such case. The outcome of the present procedure also demonstrates that it is challenging to prove that all of the conditions set out in the law are fulfilled. The present case also supposes enforcement difficulties, as the Court could not meet the 48-hour deadline imposed by the Infox Law. However, it is an interesting aspect of the proceedings that the Court placed great emphasis in its reasoning on the ability to manipulate public opinion, with respect to which it did not only take into account the specific terms used by the Minister, but also the evolution of the public debate surrounding the case.

3.2.1. Elections (previous and future)

As it has been already mentioned, the French have several national elections, legislative, presidential, European elections and referendums. Unless it is necessary to organize early national elections, there will be no need to apply special rules for the electoral period in 2025. The next elections will be held in March 2026, when the French will vote for municipal representatives, but the Infox Law is not applicable in this case. The mandate of the current president lasts until 2027.⁷⁶ Consequently, no new developments are expected until 2027 on the application of disinformation provisions in relation to elections prescribed by the Infox Law and the SREN Law.

3.3. Other initiatives to counter online disinformation

In recent years, the rise of disinformation and fake news encouraged the development of several dedicated fact-checking organizations and initiatives in addition to the above-mentioned *Pharos* platform, or the media literacy tools for teachers by *Eduscol*.⁷⁷ The following platforms aim to verify information, debunk false claims, and promote media literacy among the public. This chapter explores some prominent French fact-checking organizations and their approaches to combating disinformation.

One of the earliest fact-checking platforms in France is *HoaxBuster*,⁷⁸ a collaborative platform founded in 2000. HoaxBuster primarily focuses on debunking viral internet rumors and urban legends. It operates through community contributions, where users can submit suspicious

⁷⁶ <https://www.service-public.fr/particuliers/vosdroits/F1939>

⁷⁷ <https://eduscol.education.fr/1531/education-aux-medias-et-l-information>

⁷⁸ <https://www.hoaxbuster.com/>

claims for investigation. The platform's approach is straightforward: separating verified facts from unsubstantiated claims and providing clear explanations to dispel misinformation.

Another notable initiative is Les Décodeurs,⁷⁹ launched in 2014 by Le Monde, one of France's leading newspapers. Les Décodeurs investigates questionable claims made by public figures and analyses trending news topics to identify disinformation. It is really active during election periods, when misinformation campaigns intensify. The platform launched a tool called Décodex in 2017, which helps users assess the reliability of online sources by providing a database of websites rated based on their credibility, but this tool is no longer available.

The *CrossCheck* project was launched by Google News Lab and First Draft News in collaboration with several regional, national and international French media organizations during the 2017 French presidential election. CrossCheck is a collaborative fact-checking initiative that brought together journalists from different outlets to investigate suspicious claims and provide verified information. By fostering collaboration across media organizations, CrossCheck aimed to reduce the risk of political bias and enhances the credibility of fact-checking efforts.⁸⁰ In-depth interviews found that the collaboration was credible, fostering media literacy as it taught verification skills to users. The project's success during the election paved the way for similar initiatives in other countries.⁸¹

In addition to such platforms run by media outlets, there are independent content creators such as Hugo Travers, who runs the YouTube channel *Hugo Décrypte*,⁸² who have gained significant influence.⁸³ His content targets younger audiences, providing concise and easily digestible explanations of current events while debunking misinformation.

Les Observateurs, launched by *France 24*, takes a different approach by encouraging citizen journalism.⁸⁴ It invites viewers to submit photos and videos for verification by professional journalists. This participatory model empowers citizens to take an active role in combating disinformation while promoting transparency in the media.

In conclusion, there are various fact-checking organizations and initiatives which fight against disinformation. From traditional media outlets such as *Le Monde* or *France 24* to governmentbacked tools like Pharos, these platforms adopt diverse strategies to combat fake news. Their efforts emphasize the importance of collaboration, media literacy, and adapting fact-checking techniques to different audiences.

4. Conclusions

The legislative process of the Infox Law provoked a debate not just within the parliamentary debate or in the Council of the State but among other professionals as well. At the end of

⁷⁹ <https://www.lemonde.fr/les-decodeurs/>

⁸⁰ David DIEUDONNÉ: Fact Checking the French Election with CrossCheck, A Collaborative Reporting Initiative. *Medium.com*, 8 February 2017. <https://medium.com/google-news-lab/fact-checking-the-french-election-with-crosscheck-a-collaborative-reporting-initiative-22f14204415c#.75snz7f9s>

⁸¹ First Draft : Research on CrossCheck journalists and readers suggests positive impact for project. *Firstdraft-news.org*, 16 November 2017. <https://firstdraftnews.org/articles/crosscheck-qualitative-research/>

⁸² <https://www.youtube.com/@hugodecrypteactus>

⁸³ More than 3 million subscribers.

⁸⁴ <https://observers.france24.com/fr/>

the day, the Infox Law introduced new instruments to tackle disinformation, such as special procedures to conclude or suspend broadcasting licenses, judicial procedures applicable in campaign periods, reporting mechanisms for online platforms or media literacy tools.

During the first few years after the legislation came into force, there were not many examples of the use of the special procedures of the law. This did not necessarily mean that the procedures are ineffective. Still, the above-mentioned case study proved in France that the application of EU Law is more of a determining factor in these areas than the French approach to regulating online disinformation.

Regarding the monitoring exercises carried out by Arcom, the series of reports illustrates a growing recognition among online platforms of their role in combating disinformation. While significant progress has been made, Arcom stresses that there is still room for improvement, particularly in terms of transparency and accountability. The law does not give the regulator any coercive powers to force operators to meet their obligations. Therefore, Arcom is taking action through open and demanding dialogue with the platforms.⁸⁵ The Authority remains committed to monitoring platform activities and providing guidance to enhance the effectiveness of the measures implemented, but the harmonized EU provisions have complemented or taken the place of national law.

While progress has been made, combating online misinformation requires sustained effort, cooperation, and innovation. According to the examples above, regulatory authorities, platforms, and civil society need to collaborate to ensure that digital spaces become more reliable and trustworthy sources of information.

References

- Arcom (2022a) = Lutte contre la manipulation de l'information sur les plateformes en ligne – Bilan annuel des moyens et mesures mise en oeuvre par les opérateurs de plateforme en ligne en 2021. Arcom, 2022. <https://www.arcom.fr/se-documenter/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/lutte-contre-la-manipulation-de-linformation-sur-les-plateformes-en-ligne-bilan-2021>
- Arcom (2022b) = Rapport sur les campagnes électorales: élection à la présidence de la République et élections législatives. Arcom, 2022. <https://www.arcom.fr/se-documenter/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/rapport-sur-les-campagnes-electorales-2022-election-la-presidence-de-la-republique-et-elections-legislatives>
- Arcom (2022c) = Lutte contre la manipulation de l'information sur les plateformes en ligne – Retour sur trois années d'application de la loi du 22 décembre 2018. Arcom, 2022. <https://www.arcom.fr/se-documenter/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/lutte-contre-la-manipulation-de-linformation-sur-les-plateformes-en-ligne-bilan-2021>
- Arcom (2024) = Rapport sur les campagnes électorales. Arcom, 2024. <https://www.arcom.fr/se-documenter/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/rapport-sur-les-campagnes-electorales-2024-elections-europeennes-et-elections-legislatives>

⁸⁵ Arcom (2022c): 4.

- CSA(2020) = Lutte contre la diffusion de fausses informations sur les plateformes en ligne – Bilan de l'application et de l'effectivité des mesures mises en oeuvre par les opérateurs en 2019. Conseil supérieur de l'audiovisuel, 2020.
<https://www.vie-publique.fr/files/rapport/pdf/275995.pdf>
- CSA (2021) = CSA: Lutte contre la manipulation de l'information sur les plateformes en ligne – Bilan des mesures mises en oeuvre par les opérateurs en 2020. Conseil supérieur de l'audiovisuel, 2021. <https://www.vie-publique.fr/rapport/283257-manipulation-de-l-information-sur-les-plateformes-en-ligne-bilan-2020>
- STELLA (2019) = Elie STELLA: *L'adaptation du droit pénal aux réseaux sociaux en ligne* [Doctoral thesis]. Metz, Université de Lorraine, 2019.
<https://hal.univ-lorraine.fr/tel-02985468>
- TENOVE (2020) = Chris TENOVE: Protecting democracy from disinformation. Normative threats and policy responses. *The International Journal of Press*, Vol. 25., No. 3. (2020) 517–537. <https://doi.org/10.1177/1940161220918740>

Scaremongering – Criminal law responses to fake news and the problem of restrictions on freedom of expression defending public peace in Hungary

ZSUZSANNA MONORI

“A lie can travel halfway around the world while the truth puts on its shoes.”
– Mark Twain

Lying in itself is not illegal. It is a question of who, when, where, and what is being lied about and what fundamental values or constitutional rights might be violated as a result. The pandemic outbreak challenged Hungarian legislators in several ways, but one of the most crucial issues was the explosion of false information which could potentially sabotage the effectiveness measures to protect against the epidemic. One of the most significant legal responses to this problem in Hungary has been the codification of a new criminal offence, which has been the subject of much criticism to date and raises constitutional questions about the restriction of freedom of expression in general.

In the matrix of law, criminal liability should indeed be *ultima ratio*, i.e. the very last resort. By declaring a type of statement as a criminal offence, the legislator sends a serious message and therefore needs to exercise due care in considering the narrow pathways via which freedom of speech can be limited. This paper will explore the question of the uppermost limits: namely which false or misleading verbal expressions and under what conditions reach such a level of danger to society that they require a criminal law response from the legislator – a response imposing the strictest form of liability under a state’s *ius puniendi*. The relevant questions will be examined in the context of the current Hungarian constitutional criminal law (closely connected to constitutionally protected fundamental rights).

This study focuses on the new form of scaremongering as a criminal act; this concerns a criminal regulation which aims to protect the public interest, as an integral and inherent part of the Hungarian criminal law tradition. After considering the elements and public cases of scaremongering, it is necessary to examine the constitutional content of this offence, especially the Hungarian Constitutional Court’s relevant decisions, which conclude on the correct interpretation of the offence and help to respond to the constitutional concerns raised in relation to the offence of scaremongering.

1. Basic principles of criminal law restrictions on freedom of expression

In the European legal mindset, freedom of expression is a fundamental civil right that implies the right to hold and express one's opinions, as well as to receive and impart information and ideas, without prior authorization (censorship) or other forms of interference by the government or any form of public authority.¹ Freedom of speech is one of the most essential fundamental communication rights, although, it is not unlimited. Restrictions upon freedom of speech raise global questions in a marketplace of ideas which has now been operating at a global level for quite some time.

1.1. The constitutional nature of freedom of expression and its restrictions

Freedom of expression is enshrined in the Fundamental Law of Hungary. However, the Law adds that its exercise must not be aimed at violating the human dignity of another, as well as the dignity of the Hungarian nation or any national, ethnic, racial or religious group (Article IX). Although the Fundamental Law mentions the right to human dignity as the only absolute limit to freedom of expression, Hungarian criminal law has a long tradition of restriction in the public interest (namely public peace). Article IX of the Fundamental Law of Hungary states:

- “(1) Everyone shall have the right to freedom of expression.
- (2) Hungary shall recognize and protect the freedom and diversity of the press and shall ensure the conditions for the free dissemination of information necessary for the formation of democratic public opinion.
- (3) In the interest of the appropriate provision of information as necessary during the electoral campaign period for the formation of democratic public opinion, political advertisements may only be published in media services free of charge, under the conditions guaranteeing equal opportunities laid down in a cardinal Act.
- (4) The right to freedom of expression may not be exercised to violate the human dignity of others.
- (5) The right to freedom of expression may not be exercised to violate the dignity of the Hungarian nation or any national, ethnic, racial or religious community. Persons belonging to such communities shall be entitled to enforce their claims in Court against the expression of an opinion that violates their community, invoking the violation of their human dignity, as provided for by an Act.
- (6) The detailed rules relating to the freedom of the press and the organ supervising media services, press products and the communications market shall be laid down in a cardinal Act.”

From a Hungarian legal perspective, it is essential to clarify a conceptual point: freedom of *expression* is not synonymous with freedom of *opinion*, *speech* or *press*, although democratic

¹ CASSESE–COMBA–CARIA (2022): 12.

public opinion can be developed and upheld by these rights of communication.² The Hungarian Constitutional Court – while dealing with the criminal law assessment of hate speech – formulated the category of *communications-related fundamental rights* and stated freedom of expression as its preceding right, or as Hungarians call it, its “preceding right”:

“[T]he freedom of expression has a special place among constitutional fundamental rights. It is the ‘mother right’ of several freedoms, the fundamental right of ‘communication’. The rights deriving from this ‘mother right’ are the right to free speech and the right to the freedom of the press, with the latter encompassing the freedom of all media, the right to be informed, and the right to obtain information freely. In a broader sense, freedom of expression includes artistic and literary freedoms, the freedom to distribute and disseminate works of art, scientific research and the freedom to teach scientific knowledge.”³

In any case, even without prior governmental restraints, freedom of expression is not absolute: it can be lawfully restricted to balance it against other fundamental rights according to the proportionality test. In the case of a proportionality test, the elements required to restrict freedom of expression are the following at the very least:

1. The restriction is prescribed by law;
2. It pursues a legitimate aim (e.g. protection of the rights and freedoms of others, territorial integrity, security of the State, public security, public health, morals, prevention of disorder or crime);
3. The restriction is necessary and suitable for the legitimate aim pursued. In general, under the proportionality test, any restrictions must be necessary to achieve an objective on a constitutional basis.

These points match Article 10(2) of the European Convention of Human Rights (ECHR), according to which limitations on freedom of expression can be envisaged: in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.⁴ The constitutional basis for the above-mentioned limitations means that freedom of expression can be balanced against other constitutionally protected interests.⁵ The Fundamental Law of Hungary (in Article I Section 3) thus defines the conditions for the restriction of fundamental rights in line with the above European legislation:

“The rules for fundamental rights and obligations shall be laid down in an Act. A fundamental right may only be restricted to allow the effective use of another fundamental right or to protect a constitutional value to the extent necessary, proportionate to the objective pursued and with full respect for the essential content of that fundamental right.”

² Constitutional Court Decision No. 30/1992. (V. 26.).

³ Constitutional Court decision No. 30/1992. (V. 26.).

⁴ See also ECtHR, *Observer and Guardian v. the UK* (26 November 1991).

⁵ CASSESE–COMBA–CARIA (2022): 12.

As it is mentioned in *A Commentary of the New Hungarian Constitution*, this phrasing also clears the dogmatic uncertainty emerging from the phrase “constraints on the essential contents” derived from the old Constitution. It also identifies the possible “reasons of necessity” of such limitations as protecting another fundamental right or a constitutional value. Therefore, “abstract public good” alone is unacceptable for limiting fundamental rights, as explained below in the context of offences against public order. The Fundamental Law expressly mentions respect of the essential contents of the fundamental rights, from which it can be understood that fundamental rights have a dimension which is subject to no limitations at all: human dignity, which is inviolable.⁶

“[F]reedom of expression protects the opinion regardless of its value and truth. This freedom only has external limits; unless it collides against such a constitutionally drawn external limit, the possibility of expressing an opinion, and the fact that someone has done so, remains protected irrespective of its content.”⁷

In Hungarian literature, *Koltay* writes that freedom of speech (and of the press) protects all opinions, views, statements and expressions that are not limited by law, and these restrictions must be duly justified and sufficiently narrow in all cases. The constitutionality of a restriction of a fundamental right is conditional not only on its necessity and proportionality but also on its having a sufficiently compelling reason. In the case of freedom of expression, the basis for a restriction can only be an external reason, never the content of the opinion.⁸ Therefore, Hungarian constitutional law distinguishes between criticisms containing factual statements and opinions. The freedom of expression protects an opinion on the honor and goodwill of public actors. Factual statements do not enjoy such protection if the person who made the statements knew that the statements were untrue or they did not know because of their failure to meet professional standards applicable to their conduct.⁹

1.2. The unique role of criminal law in the Hungarian legal system – the last resort

The connection between the Constitution and criminal law in the Hungarian legal system is very complex. The Fundamental Law is a supreme source of Hungarian law, while criminal law is an area of law with a substantive system of liabilities and sanctions, which may be conceived as an essential limitation on the basic rights specified under the Fundamental Law. Meanwhile, the constitutionality of criminal law means that the penal law must be based on and adjusted to the Fundamental Law, as the basic norm formulating and mediating basic values.¹⁰

⁶ ÁDÁNY et al. (2015): 82.

⁷ Constitutional Court decision No. 36/1994. (VI. 24.).

⁸ KOLTAY (2019): 240.

⁹ ÁDÁNY et al. (2015): 100.

¹⁰ KARSAI–SZOMORA (2015): 42.

Hungarian criminal law belongs to the family of public law systems and has been significantly influenced by the German and Austrian models.¹¹ The primary source of statutory criminal law is Act C of 2012 on the Criminal Code (from now on referred to as the Criminal Code), although, in a broader sense, Hungarian criminal law also includes procedural and enforcement laws. This paper will only briefly review the provisions of the Criminal Code since its provisions determine what conduct constitutes a criminal offence.

Hungarian law (like other continental legal systems) does not recognize any legal consequence more severe than criminal liability. Even though the compensation in a civil case may be higher than the fine imposed in a criminal proceeding (e.g. in the case of defamation), the emphasis is on the nature of the liability. However, to understand the gravity of criminal liability, it is necessary to briefly explain *the ultima ratio* and *subsidiarity*, which are unique characteristics of criminal law.

In the continental legal system, legality is the basic principle in criminal substantive law (*nullum crimen - nulla poena sine lege*).¹² In the current legal environment, the principle of legality is not only declared by the Criminal Code (Article 1) but also by the Fundamental Law: Article XXVIII paragraph 4 states that “nobody can be declared guilty and be punished for conduct that was, at the time of its perpetration, not a criminal offence either under Hungarian law or provided for by an international convention or a legal instrument of the European Union”. As the Hungarian Constitutional Court stated in the following decision, criminal law might be used only as a method of last resort (*ultima ratio*) among other legal institutions and sanctions:

“Criminal law is the *ultima ratio* in the system of legal liability. Its social function is to be the sanctioning keystone of the legal system. The criminal law sanction, as well as the role and purpose of punishment, is to maintain the integrity of the law and moral norms when sanctions of other branches of law no longer help [...] It is a substantive requirement that the legislator may not act arbitrarily when determining the conduct to be punished. A strict standard must judge the need to criminalize any conduct: the use of criminal law instruments, which inevitably restrict human rights and freedoms to protect various situations under moral and legal standards, is only justified in an essential case and to a proportionate extent, if the protection of state, social, economic objectives, values, which are constitutional or traceable to the Constitution, is not possible in any other way.”¹³

The rule of law requirements arising from the nature of the *ultima ratio* are to resolve conflicts in society, namely branches of law outside criminal law, primarily. In the event of an infringement, legal integrity must be ensured by applying legal sanctions governed by other branches of law. The aim of criminal law is, therefore, nothing more than to redress the social imbalance created by an unlawful act. This imbalance entails the violation or endangerment of interests and values listed in the Fundamental Law. Also, under the general test for the restriction of a fundamental right, the basis for the restriction must always be in the in-

¹¹ KARSAI-SZOMORA (2015): 20.

¹² BELOVICS (2020): 140.

¹³ Constitutional Court Decision No. 30/1992. (V. 26.).

terest of some other value. These protected values, which are harmed or endangered by a particular criminal act, are called “legal objects”.

In Hungarian criminal law, the starting point for analyzing a given offence is always the legal object. As *Géza Tokaji* writes: “The basis for social endangerment is that the act offends or endangers a legal object [...] an act which neither offends nor endangers the legal object cannot be considered socially dangerous and consequently cannot be considered a criminal offence”.¹⁴ *Ferenc Nagy* states, “The legal object is the legally protected abstract value of the social order, the enforcement of which is in the interest of society and of the community”.¹⁵ In Hungarian criminal law, another legislative function of the legal object is determining the place, structure and possible sanction of a given criminal offence in the system. In contrast, the legal practice can draw from it in interpreting the elements of the offence, and correct interpretation is unthinkable without the definition of the legal object (this is the so-called ‘teleological interpretation’).

Social dangerousness is the primary statutory component of a criminal offence, which is also evident from the general legal definition of the offence. As Section 4 of the Criminal Code states:

1. A criminal offence shall mean an act committed intentionally or, where negligent commission is punishable under this Act, negligently, which is a danger to society and is subject to punishment under this Act.
2. An activity or omission shall constitute an act that is dangerous to society if it harms or endangers the person or rights of others or the social, economic or State order of Hungary as laid down by the Fundamental Law.

Thus, the general legal objects underlying the danger to society are the constitutional, social, economic and State order and the persons and rights of others.¹⁶ All the individual and collective legal objects protected by the specific part of the Criminal Code can be derived from these. Suppose the Court concludes that the act of committing the offence for some reason does not violate or endanger the legal object of the offence. In that case, the defendant may be acquitted based on the lack of danger to society (for example, in case¹⁷ of the use of prohibited symbols was not intended to express identification with the dictatorial ideals they embodied but to dissociate oneself from them.)¹⁸

1.3. Criminal law restrictions on the grounds of public interests

Wiener raises the question concerning all constitutional fundamental rights, whether it is possible, and under what conditions, to restrict those rights and what criteria should be used for establishing priority in case of their collision.¹⁹ The external reason for restriction may be the protec-

¹⁴ TOKAJI (1984): 103.

¹⁵ NAGY (2010): 104.

¹⁶ SZEMÁN (2022): 105–120

¹⁷ Curia decision No. BH2009. 131.

¹⁸ SZOMORA (2013): 42.

¹⁹ WIENER (1997): 27.

tion of another fundamental right, the safeguarding of its exercise, or another value protected by the Fundamental Law. *Koltay* defines three main categories of interests justifying restrictions and mentions that the different possible grounds for restriction will carry different weight in the necessity test, depending on what they are intended to protect and how:

- a) the interests of the community (society) as a whole (e.g. protection of constitutional order, external and internal security, protection of state symbols, protection of public morals, protection of plural and diverse public opinion);
- b) the interests of particular social groups (e.g. the prohibition of the expression of racial, religious or ethnic hatred);
- c) the interests of individuals (e.g. protection of individual rights and human dignity).²⁰

While the present paper focuses on restrictions on freedom of expression for reasons of public interests (namely public peace), it is essential to note that individual interests, namely the right to human dignity, often appear as an “underlying” or “secondary” protected value (e.g. in the case of banned symbols, the right of the relatives of victims of extremist regimes, or the dignity of a specific group in the case of incitement of hatred). A coherent framework of constitutional standards requires a clear distinction between the different levels of protection: public peace per se and subjective (individual) rights. In the case of the first, stricter conditions are required for restriction than legislation protecting “underlying” individual rights.

Based on an analysis of the relevant Constitutional Court decisions, *Koltay* concludes that the limitation threshold differs between the protection of individual and public interests. While the protection of public order alone would be a constitutional basis for restricting freedom of speech only if the risk of infringement is clear and present, the standard is lower in the case of individual rights affected by the exercise of freedom of speech, while its endangerment is a condition for the occurrence of a criminal offence. “In any case, in the enforcement of the law, it is not possible to separate acts that violate or threaten public peace from those that violate or threaten individual rights; the threat to individual rights entails a threat to public peace, and vice versa: public peace can be threatened by the threat to individual rights.”²¹ As the Constitutional Court stated:

“[a] restrictive law to be weighed against freedom of expression has greater weight if it is directly aimed at the enforcement and protection of another fundamental right, less if it is only aimed at such rights in an underlying manner [...] and least if it is only aimed at an abstract value in itself (e.g. public peace). [...] An abstract, potential threat to public peace is not sufficient justification for a criminal law to restrict the fundamental right to freedom of expression, which is essential to the functioning of a democratic state governed by the rule of law.”²²

This significant 1992 decision of the Constitutional Court indicates the stricter conditions: the risk of harm to public peace must be *clear* and *present*. The Constitutional Court later applied this test when it held that the prohibition of scaremongering was constitutional only in the event

²⁰ KOLTAY (2019): 239.

²¹ KOLTAY (2020a): 325.

²² Constitutional Court Decision No. 30/1992 (V. 26.).

of public danger or in times of war, where there is a clear and present danger of public order being violated.²³ However, this is not the test e.g. in the case of incitement to hatred. Since the incitement also endangers individual rights (it also violates the human dignity of a particular group), even a somewhat less severe form of danger may justify a restriction. Consequently, as *Bernát Török* points out, the “underlying” protection of the subject’s rights is greater than that of the abstract value of public peace; thus, a lesser degree of danger may lead to action.²⁴

2. Scaremongering – a controversial regulation against fake news

During the pandemic, there was a broad circulation of news, which, to a certain extent, made it difficult to determine the accuracy of the news. However, whether real or hoaxes, news can be extremely disturbing to people’s peace of mind, which had the potential to indirectly hamper the response to the pandemic. The effective detection of these crimes has remained a priority for society, which is why the Hungarian legislator made it a criminal offence to make untrue statements (which thus also fell into the category of fake news) about the epidemic that could disturb public peace.²⁵ Therefore – and because Paragraph 1 was not suitable for sanctioning scaremongering in respect of the epidemic – a second scaremongering statutory element was introduced into the act in the spring of 2020, which is now in Paragraph 2. The Constitutional Court has laid down as an explicit requirement that statements are only punishable if the perpetrator must have known at the time of committing the offence that the statements were false or if the perpetrator distorted them, and if the statements had the potential to hinder or defeat measures in the special legal order to protect against the epidemic.

“The Court may conclude that the subject of this criminal offence may not be the criticism of certain measures taken by the government during the special legal order, forecasting future events, or speculating on data not disclosed in the context of the special legal order. [...] Scaremongering may only be committed intentionally. Consequently, the perpetrator must be aware that they are committing their act at a time of a special legal order, that the fact they allege is untrue or has significantly distorted the true fact, and that the communication of their allegation is (objectively) capable of hindering or frustrating the effectiveness of protection. Furthermore, the perpetrator’s intent must include committing the above-mentioned conscious communication in front of a large audience. If his consciousness or intention does not extend to any of these elements, or if the allegation is not objectively capable of hindering or frustrating the effectiveness of the protection, the offence is not committed according to the rules of criminal law.”²⁶

The prevailing view is that, under the constitutional protection of freedom of expression, lying generally cannot be prohibited simply because it is untrue. However, legal action

²³ Constitutional Court Decision No. 18/2000 (VI. 6.).

²⁴ TÖRÖK (2013): 66.

²⁵ AMBRUS (2020): 13.

²⁶ Constitutional Court decision No. 15/2020. (VII. 8.).

against falsehoods is still possible in certain circumstances, such as defamation, perjury and Holocaust denial. Such prohibitions also cover scaremongering, i.e. if the relevant offence is committed, the untrue statement becomes punishable under certain conditions.²⁷ Scaremongering can be committed by making untrue statements or rumors and distorting (misrepresenting) a fact. Distorting a fact is not merely a shocking, sensationalist reporting of facts but also requires the perpetrator to mix the initially true facts with untrue elements so that the combined effect is untrue.²⁸ Under Section 337 of the Criminal code, scaremongering means:

1. A person who, at a site of public danger and in front of a large audience, states or disseminates any untrue fact or any misrepresented fact concerning the public danger that is capable of causing disturbance or unrest in a larger group of persons at the site of public danger is guilty of a felony and shall be punished by imprisonment for up to three years.
2. A person who, during the period of a special legal order and in front of a large audience, states or disseminates any untrue fact or any misrepresented fact that is capable of hindering or preventing the efficiency of protection is guilty of a felony and shall be punished by imprisonment for one to five years.

This offence has long been part of the Hungarian Penal Code, and quite a few constitutional court decisions have been made about its interpretation and its relations with fundamental rights.²⁹ Two of these decisions are essential to highlight here: Constitutional Court Decision No. 18/2000 (VI.6.), which (in the 1978 Criminal Code) annulled the old definition of scaremongering, and Decision No. 15/2020 (VII.8.) on the constitutionality of the amendment called into force by the pandemic in the current Criminal Code.

2.1. Defining the crime of fake news: the statutory elements of scaremongering

Since the second paragraph is relevant to this paper, its element shall be analyzed in greater detail below, such as the falsity of a statement, the meaning of special legal order, the capability of hindering or preventing any protection and what constitutes a large audience. The elements contained in the Criminal Code of 1978 were amended in 1989 at the time of the political transition, which introduced the requirement that the offence be committed in front of a large audience as a mandatory legal element, thus narrowing the scope of punishable acts. The constitutionality of the prohibition in Section 270 of the old Penal Code³⁰ was examined much later when the provision was annulled in its entirety. The reasoning

²⁷ KOLTAY (2020a): 322.

²⁸ AMBRUS (2020): 13.

²⁹ For the history of this offence see DAVID (2021) 100–104; and KOLTAY (2020b): 124–126.

³⁰ Section 270 (1): A person who, in front of a large audience, states or disseminates any untrue fact or any misrepresented fact that is capable of causing a disturbance is guilty of a misdemeanour, it shall be punished by imprisonment for up to one year, community service or a fine. (2) The punishment shall be imprisonment for a term not exceeding three years for a criminal offence if the scaremongering is committed in public danger or during a war.

was based on the fact that public peace is not a specific right or interest that can generally justify a restriction on freedom of expression. According to the Constitutional Court, such restriction is only constitutionally possible if the intensity of the disturbance of the public peace rises to the level of a clear and present danger, i.e. if there is a manifest and imminent danger to the public peace. The offence was re-codified by the legislator in Act CXXV of 2000, with the new provision considering the intentions of the Constitutional Court and penalizing only communications at a scene of public danger that could cause confusion or anxiety among a larger group of people.³¹

The purpose of the regulation is to protect the public mood and keep it free from untrue or distorting news which causes disturbance or unrest. Therefore, objective (i.e. factual) information cannot constitute a criminal offence even if it causes disorder among a large section of the population.³² The protected legal object in the crime of scaremongering is public peace. Public peace is a social atmosphere in which respect for law and order, mutual respect and recognition of the legitimate interests of citizens prevail. The result of offences against public order is public disturbance and disorder.³³ Two types of conduct can constitute a criminal offence: stating or disseminating false or misrepresented facts. When a fact is false, i.e. untrue, it means that the fact did not happen, did not exist and does not exist. Distortion of a fact means the communication of an objectively existing fact in such a way that the perpetrator takes it out of its context in a biased way, taking the favorable or unfavorable aspects according to his point of view. The offence is committed by someone who either states facts which he has invented or distorted himself, or who passes on such statements which he has heard from others.³⁴ The most thought-provoking thing about the definition is that it requires *capability* (in other words: potential), not the actual hindrance of the prevention. This also means that the offence can be committed even if there has been no failure of the effective protection against the circumstance giving the ground for the special legal order.

The second version of scaremongering can be committed only during the period of a special legal order, which determines both the place and the time of the offence. In its Article 48, the Fundamental Law declares that special legal orders shall include state of war, state of emergency and state of danger. The reason for the introduction of the special legal order in Hungary on 11 March 2020 was that the WHO declared the COVID-19 pandemic a global pandemic on 11 March 2020 [declared by Government Decree No.40/2020 (11.III.)]. While the first version of the offence under paragraph 1 stated that the offence could only be committed in a place and connection with a public danger, the new paragraph 2 extends the place of Scaremongering to the whole country.³⁵ The Explanatory Memorandum to the Bill on the Protection Against Corona Virus itself emphasizes that the line between general threats and public danger is difficult to draw. That is why the legislator has chosen to place public peace under general criminal law protection throughout the country, regardless of the exact place of the public danger.³⁶ However, this broader protection is not limited to the time of the pandemic, as the legislator may have in-

³¹ KOLTAY (2020b): 9–10

³² MEZŐLAKI (2013): 701.

³³ BELOVICS (2015): 559.

³⁴ MEZŐLAKI (2013): 702.

³⁵ DOMOKOS (2020): 81.

³⁶ <https://www.parlament.hu/irom41/09790/09790.pdf>

tended that – even after the epidemic has passed – other circumstances may arise later on which justify a higher protection of public peace [subsequently, because of the armed conflict and humanitarian crisis on the territory of Ukraine, a state of emergency was declared by Government Decree No 424/2022 (28.X.)].

According to Hungarian judicial practice, when a crime is committed “in front of a large audience”, it means that the offence is committed in front of an indefinitely large number of people – the number of whom cannot be determined by a mere glance – or that there is a possibility (e.g. the constant turnover of people present or because any type of media presence) that a more extensive or an indeterminable number of persons may become aware of the facts.³⁷ Section 459(1)(22) of the Criminal Code includes offences committed through a media product, media service, reproduction or publication on an electronic communications network (e.g. the Internet). It may seem to be a cliché, but the world of the Internet has not only amplified the power of the spoken word but has also escalated the damage it can do; national legislators have to respond to this. The “clickbait methods” used in online journalism have the potential to be considered illegal,³⁸ whether we think of some form of hate speech or scaremongering (or even defamation or slander), the vast majority of these – like all our everyday communications – occur on the Internet, where the condition for the “large audience” is met.

The other essential element is the intent of the perpetrator: the Hungarian criminal law is based on subjective liability, where proof of the perpetrator’s intent (or lack thereof) is essential. The offender’s knowledge and intent must cover each definitional element of the offence, including the falsity or dangerousness of the alleged fact. In this case, negligent acts are not punishable by law.

2.2. Critiques and the constitutionality

After the introduction of the new version of scaremongering, critical voices have been raised, arguing that the new offence is unconstitutional, mainly because the new definition does not meet the requirement of clarity and ultimately constitutes an unnecessary and disproportionate restriction on freedom of expression. Critics focused on the uncertain elements of the new prohibition: what makes a communication potentially hindering in the protection and how this will be proven in a criminal procedure. This is linked to the “endangering nature” of the offence since it does not require real obstruction to proper protective measures. In its decision No 15/2020 (VII. 8.), the Constitutional Court ruled that the new criminal law legislation on Scaremongering is not unconstitutional. This chapter discusses the main arguments and relevance of this decision.

According to the constitutional complaint,³⁹ the new legislation restricted the right to freedom of expression, was completely unpredictable, provided wide scope for arbitrary application of the law, was incapable of interpretation and left the norm’s recipients in un-

³⁷ Curia decision No. BH1983.108.

³⁸ HERKE (2024): 42–44.

³⁹ [http://public.mkab.hu/dev/dontesek.nsf/0/bd83430c4d2a942ac125855e005c4028/\\$FILE/IV_699_0_2020_inditvany_anonim.pdf](http://public.mkab.hu/dev/dontesek.nsf/0/bd83430c4d2a942ac125855e005c4028/$FILE/IV_699_0_2020_inditvany_anonim.pdf)

certainty. Thus, the legislation did not comply with the principles of constitutional criminal law and the relevant practice of the Constitutional Court, consequently, it should be annulled. In the view of the complaint, the restriction of freedom of expression to 5 years' imprisonment was neither necessary nor proportionate during the period of the special legal order. It also raised the issue of negligent lying: if a person makes a statement in good faith, i.e. did not know of the untrue basis of the statement or news.⁴⁰ The petitioner agreed that the purpose of the law is constitutionally justified and that panic should not be allowed while a special legal order is in force. However, Section 337(2) of the Criminal Code leaves everyone concerned in total uncertainty: practically anyone who speaks out on public affairs during a special legal order or even just uses social media. It also draws attention to the fact that the investigating authority is not in a position to decide whether an expression, even a false or distorted one, is currently capable of hindering a protection.⁴¹

In its decision, the Constitutional Court found that the contested provision applies only to a narrow range of statements. The prohibition applies only to *knowingly* untrue or distorted statements, not to critical opinions. Therefore, the criminal law restriction does not cover false statements if the perpetrator did not know them to be untrue. However, where a person knowingly makes false allegations which may hinder the protection, the restriction of freedom of expression is necessary and proportionate for the public interest (namely public peace). The Constitutional Court has underlined that the perpetrator must be aware of the falsity of the facts *at the time of the offence*, and the offender must be aware that the offence is or may be (i.e. *capable of being*) an obstruction to the protection during the special legal order. If a fact is disputed at the time of the offence and only later proved to be false, it does not constitute scaremongering and therefore the person cannot be held criminally liable.

In the Hungarian legal literature, several opinions have been published on the new rule and the decision of the Constitutional Court. As a general principle, *Tamás Klein* states that during the special legal regime, including in times of emergency, it may be necessary to restrict freedom of expression more extensively than during normal constitutional times. However, such a restriction of fundamental rights can only be considered constitutionally justified if it is necessary to avert the emergency, to avert danger, to protect life and property, to remedy the damage caused and to restore the normal constitutional legal situation.⁴² According to *Philip Gubis* “preventing” in the definition is unnecessary, since the capacity to hinder includes the capacity to make the result of the protection completely impossible, i.e. to obstruct it. He notes that “protection” is a very broad concept, covering the whole spectrum of government action across society from health, education and border restrictions to economic measures. Gubis argues that the very concept of protection is itself very broad and that if we add to this the fact that only “capacity” is required to prevent or defeat it, the question arises of whether the criminal law is overly strict. He does, however, indicate that supposed difficulties of interpretation will be avoided.⁴³ In *Andrea*

⁴⁰ *Nem alaptörvény-ellenes a rémhírtérjesztéssel kapcsolatos új – különleges jogrendben alkalmazandó – büntetőjogi szabályozás* [Press release]. The Constitutional Court of Hungary, 17 June 2020. <https://www.alkotmanybirosag.hu/kozlemeny/nem-alaptorveny-ellenes-a-remharterjesztessel-kapcsolatos-uj-kulonleges-jogrendben-alkalmazando-buntetojogi-szabalyozas/>

⁴¹ DOMOKOS (2020): 82.

⁴² KLEIN (2020): 222.

⁴³ GUBIS (2021): 61.

Domokos' opinion, the authentic interpretation of the concepts of public danger and epidemic are both missing from the Criminal Code. Both are needed and the same concepts must be used throughout the legal system to ensure a coherent view. Concerning scaremongering, she points out that there is also a place for statutory definitions of endangerment in the Criminal Code, but the requirement of clarity can be expected to produce a jurisprudence that is consistent with the rule of law and satisfies society's sense of justice.⁴⁴

According to *András Koltay*, the legal concepts may indeed be misleading if we start from the general grammatical meaning of the words, but the decision analyses in detail the requirement of intentionality, i.e. what exactly the perpetrator's intention must be when stating an untrue fact. The decision establishes a constitutional requirement that it applies to a range of relevant statements and strengthens the protection of freedom of expression. The criterion of intentionality leaves no doubt as to whether the sole proper interpretation is in favor of freedom of expression. He stresses, however, that the case law must make it clear from the beginning that it cannot silence public debate and critical opinions. This requirement is clear from the Constitutional Court's rulings, as is the fact that scaremongering is now punishable only in specific situations, on a particularly justified and narrow basis, leaving further clarification to judicial practice.⁴⁵

In a recent publication, *Zsanett Dorang* presented the factual elements of scaremongering, the Constitutional Court decision 15/2020 (VII.8.) and the criticism of that decision. In her opinion, the critics surrounding the amendment are not accurate, as the final decisions have not proven that the real purpose of the regulation was to silence voices critical of the government and its actions. Although, as the above examples show, criminal prosecutions started about political speech or criticism of government action, no prosecutions have been brought in any of these cases, and the authorities consistently maintained that such communications do not fall within the offence. However, it also raises the question of the preventive (i.e. chilling) effect of the possibility of criminal proceedings (becoming a suspect, interrogation, search, search and seizure of electronic devices) in cases of statements, regardless of whether the authorities subsequently terminate the proceedings for lack of offence.⁴⁶ It is important to note here again that in the matrix of different areas of law, criminal liability should be an indeed *ultima ratio*, i.e. the very last resort. By declaring a type of statement as a criminal offence, the legislator sends a serious message and therefore needs to exercise due care in considering the narrow pathways via which freedom of speech can be limited. As can be seen from the examples and case law below, these concerns have proved mostly unfounded.

2.3. Examples from the legal practice

Before discussing some specific cases, it is important to note that the outcome of a criminal case depends largely on the prosecutor: it is they who decides in which cases an indictment and charges will be brought, otherwise, the case will not go to court. The prosecutor is

⁴⁴ DOMOKOS (2020): 83.

⁴⁵ KOLTAY (2020a): 335–338.

⁴⁶ DORANG (2023): 290.

therefore the “master” of the case in interpreting the elements of the offences and determining how they can be committed. This is particularly relevant in the case of a criminal offence where certain elements are allegedly uncertain and the requirement of clarity of the rules may be infringed.⁴⁷ It should also be mentioned here that court decisions and prosecution documents for specific cases are not available, so the sources are press reports and official authority statements.

According to an official police statement in 2020, 87 prosecutions for scaremongering had been initiated a few months after the amendment. Of these, 38 cases were pending, 6 were at the prosecution stage and a total of 16 suspects had been questioned during the investigations.⁴⁸ Although detailed statistics for the last four years are not available, it can be stated that following the amendment, a large number of charges were brought to the investigating authorities in a relatively short period (130 prosecutions in total by 2023, according to official communications), but the number of prosecutions and convictions was very low. The fact that, despite the high number of prosecutions, only a few cases reached the court stage indicates that the police and prosecution were able to identify the criminal offences among the significant number of cases.⁴⁹

In some known cases, there was no doubt that the offence had been committed. For example, the conviction was handed down in the case of a man who, using a headline from a news agency, falsely claimed that a hospital doctor said that the “epicenter” of the coronavirus was in a nearby town and that no one should go on the streets. The defendant used his telephone to pass this message on to a friend, who in good faith forwarded it to his contacts, and the rumor spread very quickly on the Internet. According to the falsified article, a severely infected man from the village came to the hospital, so anyone around him was likely to be infected. The mayor was then forced to state denial to calm the public mood.⁵⁰ The court sentenced the man, who admitted the crime and expressed regret, to a suspended sentence of one year imprisonment.⁵¹ Similarly, in the case decided by the Ózd District Court, there was no doubt about the culpability of the defendant, who uploaded a video on YouTube with following statements: “there is no such thing as a coronavirus [...] everyone should go out into the streets, there is no disease, there is no coronavirus [...] everyone should go out into the streets, don’t be afraid of the coronavirus, because there is no such thing, there is no such disease.” The accused, through a video (viewed by 1,500 people), urged citizens to violate the current epidemic rules, and his actions were therefore capable of hindering or preventing the effectiveness of the protection; as a result, the district court imposed a suspended prison sentence of 1 year 10 months.⁵²

⁴⁷ “Az ügyészség szerepe a korábbiaknál is hangsúlyosabbá vált” [Press release]. The Prosecution Service of Hungary, 15 November 2021. <https://ugyeszseg.hu/az-ugyeszseg-szerepe-a-korabbiaknal-is-hangsulyosabba-valt/>

⁴⁸ *A rendőrség tudomásul veszi az ügyészség álláspontját* [Press release]. The Hungarian National Police, 13 May 2017. <https://www.police.hu/hu/hirek-es-informaciok/legfrissebb-hireink/kozrendvedelem/kozlemeny-36>

⁴⁹ DORANG (2023): 291.

⁵⁰ Jogászvilág: Elítéltek egy férfit álhírtérjesztés miatt. *Jogaszvilag.hu*, 2 September 2020. <https://jogaszvilag.hu/napi/eliteltek-egy-ferfit-alhinterjesztes-miatt/>

⁵¹ GUBIS (2021): 63.

⁵² *Miskolci Törvényszék – Büntetővégzés a rémhírtérjesztéssel vádolt férfi ügyében* [Press release]. Regional Court of Miskolc, 27 April 2021. <https://birosag.hu/aktualis-kozlemenyek/miskolci-torvenyszek-buntetovegzes-remhinterjesztessel-vadolt-ferfi-ugyeben>

Resulting the amendment of the criminal code, a number of cases were reported in the press, in particular those in which criminal proceedings were opened but eventually closed for lack of prosecution. It is worth highlighting two cases that have been widely covered by the media, both of where the statements in question were strongly critical of government decisions and the head of government, but did not result in guilty verdicts of scaremongering.

The case of the man from Szerencs who, in a Facebook post, called the Prime Minister a cruel tyrant and dictator and claimed that the “country’s leadership has deliberately timed the lifting of curfew to coincide with the peak of the coronavirus”, which will lead to mass outbreaks. In the post, the man also suggested that the prime minister is easing curfews to offset the strain on the budget by the deaths of large numbers of pensioners and that he is sending thousands to their deaths to promote himself. Within a short time, 19 people commented on the published post, followed by 14 sharing it. The prosecution considered that the man did not commit an offence as his act lacked a criminal element, since to commit an offence the perpetrator must have stated or reported an untrue fact or a distortion of a true fact to a large public, which is capable of prejudicing the outcome of the protection. The Facebook post in question did not contain such a statement. According to the prosecutor’s office, the post only “analysed a hypothetical situation in the case of a later event, such as the partial withdrawal of the restrictions imposed during the epidemic on 4 May 2020, and can therefore only be classified as the assumption that excludes the possibility of a criminal offence”. Furthermore, the statement was not intended to hinder the effectiveness of the protection, but rather to argue against the relief, drawing attention to the serious nature of the virus. In their official statement, the police stressed that they take note of the prosecutor’s position and decisions in specific cases in all cases and will continue to develop their practice in light of these.

A very similar case was the one in Gyula, in which the man accused spoke out against the freeing up of hospital beds, claiming that this meant “sending dying patients [...] home”. The Facebook post in the latter case was published on the day when a demonstration was held in Gyula on the subject of hospital reorganization and government communication on epidemics. The accused shared a post in a local Facebook group, in which he said – referring to the hospital evacuations ordered due to the pandemic – that “1,170 beds were also evacuated in Gyula”. The police opened an investigation because the post presumably obstructed the effectiveness of the pandemic control, but the prosecutor’s office closed the case for lack of criminality.⁵³ The accused in the case in Gyula exercised his legal right to appeal against the allegations, and his case was finally decided by the highest judicial forum, the Curia.⁵⁴

In another case – according to a press release from the Chief Prosecution Office of the Capital – in December 2020, two individuals filed a complaint with the Hungarian

⁵³ *A rémhírterjesztés miatt folytatott nyomozást a Szerencsi Rendőrkapitányság megszüntette* [Press release]. Borsod-Abaúj-Zemplén County Police Headquarters, 14 May 2020. https://www.police.hu/hu/hirek-es-informaciok/legfrissebb-hireink/bunugyek/megszuntetve?fbclid=IwAR3iuefpv-rXSPGBgH_UIn6N5FxFtO5PKNbZaED50G6JJSe388kKqtbUkE

⁵⁴ According to the decision published under No.BH2023.11.280, the arrest of a suspect for questioning by or on the order of the investigating authority in the course of criminal proceedings in respect of the offence under investigation is not an administrative act, but an investigative act, and therefore the Police Act cannot be invoked as its sole legal basis.

Medical Chamber, which was forwarded to the Budapest Police Headquarters. A businessman with a medical degree denied the existence of the coronavirus on social media and a website and spread false information concerning the virus.⁵⁵ After the investigation, the prosecution brought charges against the man. According to the indictment, at noon on 25 November 2020, the man published an article entitled “Freedom of Choice” on his website, which was also posted on other social media sites named after the accused. In the article, the man called the shopping time band for pensioners a pointless measure, described the government’s actions as viral propaganda and called the whole protection an “artificially fueled hysteria”, also stating that mandatory masks are harmful to health and do not fulfil their role as a protective device. According to the prosecution, the defendant’s deliberately untrue and unreliable statements were capable of misleading the public and creating mistrust towards the protection coordinators, as he described their activities as “participation in a virus business”. The accused achieved his goal, as his post on his website received hundreds of comments on his social media page and hundreds of shares.⁵⁶ The proceedings continued in the Nyíregyháza District Court, which on May 30, 2022 sentenced the defendant to a fine of HUF 100,000 (approximately 240 EUR) for the crime of scaremongering. According to the court, the man had distorted the facts and his behavior was likely to hinder the purpose and effectiveness of the protection measures, and since the writing appeared on the internet – on the defendant’s website – it was considered to have been committed in front of a large audience. Both the prosecution and the protection appealed.⁵⁷ The Court of Appeal in Nyíregyháza changed the previous verdict of the District Court in that it reduced the sentence to a reprimand, but found him guilty of the crime.⁵⁸ In the reasons for its judgment, it emphasized that the defendant, as a pharmacist with a university education, was aware that surgical masks had already been used to prevent the spread of various infections and that they could therefore be suitable for reducing and preventing COVID infections spread by droplet infection. For this reason, his claim that the use of masks makes no difference in preventing the spread of infection goes beyond being critical of government action. The court stressed that, in the context of nationwide protective measures, freedom of expression cannot extend to the allegation or reporting of knowingly untrue facts. The court found that, although the untrue facts alleged by the accused were not capable of preventing the effectiveness of the protection, they were

⁵⁵ *Feljelentések a koronavírust érintő rémhírtérjesztésre hivatkozva* [Press release]. Chief Prosecution Office of the Capital, 7 December 2020. <https://ugyeszseg.hu/feljelentések-a-koronavirust-erinto-remhírtérjesztésre-hivatkozva-a-fovarosi-fougyeszseg-sajtokozlemenye/>

⁵⁶ *Vádemelés rémhírtérjesztés miatt* [Press release]. Szabolcs-Szatmár-Bereg County Police Headquarters, 21 June 2021. <https://ugyeszseg.hu/vademeles-remhírtérjesztés-miatt-a-szabolcs-szatmar-bereg-megyei-fougyeszseg-sajtokozlemenye/>

⁵⁷ *Nyíregyházi Törvényszék – Rémhírtérjesztés büntette miatt ítélte el a bíróság a nyíregyházi férfit* [Press release]. Court of Nyíregyháza, 31 May 2022. <https://birosag.hu/aktualis-kozlemenyek/nyiregyhazi-torvenyszek-remhírtérjesztés-buntette-miatt-ítelte-el-birosag>

⁵⁸ According to Section 64 of the Criminal Act, a person shall be reprimanded if, at the time of adjudication, his act is no longer dangerous or the degree of danger to society is so insignificant that the application of even the lowest penalty or any other measure available under the Act, apart from confiscation, forfeiture of assets and rendering electronic data permanently inaccessible, is unnecessary. Through applying reprimand, the court or the prosecution service expresses its disapproval of the unlawful act and warns the perpetrator to refrain from committing any criminal offence in the future.

capable of hindering it, and that the accused had therefore committed the crime of scaremongering.⁵⁹ The defendant's education, which is linked to his knowledge of the facts and his intent to commit the crime, was therefore of great importance in the judging of this case. In many intentional offences, law enforcement officers examine what the defendant's knowledge is, whether by education or other knowledge and what additional knowledge they possess.⁶⁰

However, in addition to the legislator, the law enforcement authorities (police, prosecution and courts) also have a key role in the interpretation of the offence, in the course of which the standards established by the relevant Constitutional Court case law cannot be disregarded. Zsanett Dorang, agreeing with András Koltay, stresses the importance of case law: it must be made clear that public debates and critical opinions fall outside the scope of the regulation of scaremongering. It is necessary to avoid using the strictures of criminal law to discourage public debate, as the Constitutional Court emphasized when it stated that the new criminal definition does not cover the free exchange of ideas, which is essential for the development of democratic public debate.⁶¹

3. Concluding thoughts

As shown above, restricting freedom of expression in the interests of public order may be constitutional, but how it is done depends on the significance of the value protected. It can be stated that the restriction of fundamental rights to protect the public interest (namely public peace) is not unlimited. These limits are constitutional standards by which the requirements of necessity and proportionality shall be met. The two common characteristics of crimes against public peace do contribute to the danger of the act to society: the public nature of the offence and the intentionality.

Examining the criminal law definition of fake news and knowing the situational elements makes it possible to answer whether these restrictions ultimately pass constitutional tests. This is closely related to the issue of social dangerousness and legal objects discussed at the beginning of this study. The conclusion on scaremongering is that (1) it does not include opinions or (even harsh) criticism; (2) accurate, objective information is not considered a criminal offence even if it causes a disturbance in public peace; (3) the perpetrator must know that he is stating or publishing something that is untrue, finally (4) the untrue statement must be made in the context of a collective danger.⁶² Based on these conditions, the Constitutional Court ruled that the second new paragraph is a constitutional restriction on freedom of expression.

⁵⁹ SZON: Megrovásra enyhítette a törvényszék a rémhírterjesztés bűntette miatt elítélt férfi büntetését. *SZON.hu*, 11 October 2022. <https://www.szon.hu/helyi-kek-hirek/2022/10/megrovásra-enyhitette-a-torvenyszek-a-remhirterjesztes-buntette-miatt-elitelt-ferfi-buntetese>

⁶⁰ DORANG (2023): 289.

⁶¹ KOLTAY (2020a): 338.

⁶² A "collective danger" (in simple terms) is an objective situation where an unidentifiable large number of persons or things of significant value may be endangered (Curia decision No. BH1988.304).

When communication becomes punishable by criminal law, it causes a level of social imbalance that requires the harshest response. As *Schafer* wrote, a criminal code aims to safeguard society against activities socially dangerous to itself – the element of social dangerousness permeates the concept of crime, “[h]erein lies that peculiarity which distinguishes Hungarian criminal law on the one hand from the concept and structure of criminal law of the Western World on the other.”⁶³ It is also worth examining alternative means of sanctioning communications published in the online space other than criminal law and assessing what other – possibly more appropriate – legal means are available, including those in the field of civil or media law, which are also anchored to the constitutional foundations and considerations elaborated in this paper.

References

- ÁDÁNY et al. (2015) = Tamás Vince ÁDÁNY – Nóra BALOGH-BÉKÉSI – Zsolt BALOGH – Barnabás HAJAS: Rights and freedoms. In: András PATYI – Balázs SCHANDA – András Zs. VARGA (Eds.): *The basic (fundamental) law of Hungary. A commentary of the new Hungarian constitution*. Dublin[–Budapest], Clarus Press – National University of Public Service, 2015. 79–122.
- AMBRUS (2020) = AMBRUS István: A koronavírus-járvány és a büntetőjog. *MTA Law Working Papers*, 2020/5.
- BELOVICS (2015) = BELOVICS Ervin: XXXII. Fejezet. A köznyugalom elleni bűncselekmények. In: BELOVICS Ervin – MOLNÁR Gábor Miklós – SINKU Pál – BUSCH Béla (szerk.): *Büntetőjog 2. Különös rész* [4th ed.]. Budapest, HVG-ORAC, 2015. 561–583.
- BELOVICS (2020) = Ervin BELOVICS: Requirements for criminal law. *Journal of Eastern-European Criminal Law*, Vol. 7., No. 1. (2020) 136–146.
- CASSESE–COMBA–CARIA (2022) = Sabino CASSESE – Mario COMBA – Riccardo DE CARIA: *ELI report on freedom of expression as a common constitutional tradition in Europe*. Vienna, European Law Institute, 2022.
- DÁVID (2021) = Ferenc DÁVID: Scaremongering in the Hungarian criminal law. *Journal of Eastern-European Criminal Law*, Vol. 8., No. 2. (2021) 100–106.
- DOMOKOS (2020) = DOMOKOS Andrea: Egyes, a járványhoz kapcsolódó büntetőjogi szabályok különleges jogrend idején. *Glossa Iuridica*, 2020, Különszám. 73–83.
- DORANG (2023) = DORANG Zsanett: A koronavírusstagadók kriminalizálása, elemzés a rémhírtérjesztésről a vírus után. *Glossa Iuridica*, 2023/5–6. 269–292.
- GUBIS (2021) = Philip GUBIS: A koronavírus-járvány miatt aktuális egyes bűncselekményi tényállások dogmatikai áttekintése. In: AMBRUS, István (szerk.): *COVID–19 és büntetőjog. Az emberi egészség, a köznyugalom és más jogtárgyak védelme járvány idején*. Budapest, ELTE Eötvös Kiadó, 2021. 43–64.

⁶³ SCHAFER (1959): 167.

- HERKE (2024) = HERKE Csongor: Digitális csapdák. A clickbait és a bűnügyi tudományok kereszteződése. *JURA*, 2024/1. 26–47.
- KARSAI–SZOMORA (2015): Krisztina KARSAI – Zsolt SZOMORA: Hungary. In: Vanessa FRANSSEN – Mona GIACOMETTI – Stanislaw TOSZA (Eds.): *International encyclopaedia of laws. Criminal law*. Alphen aan den Rijn, Kluwer Law International, 2015.
- KLEIN (2020) = KLEIN Tamás: A veszélyhelyzeti jogalkotás alkotmányos alapjai, különös tekintettel a véleménynyilvánítás különleges jogrendi korlátozhatóságára. *Glossa Iuridica*, 2020, Különszám. 211–234.
- KOLTAY (2019) = KOLTAY András: Általános bevezető. In: KOLTAY András (szerk.): *Magyar és európai médiajog* [3rd rev'd ed.], Budapest, Wolters Kluwer, 2019. 239–240.
- KOLTAY (2020a) = KOLTAY András: A rémhírterjesztés büntetethezességének alkotmányosságáról. In *Medias Res*, 2020/2. 322–338. <https://doi.org/10.59851/imr.9.2.8>
- KOLTAY (2020b) = András KOLTAY: *The punishment of scaremongering in the Hungarian legal system. Freedom of speech in the times of the COVID-19 pandemic*. (November 23, 2020). <http://dx.doi.org/10.2139/ssrn.3735867>
- MEZŐLAKI (2013) = MEZŐLAKI Erik: A köznyugalom elleni bűncselekmények. In: KARSAI, Krisztina (szerk.): *Kommentár a Büntető Törvénykönyvhöz. Kommentár a Büntető Törvénykönyvről szóló 2012. évi C. törvényhez*. Budapest, CompLex, 2013. 691–711.
- NAGY (2010) = NAGY Ferenc: *A magyar büntetőjog általános része* [2nd rev'd ed.]. Budapest, HVG-ORAC, 2010.
- SCHAFER (1959) = István SCHAFER: *Some basic principles of Hungarian criminal law. Modern Law Review*, Vol. 22., No. 2. (1959) 164–174. <https://doi.org/10.1111/j.1468-2230.1959.tb02167.x>
- SZEMÁN (2022) = SZEMÁN Ádám: Társadalomra veszélyesség megjelenése, alakulása és fogadtatása a magyar büntetőtörvényekben. *Debreceni Jogi Műhely*, 2022/1–2. 105–120. <https://doi.org/10.24169/djm/2022/1-2/6>
- SZOMORA (2013) = SZOMORA Zsolt: A büntetőjogi felelősség. In: KARSAI, Krisztina (szerk.): *Kommentár a Büntető Törvénykönyvhöz. Kommentár a Büntető Törvénykönyvről szóló 2012. évi C. törvényhez*. Budapest, CompLex, 2013. 39–71.
- TOKAJI (1984) = TOKAJI Géza: *A bűncselekménytan alapjai a magyar büntetőjogban*. Budapest, Közgazdasági és Jogi Könyvkiadó, 1984.
- TÖRÖK (2013) = TÖRÖK Bernát: A gyűlöletbeszéd tilalmának médiajogi mércéi. *Jogtudományi Közlöny*, 2013/2. 59–72.
- WIENER (1997) = Imre A. WIENER: Constitution and criminal law. *Acta Juridica Hungarica*, 1997/1–2. 19–37.

Authors

Andrej Školka – PhD, The School of Communication and Media, Bratislava, Slovakia.

Gergely Ferenc Lendvai – Pázmány Péter Catholic University, Faculty of Law, PhD candidate.

János Tamás Papp – PhD, Associate Professor, Pázmány Péter Catholic University, Faculty of Law, Research expert, National Media and Infocommunications Authority.

Levente Nyakas – PhD, Associate Professor, Budapest Metropolitan University, Head of Media Council's Institute for Media Studies

Mónika Andok – Pázmány Péter Catholic University, Department of Communication Studies, Chair of Department, Associate Professor.

Péter Bányász – Associate Professor, Ludovika University of Public Service Faculty of Public Governance and International Studies.

Péter Nemere – Research and legal expert, National Media and Infocommunications Authority, PhD candidate, Eötvös Loránd University.

Ronan Ó Fathaigh – Assistant Professor, Institute for Information Law, Faculty of Law, University of Amsterdam.

Zsuzsanna Monori – Member of staff at the Media Council's Institute for Media Studies and Research Lecturer at the Public Law Centre of Mathias Corvinus Collegium.

Institute for Media Studies, Budapest
Responsible for publication: Levente Nyakas
Layout: Ákos Varga
Published in 300 copies, in a length of 16,5 B5 sheets
Médiatudományi Könyvtár: ISSN 2063-5222
Médiatudományi Könyvtár 63.: ISBN 978-615-5302-60-2