

Information Security Policy

Conditions, Threats and Implementation
in the International Environment



EDITED BY
PIOTR BAJOR

INFORMATION SECURITY POLICY

INFORMATION SECURITY POLICY

**Conditions, Threats and Implementation
in the International Environment**

**edited by
PIOTR BAJOR**




Kraków 2022

© Copyright by individual authors, 2022

Piotr Bajor

Jagiellonian University in Kraków, Poland

 <https://orcid.org/0000-0003-2569-2552>

 piotr.bajor@uj.edu.pl

Review: Marek Delong

Language editor:

Cover design: Marta Jaszczuk

ISBN 978-83-8138-826-9 (print)

ISBN 978-83-8138-827-6 (PDF)

<https://doi.org/10.12797/9788381388276>

The publication is available under the Creative Commons Attribution 4.0 International license. Some rights reserved by the authors. The publication was created as part of the “Public Diplomacy 2022” competition. Any use of the work is allowed, provided that the above information is preserved, including information about the license used and about the rights holders.

The opinions expressed in this publication are those of the authors and do not reflect the views of the official positions of the Ministry of Foreign Affairs of the Republic of Poland.



Ministry
of Foreign Affairs
Republic of Poland

Public task financed by the Ministry of Foreign Affairs of
the Republic of Poland within the grant competition
“Public Diplomacy 2022”.



The project was implemented by the Foundation Center for Research
on the Contemporary Security Environment, in cooperation with
a research team from the Jagiellonian University in Kraków.

KSIĘGARNIA AKADEMICKA PUBLISHING

ul. św. Anny 6, 31-008 Kraków

tel.: 12 421-13-87; 12 431-27-43

e-mail: publishing@akademicka.pl

<https://akademicka.pl>

Table of contents

Introduction	7
MAGDALENA DANEK	
Social Media as a Recipient and Creator of Political Actions in the Context of the Security Crisis.....	9
AGNIESZKA NITSZKE	
The European Union versus Russian Disinformation	35
MICHAŁ MAREK	
Information Security and Mechanisms Used by the Russian Federation to Shape Polish Public Opinion.....	53
MONIKA ŚLUFIŃSKA	
The Russia-Ukraine War. Two Strategies of Communication?	67
ADRIAN TYSZKIEWICZ	
The Russian Narrative Construct towards Ukraine.....	83
PIOTR BAJOR	
Information Security Policy of Ukraine – Assumptions and Effectiveness	99
Index of names	123

MAGDALENA DANEK 

Jagiellonian University in Kraków

Social Media as a Recipient and Creator of Political Actions in the Context of the Security Crisis

ABSTRACT: Social media is not only an increasingly popular communication channel or business tool, but also one of the arsenals of information warfare. The next phase of Russia's war against Ukraine, launched on February 24, 2022, showed once again that the content disseminated through it is used not only to provide actual information or to improve the organisation of assistance to refugees, but also to spread disinformation and propaganda. The aim of the article is to analyse the current status of social media platforms as tools of influence and power – particularly during the war in Ukraine – as well as activities aimed at combating disinformation, especially in the context of the activities of the EU, selected state actors and the owners of these platforms themselves (in this aspect, the analysis will include the activities of Meta and Twitter). The research hypothesis is based on the assumption that social media is, in the scope of the present issue, not only the recipient of political decisions made by legitimised actors, but – by virtue of their power over the flow of a significant amount of information – it become an important actor in these activities in terms of influencing political processes and decision-making centres (e.g., by

arbitrarily deciding on the visibility of hate speech content in situations of armed conflict).

KEYWORDS: social media, disinformation, Ukraine, power, influence, hybrid war

Introduction

With the emergence and development of the internet and the applications based on it, including social media, expectations could be seen in both public and academic discourse about the opportunity to make them tools to strengthen democratic processes and political engagement of citizens. Thus, television, which Robert Putnam saw as one of the main causes of generating civic passivity and the erosion of social capital,¹ was to give way to an egalitarian and interactive space based on free access to information and free expression of opinion. Jan van Dijk concludes that hopes for the impact of ICT on politics were linked primarily to increasing the acquisition and exchange of information between government and administrative representatives and citizens, enhancing public debate, deliberation, the formation of communities and citizen participation in decisions of public importance.² These approaches were based on the belief that increasing citizens' access to information benefits both society itself and democratic procedures.

The massive proliferation of false content on social media, including the notorious disinformation campaigns accompanying events of particular significance – such as the 2016 US presidential election, the campaign for Britain's exit from the European Union (EU), the COVID-19 pandemic or the next round of the war in Ukraine launched on February 24, 2022 – are shifting the focus towards seeing the internet and the applications that function within it as a space not primarily for debate, but information warfare.³

¹ R. Putnam, *Samotna gra w kręgle. Upadek i odrodzenie wspólnot lokalnych w Stanach Zjednoczonych*, transl. P. Sadura, S. Szymański, Wydawnictwo Akademickie i Profesjonalne, Warszawa 2008, p. 384.

² J. van Dijk, *Społeczne aspekty nowych mediów*, transl. J. Konieczny, Wydawnictwo Naukowe PWN, Warszawa 2010, p. 150.

³ A. Guess, B. Nyhan, J. Reifler, *Selective Exposure to Misinformation: Evidence from the Consumption of Fake News during the 2016 US Presidential Campaign*, 9.01.2018, [on-line:] <https://about.fb.com/wp-content/uploads/2018/01/fake-news-2016.pdf>, 20 November 2022; M.T. Bastos, D. Mercea, "The Brexit Botnet and User-generated Hyperpartisan News", *Social Science Computer Review*, vol. 37, no. 1 (2019), pp. 38–54; Y.M. Rocha et al., "The Impact of Fake News on Social Media and Its Influence on Health during the COVID-19 Pandemic: A Systematic Review", *Journal of Public Health* (2021), pp. 1–10.

The information space has, since ancient times, been seen as a vital pillar of security and, at the same time, a tool for confrontational action. This is expressed in the words of Sun Tzu, who – in his treatises – indicated that *war is about being misled*.⁴ Coherent management of the information space, especially in its digital dimension, is one of the key elements of state security.

The approach to war as not only a kinetic clash, but a whole range of diverse actions – including a special role for the information sphere – has been particularly popularised in the context of the notion of hybrid warfare. The clearest emanation of this phenomenon was Russia's actions towards Ukraine with the annexation of Crimea and the start of fighting in the Donbas region in 2014. The Russian Federation's point of view on the modern battlefield can be reconstructed from an article by the Chief of the General Staff of the Russian Armed Forces (Valery Gerasimov), where he states that nowadays, the fundamental principles of war have changed, and the role of non-military means of achieving political and strategic goals has significantly increased – often exceeding the power and effectiveness of kinetic weapons.⁵

In this sense, hybrid warfare, also referred to as *a war of controlled chaos*, encompasses the entire range of actions implemented to destabilise the economic and political situation, disintegrate and limit sovereignty, and consequently change political power to that controlled by the aggressor.⁶ Although the concept of hybrid warfare does not have a clearly defined scope of meaning – and, thus, faces accusations of blurring the boundaries between times of war and times of peace, lowering preparedness for an appropriate response – it is noted that it points to key current and future security and defence challenges.⁷

Russia's ongoing war against Ukraine, which has been continuing since February 24, 2022, despite being – in its significant dimension – an example of a kinetic type of clash, is also marked by a meaningful potential for other acts with the hallmarks of a hybrid impact. It should be emphasized that, they are targeted not only at Ukraine,

⁴ Sun Tzu, *Sztuka wojny*, transl. J. Zawadzki, Hachette, Warszawa 2009, p. 31.

⁵ V. Gerasimov, "The Value of Science Is in the Foresight New Challenges Demand Rethinking the Forms and Methods of Carrying out Combat Operations", *Military Review* (2016), [on-line:] https://www.armyupress.army.mil/portals/7/military-review/archives/english/militaryreview_20160228_art008.pdf, 22 November 2022.

⁶ O. Wasiuta, "Geneza pojęcia i zmiany podejścia do wojny hybrydowej w zachodnim dyskursie politycznym i wojskowym", *Przegląd Geopolityczny*, no. 17 (2016), p. 28.

⁷ A. Bilal, "Hybrid Warfare – New Threats, Complexity, and 'Trust' as the Antidote", *NATO Review* 2021, [on-line:] <https://www.nato.int/docu/review/articles/2021/11/30/hybrid-warfare-new-threats-complexity-and-trust-as-the-antidote/index.htm> (20.11.2022).

but also at Western countries. They manifest themselves, among other means, in large-scale propaganda actions or economic blackmail related to access to energy resources.

The aim of the article is to analyse the current status of social media platforms as tools of influence and the space of power making, in particular during the war in Ukraine, as well as activities aimed at combating disinformation – especially in the context of the activity of the EU, selected state actors and the owners of these platforms themselves (in this aspect, the analysis will include Meta and Twitter activities). The theoretical framework for the undertaken research will be the concept of the network society, and power understood as an influence on the management of communication processes by Manuel Castells. The research hypothesis is based on the assumption that in the scope of this issue, social media is not only the recipient of political decisions made by legitimate actors, but thanks to their control over the flow of a significant amount of information, they become an important actor of these activities in terms of influencing political processes.

In the course of the analysis carried out, three key reference levels were distinguished. The first relates to using social media as a new and increasingly crucial war arsenal in information warfare. The second one concerns regulatory action and the pressure exerted by political actors such as states and international organisations on social media platforms, which resultantly become the object of political action – especially in the context of the fight against disinformation. The last, but extremely important, dimension covers actions taken in relation to the conflict by the managers of social platforms, making them important actors of decision and political impact.

Information management in cyberspace as an emanation of power

Manuel Castells, analysing the contemporary transformations of social, economic and political structures in the era of dynamic ICT development, introduces the concept of a network society (i.e., social structure) whose main features are the presence of digital network communication technologies, as well as the reproduction and institutionalisation of the connections created thanks to them through society itself. In this way, according to the researcher, a *new social morphology* is being created.⁸ One of the essential dimensions of this emerging social structure are the power relationships

⁸ M. Castells, *Spółeczeństwo sieci*, crowd. M. Marody et al., transl. M. Marody, Wydawnictwo Naukowe PWN, Warszawa 2011, p. 491.

that are *exercised primarily by constructing meanings in people's minds*.⁹ This approach reflects the concept of power as invisible, circulating, devoid of a specific location and, at the same time, omnipresent (as presented by Michel Foucault in his works).¹⁰ For Castells, this process takes place via multimodal networks, where the type of communication appropriate for the era of new media is implemented. He describes this as *mass self-communication*.¹¹ Its essence is a potentially opposing combination of the global possibility of spreading messages with individualised and independent creation and reception based on personal selection.

Power, meaning a relational connection based on exerting influence, thus shifts in a network society from using physical violence to constructing meanings and media narratives. This process occurs mainly in communication networks assuming the role of new power centres. From this point of view, new media is not just neutral technical creations; access to them and the ability to use their networked architecture is one of the main sources of power. Under the concept of Castell, the nation-state coexists as one of the many sources of power and authority. At the same time, it is a node of a deeper network of power which consists of forces of capital, communication, international institutions, social movements, terrorist organisations, as well as local and regional authorities.¹² Due to the dynamic development and ubiquity of new media, the space of political competition is dominated by the media, which is becoming *the privileged realm of politics*.¹³

Therefore, in this sense, social media is not only platforms for the mass dissemination of information, but also active subjects for the exercise of power and influence (i.e., political actions), which is particularly evident during crises, including armed conflicts. Several factors can be distinguished to qualify them in this way. The first is its high and growing popularity. According to data as of January 2022, it is used by over 4.6 billion users worldwide,¹⁴ which – at the same time – means an annual growth of 10%. The purpose of using these platforms is not only to contact friends, but also to obtain information and express one's own beliefs. However, it is not only

⁹ Idem, *Władza komunikacji*, transl. J. Jedliński, P. Tomanek, Wydawnictwo Naukowe PWN, Warszawa 2013, p. 409.

¹⁰ M. Foucault, *Nadzorować i karać. Narodziny więzienia*, transl. T. Komendant, Wydawnictwo Aletheia, Warszawa 2009, p. 189.

¹¹ M. Castells, *Władza...*, p. 415.

¹² Idem, *Siła tożsamości*, transl. S. Szymański, Wydawnictwo Naukowe PWN, Warszawa 2009, pp. 323–331.

¹³ *Ibidem*, p. 339.

¹⁴ S. Kemp, "Digital 2022: Global Overview Report", *Data Reportal*, 26.01.2022 [on-line:] <https://datareportal.com/reports/digital-2022-global-overview-report> (23.09.2022).

the scale, but – most of all – the structure and convention of the operation of social media platforms based on the monetisation of users' attention that is the main aspect of their enormous impact.

In September 2021, the *Wall Street Journal* accessed Facebook's internal guidelines for content moderation, which it obtained from former employee Francis Haugen. These documents, submitted to the US Congress and referred to as *Facebook Papers*, indicated that the owners of the platform put their own profit much higher over security, which resulted in the admission to the public circulation of content that is controversial and arouses social unrest because it generates more traffic and user activity. Consequently, it brings more profit. It was pointed out that the loosening of moderation restrictions after the presidential elections in the US in 2020 allowed for a rapid increase in the popularity of radical groups. It could have stimulated the Capitol attack in early 2021.¹⁵ The obtained information also indicates that the platform's algorithm model strengthens social polarisation by promoting emoticons expressing emotions (including anger, for example) five times more than the popular 'like'.¹⁶ The lack of an appropriate number of moderators operating in different languages also means that there is a fundamental disproportion in combating disinformation content occurring in non-English-speaking countries.¹⁷

Another factor that makes social media platforms with billions of users the real subjects of power and political action is the arbitrariness and frequent lack of transparency in their decision-making. They combine, quite paradoxically, with the exclusion of social media liability for the content published by the users of their services, which is primarily guaranteed in US law, namely Section 230 *Communication Decency Act*.¹⁸

On the one hand, this means that social media platforms are not treated as media publishers, which allows them to maintain the possibility of free expression but, on the other hand, does not restrict their owners from making arbitrary decisions. Henry Kissinger stated that a laptop can make global consequences, thus referring to the

¹⁵ C. Lima, "A Whistleblower's Power: Key Takeaways from the Facebook Papers", *Washington Post*, 26.10.2021, [on-line:] https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/?itid=co_facebookunderfire_1 (21.11.2022).

¹⁶ J.B. Merrill, W. Oremus, "Five Points of Anger, One for a 'like': How Facebook's Formula Fostered Rage and Misinformation", *The Washington Post*, 26.10.2021, [on-line:] https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/?utm_campaign=wp_main&utm_medium=social&utm_source=twitter (20.11.2022).

¹⁷ M. Scott, "Facebook Did Little to Moderate Posts in the World's Most Violent Countries", *Politico*, 25.10.2021, [on-line:] <https://www.politico.com/news/2021/10/25/facebook-moderate-posts-violent-countries-517050>, 29 September 2022.

¹⁸ *The Communication Decency Act of 1996*, 47 U.S. Code § 230.

situation where users of modern mobile phones and computers have [in their hands] an unprecedented potential for gathering, aggregating and analysing information – which was beyond the reach of many intelligence agencies in the previous generation. However, greater access to information may also be used in a direction that is undesirable from the point of view of democratic processes. Media corporations who collect a lot of detailed data about users and are able to track and even influence them, which is even beyond the capabilities of many modern countries.¹⁹

One of the co-founders of the most popular platform (Facebook) and the coordinator of Barack Obama's presidential campaign online in 2008 (Chris Hughes), in a famous column for *New York Times*, stated that the influence in the hands of Mark Zuckerberg is enormous and far beyond that of other entities in the private or public sector. The head of Facebook, operating since October 2021 as part of the Meta Group, can practically independently decide on the configuration of the platform's algorithm and, thus, determine what almost 3 billion users see and how their data is processed.²⁰ Mass spreading of false content – often as a result of specialised disinformation campaigns and monetisation of negative emotions by the platforms themselves – thus has negative consequences for democracy by strengthening epistemic cynicism (based on lowering trust in institutions and authorities), polarising discussions and give the feeling that mutual debate does not make sense due to the participation of many unidentified entities, especially bots and trolls.²¹

Social media as an environment for information warfare and the creation of conflict narratives

The ongoing renewal of the war in Ukraine (since February 24, 2022) involves communication on social media platforms on an unprecedented scale. As a result, the conflict is referred to (in public discourse) as the most *viral* war or even *the first TikTok*

¹⁹ H. Kissinger, *Porządek światowy*, transl. M. Antosiewicz, Wydawnictwo Czarne, Wołowiec 2017, p. 323.

²⁰ Ch. Hughes, "It's Time to Break Up Facebook", *New York Times*, 9.05.2019, [on-line] <https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html>, 17 September 2022.

²¹ S. McKay, C. Tenove, "Disinformation as a Threat to Deliberative Democracy", *Political Research Quarterly*, vol. 74, no. 3 (2021), pp. 703–717.

war due to the high popularity of this channel in accounts of the theatre of war disseminated – particularly by civilians.²²

Posting war content on social media has a variety of goals. In the case of Ukraine, these platforms are often a tool for a quick transmission of important information from the point of view of the civilian population both in Ukraine and abroad (e.g., alerts about missile attacks or information on points and the refugee aid organisation system). Above all, however, these channels allow internet users to report their wartime experiences, and viewers from almost all over the world witness them on a regular basis. In this way, expression in social media helps to maintain determination and unity in resisting, which is often reinforced with a message in the form of memes and a content based on situational comedic effect that reveal the weaknesses of the Russian army and the strength and successes of Ukraine (e.g., videos and memes about the ‘theft’ of Russian tanks by local farmers, the famous war song performed by the Ukrainian military called *Bayraktar*). The materials published on these channels are also used to build and maintain the will to help Ukraine, especially by Western states and societies. This is the aim of showing the crimes committed by Russian soldiers and shaping the narrative of the war as an existential threat to the whole of Europe, not just a local conflict.²³

The very person of President Zelensky, who, in his communication on Twitter (he has 6.5 million followers), is also important for the implementation of the information policy on the part of Ukraine²⁴ because his account contains both official statements as well as non-professional video recordings that create a sense of closeness and ‘naturalness’ with the audience. The message coming from this communication is largely directed at Western countries and is aimed at influencing the provision of further assistance to Ukraine, primarily in the area of the supply of military equipment.

During the war in Ukraine, social media is also becoming a tool for disseminating opinions and sharing knowledge by analysts from the white intelligence community, the so-called “OSINT” (*open-source intelligence*) such as Belling Cat, Rochan Consulting or those who track disinformation in major Russian news channels (e.g., journalist Julia Davis, who was sanctioned by the Russian authorities for her activities).²⁵

²² M. Połowaniuk, “Rosyjska inwazja to pierwsza ‘wojna TikTokowa’”. Ukraina pokazuje, jak wygrać bitwę w sieci”, *Spider’s Web*, 28.02.2022, [on-line:] <https://spidersweb.pl/2022/02/jak-wygrac-wojne-w-internecie.html>.

²³ D. Ciuriak, *The Role of Social Media in Russia’s War on Ukraine*, 5.05.2022 [on-line:] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4078863 (28.09.2022).

²⁴ Status on 28.09.2022, [on-line:] <https://twitter.com/ZelenskyyUa>.

²⁵ Julia Davis’ Twitter account, where she publishes the results of her analysis of the narratives of major Russian news channels (entitled *Russia Media Monitor*) is followed by 353,000 users (as of September 28, 2022).

Social media platforms are also a mobilisation environment for grassroots support for Ukraine in respect of the humanitarian, military and informational dimensions. This is the dimension of the group created by internet users called NAFO (North Atlantic Fellas Organization), symbolised by the dog Shiba Inu. Its aim is to expose Russian propaganda and raise funds to support the Ukrainian army.



Figure 1. A tweet from the Ukrainian Ministry of Defence profile expressing gratitude to the NAFO movement for its activities. Source: Profile of the Ministry of Defence of Ukraine on Twitter, [online:] <https://twitter.com/defenceu/status/1563851548643426304>, (28.09.2022)

As for information warfare, social media became – especially from Russia’s point of view – an ideal tool for the implementation of disinformation campaigns, large-scale actions based on inauthentic activity and the dissemination of propaganda messages from the Russian authorities (for example, through the famous and aggressive in its message posts of the Vice-President of the Security Council of the Russian Federation, Dmitry Medvedev, on Telegram). In Russia’s confrontational actions, social media is another channel through which the fog of war can be generated, to spread panic in Western societies or reinforce uncertainty about who is guilty of committing war crimes (such as those in the maternity hospital in Mariupol or in Bucha). Reports from social media platforms regarding identified misinformative profiles and content often point

to accounts linked to Russia. Only in September 2022, Meta Concern (the owner of Facebook) announced that it had detected two extensive disinformation networks related to China and Russia. In the case of Russia, disinformation activities launched in May 2022 were based on the creation of around 60 websites imitating well-known media portals (e.g., *The Guardian*, *Bild* or *Spiegel*). The propaganda content published there was aimed at criticising Ukraine and strengthening antagonism towards refugees, building a positive Russian image and emphasising that the sanctions imposed on it are counterproductive. The content was then disseminated both organically and through purchased advertising on various social media platforms, including Facebook (with 1,633 accounts involved), Instagram, Twitter, Telegram and YouTube. This campaign, the largest since the start of the war in Ukraine, targeted audiences in Germany, Italy, the UK, France, Latvia and Ukraine. Significantly, this initially diverse audience has, over time, been limited to influencing mainly German audiences – indicating them as a key target of Kremlin propaganda.

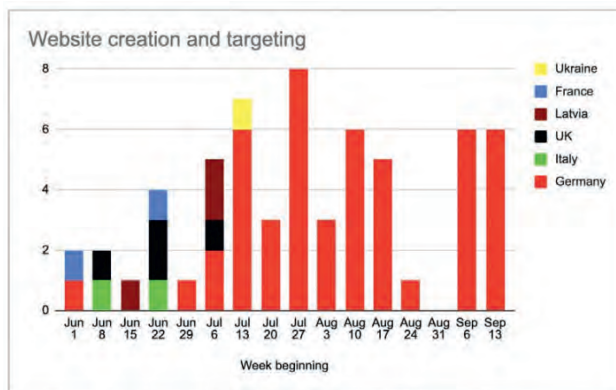


Chart 1. The number of popular media imitation sites created and their target audience. Source: B. Nimmo, M. Torrey, "Taking Down Coordinated Inauthentic Behavior from Russia and China", Meta Detailed Report, [on-line] <https://about.fb.com/news/2022/09/removing-coordinated-inauthentic-behavior-from-china-and-russia/> (28.09.2022)

In its August 2022 report, Twitter also shared information about the identification and deletion of 38 accounts linked to pro-Kremlin media outlets. Before disseminating the Russian narrative about the war in Ukraine, their activities focused on propaganda related to the COVID-19 pandemic.²⁶

²⁶ The analysis was based on data shared by Twitter with The Stanford Internet Observatory as part of the Moderation Research Consortium. See: *A Front for Influence: An Analysis of a Pro-Kremlin*



Figure 2. A screenshot from the authentic website of The Guardian – theguardian[.]com (top) versus from a website created by the Russian disinformation network to imitate this medium – theguardian[.]co[.]com (bottom) containing an article accusing Ukraine of staging the Bucha crime. Source: B. Nimmo, M. Torrey, “Taking Down Coordinated Inauthentic Behaviour from Russia and China”, Meta Detailed Report, [on-line:] <https://about.fb.com/news/2022/09/removing-coordinated-inauthentic-behavior-from-china-and-russia/> (28.09.2022)

In turn, an example of a widespread disinformation campaign, mainly in the Polish-language Twitter space, is the sudden increase in popularity of the hashtag #StopUkrainizacjiPolski, attached at the end of August 2022 to content expressing opposition to aid for refugees and generating negative attitudes towards them. The hashtag, which previously appeared on social media but never before with such popularity, was especially popular with the activists of the right-wing political party Confederation of the Polish Crown. Winning the position of hashtag leader among Polish Twitter users at the end of August this year, however, was due – as indicated by DFRLab analysis – to the artificial manipulation of traffic generated by a group of hyperactive accounts (according to this analysis, ten accounts accounted for as much as 16% of all, approximately 46,000 hashtag-related activities undertaken between 24–27 August 2022).²⁷ An analysis of the further popularity of the hashtag made by the author of the article shows that on September 10–17, activity related to it decreased

Network Promoting Narratives on COVID-19 and Ukraine [on-line:] <https://cyber.fsi.stanford.edu/io/news/sio-aug-22-takedowns-ua> (28.09.2022).

²⁷ G. Gigitashvili, *Twitter Campaign Pushes Anti-Ukraine Hashtag into Poland's Trending List*, 8.2022 [on-line:] <https://medium.com/dfrlab/twitter-campaign-pushes-anti-ukraine-hashtag-into-polands-trending-list-90ccc9474a60> (22.11.2022).

(a total of 5,857 activities were recorded, including 743 tweets, 3,388 retweets and 1,726 replies). However, it was noted that more than 30% of the tweets came from two profiles whose activity clearly increased during the peak of the hashtag's popularity. At the same time, it was shown that both of these accounts were publishing about 30 tweets a day, which suggests suspicions of inauthentic or propaganda-oriented activity.²⁸

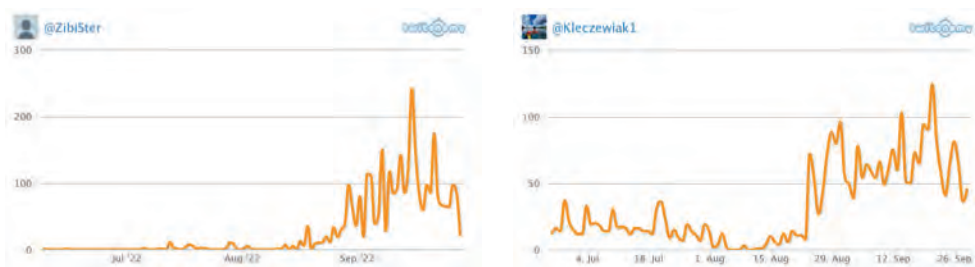


Chart 2. Three-month activity (tweets, retweets and replies) of two accounts publishing the most tweets with the hashtag #StopUkrainalizacjiPolski on 10-17.09.2022. Source: own analysis using the Twitonomy tool

Social media as an object of political activities, especially in the fight against disinformation

Numerous disinformation campaigns, as well as the Cambridge Analytica case (when data from 87 million Facebook user accounts were used for political micro-targeting) were important triggers for various actors to combat this practice, including increasing the accountability of social media platforms for their content.²⁹ They included both legislative solutions (e.g., laws adopted in France, Germany or Austria) as well as activities related to the early detection of disinformation campaigns and increasing social resistance to their impact. The large-scale fight against *fake news* contributed to the emergence of specialised entities such as the Centre Against Terrorism and Hybrid Threats (CTHH) in the Czech Republic or the European Centre of Excellence for Countering Hybrid Threats established in April 2017 (*Hybrid CoE*), an embodiment

²⁸ Data on number of tweets was obtained from the Twitter API via the MAXQDA software. The analysis of the activity of individual profiles was carried out using the Twitonomy tool.

²⁹ F. Saurwein, C. Spencer-Smith, "Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe", *Digital Journalism*, vol. 8, no 6 (2020), pp. 820–841.

of the interaction between the EU and NATO.³⁰ Furthermore, the EU, in March 2015, had already taken steps to combat Russian disinformation, particularly in the Eastern Partnership countries, by establishing a task force within the framework of the European External Action Service (EEAS) *East Strat Com*. Its flagship project is *EUvsDisinfo*, aimed at monitoring the pro-Kremlin media and the propaganda they spread, which is currently of particular importance during the ongoing war in Ukraine and intensified disinformation activities at that time.³¹ It is also worth to mention, an important role of non-governmental organisations in countering disinformation. In this regard this role is played especially by the international community of fact-checking organisations, *International Fact-Checking Network*, which consists of around 100 entities to verify the content on online platforms.

An important aspect of the actions taken in the fight against disinformation was the attempt to find a balance between two key values, namely freedom of expression and ensuring security. While the sphere of combating disinformation by detecting and giving an early warning [to the public] of false narratives does not enter into this dilemma so much, actions taken in the legislative arena are often met with concerns about restrictions on freedom of speech. The lack of a strictly defined meaning of the terms such as disinformation or *fake news* both in the media and scientific discourse³² are often used – in particular, under authoritarian regimes to limit the freedom of citizens. This is the case, for example, in Russia, where since 2019, dissemination of false information that may harm life and health, public order and safety, as well as content that expresses disrespect for the authorities and state symbols of the Russian Federation is penalised with a fine or 15-day detention.³³

Dilemmas over the scope of the solutions introduced in this regard are also present in democratic states, which have introduced new obligations on social media platforms to combat disinformation or hate speech in their legislation. This was the case, for example, in Germany, where the regulations of the *Network Enforcement Act*

³⁰ M. Danek, "Modele walki z dezinformacją: od restrykcji do współpracy", in M. Bernaczyk, T. Gąsior, J. Misiuna, M. Serowaniec (eds), *Znaczenie nowych technologii dla jakości systemu politycznego. Ujęcie politologiczne, prawne i socjologiczne*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2020, p. 138.

³¹ The project's website contains both research and reports presenting the results of conducted monitoring, as well as a database allowing users to search for individual media content on their own. See Project website [on-line:] <https://euvsdisinfo.eu/pl/#> (28.09.2022).

³² E.C. Tandoc Jr., Z.W. Lim, R. Ling, "Defining 'Fake News': A Typology of Scholarly Definitions", *Digital Journalism*, vol. 6, no. 2 (2018), pp. 137–153.

³³ "Russia Passes Legislation Banning 'Disrespect' of Authorities and 'Fake News'", *The Moscow Times*, 7.03.2019 [on-line:] <https://www.themoscowtimes.com/2019/03/07/russia-passes-legislation-banning-disrespect-of-authorities-and-fake-news-a64742> (29.09.2022).

(NetZDG), in force since 2018, were met with criticism from the opposition.³⁴ *The Online Safety Bill*, currently under consideration in the UK, is also controversial due to the requirement for social media platforms to tackle not only illegal content, but also content that is described as legal but harmful, which has led to fears of censorship.³⁵

The war in Ukraine emphasised this dilemma even more. The extraordinary circumstances related to the security crisis created expectations to cross the existing borders in the context of increasing security and ensuring freedom of expression. There has been an expectation and even a pressure towards social media platforms on the part of Western countries – as well as Ukraine – to act not so much as a neutral actor, but as an active influence entity in the fight against Russian propaganda. Responding to these expectations, both Meta and Twitter started, from the beginning of the conflict, to specifically flag and restrict the visibility of content from Russian media. Then, following the introduction of EU-level sanctions, they blocked the accounts of two key Kremlin tubes (Sputnik and Russia Today).³⁶ In response to the actions taken by the platforms, Facebook, Instagram and Twitter were blocked in Russia by the decision of *Roskomnadzor*.³⁷

In turn, Ukrainian political actors emphasise that these platforms must give up their position of technical neutrality in favour of becoming allies in shaping the information environment of the ongoing war. In this approach, more emphasis is placed on freedom of speech, especially in terms of reporting on Russian war crimes. As President Zelenskiy noted in one of his tweets, “War is not only a military opposition on UA land. It is also a fierce battle in the informational space.”³⁸ The politician then expressed his gratitude to Meta and other social media platforms for their solidarity with Ukraine in this field. However, one can notice that content reporting on events in Ukraine is often blocked due to automated algorithms or the activity of Russian-

³⁴ M. Danek, “Komunikowanie polityczne w dobie fake newsów – walka z dezinformacją w sieci”, *Krakowskie Studia Małopolskie*, vol. 23 (2018), pp. 210–228.

³⁵ “Online Safety Bill to Return as Soon as Possible”, *BBC*, 20.09.2022, [on-line:] <https://www.bbc.com/news/technology-62908598> (29.09.2022).

³⁶ “Kalendarium – sankcje UE wobec Rosji w sprawie Ukrainy”, *Strona Rady Europejskiej i Rady Unii Europejskiej*, [on-line:] <https://www.consilium.europa.eu/pl/policies/sanctions/restrictive-measures-against-russia-over-ukraine/history-restrictive-measures-against-russia-over-ukraine/> (29.09.2022).

³⁷ “Facebook i Twitter zablokowane w Rosji”, *Wirtualne Media*, 5.03.2022, [on-line:] <https://www.wirtualnemedia.pl/artykul/rosja-blokada-facebook-twitter> (29.09.2022).

³⁸ Quoted from: Volodymyr Zelensky’s tweet of March 13, 2022, [on-line] https://twitter.com/ZelenskyyUa/status/1503046528071618562?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1503046528071618562%7Ctwgr%5E69f0efac3b90149cdb32cd47fa434406cf5c877%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fdorzechy.pl%2Fopinie%2F275323%2Fzelenski-dziekuje-mecie-wojna-to-nie-tylko-konfrontacja-militarna.html (29.09.2022).

linked accounts that report it as material that does not comply with the platforms' terms of service. This was pointed out by Mykhailo Fedorov, Deputy Prime Minister and Minister of Digital Transformation of Ukraine, who – in a July 2022 letter to Nick Clegg, CEO of Global Policy Meta Corporation, on the moderation of Ukrainian content – stressed that the truth about the Russian invasion could not resonate enough when social media policy is created *only for the time of peace*.³⁹ He also pointed out the need to provide the Ukrainian side with a so-called “market-specific slur lists” (i.e., an internal and periodically updated document by Meta to help moderators adjudicate violations of the rules of procedure in relation to the cultural context).⁴⁰ According to Fedorov, this will allow to limit the cases of excessive and incorrect imposition of restrictions on content.

Finally, it should be noted that during the ongoing armed conflict, the fight against disinformation entered the next level in the context of EU actions. It is the result of a policy that has been in place for several years in this area, and was intensified – in particular – before the European Parliament elections in 2019. The previously adopted acts, especially in the form of self-regulation of digital platforms, such as the Code of Practice on Disinformation,⁴¹ were supplemented by the Digital Service Act (DSA)⁴² and Digital Market Act (DMA) regulations adopted in 2022, regulating the digital platform market. The EU's work also resulted in a clear definition of disinformation, which excluded satire, parody, reporting errors or biased opinions.⁴³

³⁹ “Мінцифра звернулася з листом до Meta щодо модерації українського контенту”, *The Digital*, [on-line:] <https://thedigital.gov.ua/news/mintsifra-zvernulasya-z-listom-do-meta-shchodo-moderatsii-ukrainskogo-kontentu> (29.09.2022).

⁴⁰ “Jak tworzymy i wykorzystujemy listy zniewag typowych dla danego rynku”, *Transparency Center*, 12.08.2022, [on-line:] <https://transparency.fb.com/pl-pl/enforcement/taking-action/how-we-create-and-use-market-slurs> (29.09.2022).

⁴¹ The first version of the Code was implemented in 2018, and after periodic evaluation, it was decided to strengthen and specify its provisions. It obliges its signatories, i.a. Meta, Google, Twitter, TikTok and Microsoft to increase the transparency of their activities, demonetize disinformation, identify fake accounts increase, transparency of political advertisements, cooperate with fact-checking organizations and the scientific community. Currently, 34 entities are signatories to it, and in addition to digital platforms, it has also been joined by non-governmental organizations dealing with the fight against disinformation. See: *2022 Strengthened Code of Practice on Disinformation*, 16.06.2022, [on-line:] <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation> (29.09.2022).

⁴² COM(2020) 825 final, *Rozporządzenie Parlamentu Europejskiego i Rady w sprawie jednolitego rynku usług cyfrowych (akt o usługach cyfrowych) i zmieniające dyrektywę 2000/31/WE*, 15.12.2020 [on-line:] <https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:52020PC0825> (29.09.2022).

⁴³ Disinformation was defined as *verifiably false or misleading information created, presented and disseminated for the purpose of obtaining economic advantage or misleading the public, which is likely to cause public harm*. Public harm, in turn, was linked to threats to democratic political processes

Thanks to the actions taken, social media platforms have been obliged to increase the transparency of their activities by publishing periodic reports on the rules and effects of content moderation or the number and type of requests for access to user data submitted by state authorities.⁴⁴ The platforms were also obliged to create ads libraries which allow, in particular in the aspect of political messages, them to be analysed in terms of sponsors, financial resources and targeting criteria (although they are presented only concerning a few main variables, such as place of residence or sex).⁴⁵

At the same time, it should be noted that social media platforms have started cooperation with third parties (scientists, journalists, NGOs, etc.) in content moderation. This has taken on a rather formalised form in the case of Meta, which has set up a so-called “Oversight Board”, composed partly of members designated by the company itself and partly by representatives from the expert community, especially academia. The board investigates complaints from Facebook and Instagram users⁴⁶ on decisions made by platforms moderators, as well as taking actions on its own behalf on the basis of content monitoring. Meta can also turn to the council to take a position on a given matter.⁴⁷ The effects of the work of this body are binding decisions regarding the removal or restoration of previously deleted content, as well as non-obligatory recommendations related to the improvement of Meta’s terms of service in the field of content moderation. Among the issues raised and considered by the board was the

and the formation of policy and to public goods. See: COM(2018) 236 final, *Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów. Zwalczanie dezinformacji w Internecie: podejście europejskie*, 26.04.2020, [on-line:] <https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:52018DC0236> (29.09.2022).

⁴⁴ *Meta Transparency Center*, [on-line:] <https://transparency.fb.com/pl-pl/>; *Twitter Transparency Center*, [on-line:] <https://transparency.twitter.com/> (29.09.2022).

⁴⁵ See: *Meta Ads Library*, [on-line:] https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=PL&media_type=all (29.09.2022). Twitter owners do not maintain an up-to-date Ad Library, as they decided in 2019 to remove politics-related advertising, arguing that political coverage should be gained, not bought. See: “Twitter będzie blokował reklamy polityczne. Dorsey kpi ze stanowiska Facebooka”, *Wirtualne Media* [on-line:] <https://www.wirtualnemedial.pl/artykul/twitter-bedzie-blokowal-reklamy-polityczne-dorsey-kpi-ze-stanowiska-facebook> (29.09.2022).

⁴⁶ A request to the Board for a case to be considered can only be made after exhausting the appeal path within the platform’s internal procedure.

⁴⁷ It is worth adding that the Board does not issue decisions on all reported cases, but only on those it deems the most controversial and key in the selection process. In practice, this number is very low compared to the reported complaints. According to the report for the first quarter of 2022, out of over 479 thousand, only three of the reported cases were selected for consideration, two of which were reported by users and one by Meta. See: “Oversight Board Q1 2022 Transparency Report”, *Oversight Board*, 8.2022, [on-line:] <https://oversightboard.com/news/572895201133203-oversight-board-publishes-transparency-report-for-first-quarter-of-2022/> (29.09.2022).

decision to reinstate a comment under a post reporting on the protests in Russia in support of Alexei Navalny, under which its author called a user who criticised their participants (claiming that shamelessly abused and mentally retarded people took part) *a cowardly bot*. The board considered that while the deletion of the comment may have been in line with community standards, this failed to take into account the wider political context, disproportionately restricting freedom of expression.⁴⁸ This position shows that the rigid application of platform regulations can be disadvantageous, especially in countries where the authorities restrict freedom of speech.

Such situations can be mitigated by the aforementioned DSA, which introduces several significant changes in the context of reducing the social impact of disinformation. These include, in particular, limiting the use of sensitive data (e.g., on political views or religion), clarifying the logic behind the algorithms or making content moderation more transparent. Users are to receive a detailed explanation of the reasons for removing or limiting the visibility of their content, and above all it's a human, not an algorithm, who must examine appeals against this decision. In addition, new EU law categorises service providers according to their social impact, identifying the major online platforms (with at least 10% of the EU population, i.e., at least 45 million users today) with the greatest obligations. In addition to upholding the order associated with the ads libraries, they must ensure an adequate level of moderation in all EU languages by hiring additional moderators, performing a periodic risk analysis of their activities or providing an alternative system for selecting content recommendations to that suggested by the algorithm.⁴⁹

In addition, under the regulation, digital service coordinators will be appointed in each EU Member State. Their role will be related not only to monitoring the implementation of DSA regulations, but also to handling user complaints regarding the infringement of their rights. The changes introduced in DSA (in the context of security crises) are, therefore, important in terms of increasing the transparency of the operation of social media platforms and the principles of targeting political advertising – which may affect more effective monitoring of their activities (also in cooperation with the scientific community) and earlier response to possible risks. More transparent moderation rules have the potential to reduce the arbitrariness of social media platforms

⁴⁸ “Pro-Navalny protests in Russia”, *Oversight Board*, [on-line:] <https://www.oversightboard.com/decision/FB-6YHRXHZR/> (29.09.2022).

⁴⁹ In July 2022, Meta launched a new news feed system on Facebook called *Home*, which allows users to independently decide on the order of displayed content. See: “Introducing Home and Feeds on Facebook”, *Meta*, [on-line] <https://about.fb.com/news/2022/07/home-and-feeds-on-facebook/> (29.09.2022).

in deciding the visibility of content, and the introduction of an alternative form of *news feed* may also⁵⁰ contribute to reducing the so-called “filter bubble mechanism” (i.e., a situation where social media users function in a closed environment of uniform and often controversial and manipulated opinions). However, this does not change the fact that the individual provisions contained in the DSA are often formulated at a high level of generality, and the shape of their implementation depends on the social media platforms themselves (e.g., the rule that online platforms suspend for a reasonable period the provision of services to audiences that frequently transmit illegal content).⁵¹ Therefore, it is necessary to constantly monitor the implementation of the content of the regulation – both in the context of the transparency of the activity of the platforms and the question of their actions towards restriction of freedom of expression as well as the whole contribution to reducing the negative impact of manipulative content. National digital service coordinators and the trusted entities appointed and cooperating with them, based on the provisions of the DSA, will play a significant role in this respect.

Social media platforms as decision-making entities in the context of the war in Ukraine

The extraordinary circumstances of the security crisis triggered by Russia’s aggression against Ukraine also provoked actions by the social media platforms managers, which were taken on their own initiative as actors of influence on political realities. On the one hand, they are an expression of the continuation of their previous policy, mainly in terms of increasing the effectiveness of content moderation but on the other one can notice the introduction of new initiatives that go beyond the applicable regulations. Both Meta and Twitter have published [on their sites] lists of actions they have taken to minimise the negative effects of war-related disinformation, as well as to help the civilian population⁵² – which is also important for building a positive image [of the platforms].

⁵⁰ It is worth noting that the algorithm will remain the default option. Thus, users themselves will have to decide whether they want to use a different option for selecting content.

⁵¹ COM(2020) 825 final, *Rozporządzenie*.

⁵² Information on the actions taken by Meta and Twitter comes from the companies’ websites devoted to the subject, unless another source is mentioned in a footnote. See: S. McSweeney, “Our Ongoing Approach to the War in Ukraine”, *Twitter*, [on-line:] https://blog.twitter.com/en_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine (29.09.2022); “Meta’s Ongoing

Meta has established a special information centre with moderators who are fluent in Ukrainian and Russian to monitor the content on the platforms continuously. In addition, the concern has increased the protection of the privacy of users from Ukraine by introducing, among others, simple functionality of blocking a profile or preventing third parties from viewing their list of friends. A dedicated *Community Help* sphere has also been set up on Facebook, especially for refugees from Ukraine, where information is shared (e.g., from UN agencies) on the availability of medical assistance in Ukraine and bordering countries. It is complemented by a hotline created by the Ukrainian state services on WhatsApp messenger, where key information related to crisis management is published on an ongoing basis. In addition, Meta, within the scope of its *Data for Good* project, shares information on users' social connectedness and mobility with various trusted entities (e.g., the World Bank, Doctors Without Borders), which helps to track the flow of refugees, for example.

Both Meta and Twitter have highlighted the importance of providing higher-quality content moderation. Even before the aforementioned blocking of the accounts of major Russian media outlets, the companies had already decided to reduce the visibility of the content they disseminate, demonetise their activities and clearly mark that they are linked to Russia. Additionally, Twitter blocked the possibility of buying ads in Russia and Ukraine, and in connection with content related to the subject of the war. In addition, the platform has introduced a rule that it will not recommend users content published by representatives of governments of countries that restrict free access to information and are involved in armed conflicts. This rule was primarily applied to Russia. Twitter also introduced changes to the system of recommending content to users from Russia and Ukraine, excluding tweets from entities that they do not currently follow. In addition, based on its internal assessments, the platform reduces the visibility of content that does not explicitly violate the platform's regulations, but may contribute to social harm. Meta was more arbitrary and decided to temporarily suspend its rules of procedure concerning hate speech and allow Ukrainian users to express it in relation to the Russian aggressors. Initial reports by Reuters in early March, based on information from internal company emails, indicated that Facebook and Instagram users from Ukraine and several countries (Armenia, Azerbaijan, Georgia, Hungary, Poland, Romania, Russia, Slovakia and the Baltic States) would be able to call for the death of soldiers – and even political leaders – led by Vladimir Putin. A few days later, the representative of Meta, Nick Clegg, stipulated that this time-limited

Efforts Regarding Russia's Invasion of Ukraine", *Meta*, 26.20.2022, [on-line:] <https://about.fb.com/news/2022/02/metas-ongoing-efforts-regarding-russias-invasion-of-ukraine/#latest> (29.09.2022).

change in the company's moderation policy would only apply to Ukraine and would allow its citizens to "express their resistance and fury at the invading military forces, which would rightly be viewed as unacceptable", while pointing out that hate speech against Russians would still be banned. In conclusion it can be stated that the social media platform made an arbitrary decision that goes beyond the regulations in the face of extraordinary circumstances. In response to these reports, Russia recognised Facebook and Instagram as extremist entities and blocked their operation in Russia. This sanction is a kind of determinant of the position of the largest social media platforms in the context of their impact on shaping the information space, which, as previously indicated, is one of the crucial fields of power and political relations in their confrontational dimension.

Summary

The conducted analysis allows to conclude that social media platforms are not neutral technical tools for disseminating content on a mass scale, but a means of exerting influence, particularly through disinformation and propaganda. This makes them a key information battlefield, especially in security crises taking the form of hybrid wars. Moreover, the owners of these platforms themselves can be described as sensitive actors of influence and power. Unlike traditional political actors (states, international organisations), they use this power and influence in the form of arbitrary and not always transparent procedures. The security crisis triggered by the war has also shown that political actors often expect social media platforms to abandon neutrality in favour of a commitment to defend higher values.

Nevertheless, social media play an important role in shaping the narrative towards the conflict, which was skilfully used primarily by Ukraine. They can be seen also as the channels of mass dissemination of information that may be of importance for crisis management activities. With regard to the spread of disinformation, it seems that actions taken by many actors, including social media platforms themselves, to limit it are needed. However, they will not completely eliminate the spread of manipulated content on a large scale as they must always balance between key values of freedom of expression and security. An experiment carried out by the NATO Strategic Communication Centre of Excellence (NATO StratCom CoE) between September and November 2021⁵³ involving the purchase of more than 114,000 inauthentic activities

⁵³ NATO StratCom CoE, "Social Media Manipulation 2021/2022: Assessing the Ability of Social Media Companies to Combat Platform Manipulation", *NATO Strategic Communications Centre of*

(posts, comments, views) on Facebook, Instagram, YouTube, Twitter, Tiktok and VKontakte platforms for €279, based on Russian social media manipulation strategies, showed that 96% of them still remained active after a four-week period. This is sufficient time to spread disinformation to even millions of users, and any remedial action taken after that time may be delayed. This study indicates that the internal mechanism of platforms in combating disinformation campaigns is unreliable, the more so that the disinformation activity initiated by bots or trolls is then reinforced and sustained by the users themselves, who also become *curators* of these narratives.⁵⁴ The problem will continue to worsen, not only due to the aggressive actions in the information space undertaken by Russia or China, but the fact that younger generations, compared to older ones, rely much more on social media as sources of information about the world. From a Polish point of view, this is all the more important because, as Eurobarometer surveys have shown, Poles are the most uncertain of all EU nations as to whether they can recognise disinformation (40% of indications against an EU average of 30%).⁵⁵

In the *National Security Strategy of the Republic of Poland*, in the aspect of information space management, an emphasis was placed on counteracting disinformation, building short-term and long-term communication strategies in cyberspace, cooperating with mass media, social media and non-governmental organisations, as well as building social resistance to campaigns of manipulated content. This approach seems to cover all the key elements of managing the new media environment, including educating citizens to operate in it safely and responsibly.⁵⁶

The experience of the war in Ukraine has underlined the critical role of the social education in respect of methods of recognising and dealing with disinformation. Is it a premise, that the social media platforms will be used more responsibly (e.g., by choosing an alternative method of displaying content to the algorithm). The decision to contextualise content moderation made by the owners of Meta in response to the war in Ukraine may be perceived, on the one hand, as taking the right side of

Excellence, 27.04.2022, [on-line:] <https://stratcomcoe.org/publications/social-media-manipulation-20212022-assessing-the-ability-of-social-media-companies-to-combat-platform-manipulation/242> (29.09.2022).

⁵⁴ Y. Golovchenko, M. Hartmann, R. Adler-Nissen, "State, Media and Civil Society in the Information Warfare over Ukraine: Citizen Curators of Digital Disinformation", *International Affairs*, vol. 94, no. 5 (2018), pp. 975–994.

⁵⁵ Survey conducted at the turn of April and May 2022, See: Flash Eurobarometr, *News&Media Survey*, [on-line] <https://europa.eu/eurobarometer/surveys/detail/2832> (29.09.2022).

⁵⁶ *Strategia Bezpieczeństwa Narodowego Rzeczypospolitej Polskiej 2020*, Warszawa 2020, [on-line] https://www.bbn.gov.pl/ftp/dokumenty/Strategia_Bezpieczenstwa_Narodowego_RP_2020.pdf (28.09.2022).

the conflict, but on the other, it also reveals a dangerous tendency to arbitrarily make decisions that have a key impact on shaping the public debate and thus the exercise of power in a networked society. Therefore, it is important for the state administration as well as NGOs to actively cooperate and monitor the activity of the social media platforms, which, among others, is expressed in the provisions of the DSA, so that a wider group of actors takes key decisions for shaping the narrative. However, the most important action should be to make citizens aware that in the era of a highly individualised media environment, they also have greater responsibility in searching for reliable sources of information and functioning in the space of many, sometimes controversial opinions.⁵⁷ Such an approach will help both reduce the negative impact of social media platforms on democratic debate and, at the same time, protect the digital space from attempts to arbitrarily restrict it in terms of freedom of speech in favour of superior and broadly understood security.

References

- 2022 *Strengthened Code of Practice on Disinformation*, 16.06.2022, [on-line:] <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>.
- Bastos M.T., Mercea D., "The Brexit Botnet and User-generated Hyperpartisan News", *Social Science Computer Review*, vol. 37, no. 1 (2019), pp. 38–54, <https://doi.org/10.1177/0894439317734157>.
- Bilal A., "Hybrid Warfare – New Threats, Complexity, and 'Trust' as the Antidote", *NATO Review* 2021, [on-line:] <https://www.nato.int/docu/review/articles/2021/11/30/hybrid-warfare-new-threats-complexity-and-trust-as-the-antidote/index.html>.
- Castells M., *Siła tożsamości*, transl. S. Szymański, Wydawnictwo Naukowe PWN, Warszawa 2009.
- Castells M., *Spółeczeństwo sieci*, crowd. M. Marody et al., transl. M. Marody, Wydawnictwo Naukowe PWN, Warszawa 2011.
- Castells M., *Władza komunikacji*, transl. J. Jedliński, P. Tomanek, Wydawnictwo Naukowe PWN, Warszawa 2013.
- Ciuriak D., *The Role of Social Media in Russia's War on Ukraine*, 5.05.2022, [on-line:] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4078863.
- COM(2020) 825 final, *Rozporządzenie Parlamentu Europejskiego i Rady w sprawie jednolitego rynku usług cyfrowych (akt o usługach cyfrowych) i zmieniające dyrektywę 2000/31/WE*, 15.12.2020, [on-line:] <https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:52020PC0825>.

⁵⁷ A. Jungherr, R. Schroeder, "Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy", *Social Media+Society*, vol. 7, no. 1 (2021), pp. 1–13.

- COM(2018) 236 final, *Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów. Zwalczanie dezinformacji w Internecie: podejście europejskie*, 26.04.2020, [on-line:] <https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:52018DC0236>.
- The Communication Decency Act of 1996*, 47 U.S. Code § 230.
- Danek M., "Komunikowanie polityczne w dobie fake newsów – walka z dezinformacją w sieci", *Krakowskie Studia Małopolskie*, vol. 23 (2018), pp. 210–228, <https://doi.org/10.15804/ksm201811>.
- Danek M., "Modele walki z dezinformacją: od restrykcji do współpracy", in M. Bernaczyk, T. Gąsior, J. Misiuna, M. Serowaniec (eds), *Znaczenie nowych technologii dla jakości systemu politycznego. Ujęcie politologiczne, prawne i socjologiczne*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2020.
- Dijk van J., *Społeczne aspekty nowych mediów*, transl. J. Konieczny, Wydawnictwo Naukowe PWN, Warszawa 2010.
- EUvsDisinfo*, [on-line:] <https://euvsdisinfo.eu/pl/#>.
- "Facebook i Twitter zablokowane w Rosji", *Wirtualne Media*, 5.03.2022, [on-line:] <https://www.wirtualnemedi.pl/artykul/rosja-blokada-facebook-twitter>.
- Flash Eurobarometr, *News&Media Survey*, 6.2022, [on-line:] <https://europa.eu/eurobarometer/surveys/detail/2832%20>.
- Foucault M., *Nadzorować i karać. Narodziny więzienia*, transl. T. Komendant, Wydawnictwo Aletheia, Warszawa 2009.
- A Front for Influence: An Analysis of a Pro-Kremlin Network Promoting Narratives on COVID-19 and Ukraine*, 24.08.2022, [on-line] <https://cyber.fsi.stanford.edu/io/news/sio-aug-22-takedowns-ua>.
- Gerasimow V., "The Value of Science Is in the Foresight New Challenges Demand Rethinking the Forms and Methods of Carrying out Combat Operations", *Military Review* (2016), pp. 23–29, https://www.armyupress.army.mil/portals/7/military-review/archives/english/militaryreview_20160228_art008.pdf.
- Gigitashvili G., *Twitter Campaign Pushes Anti-Ukraine Hashtag into Poland's Trending List*, 8.2022, [on-line:] <https://medium.com/dfrlab/twitter-campaign-pushes-anti-ukraine-hashtag-into-polands-trending-list-90ccc9474a60>.
- Golovchenko Y., Hartmann M., Adler-Nissen R., "State, Media and Civil Society in the Information Warfare over Ukraine: Citizen Curators of Digital Disinformation", *International Affairs*, vol. 94, no. 5 (2018), pp. 975–994, <https://doi.org/10.1093/ia/iiy148>.
- Guess A., Nyhan B., Reifler J., *Selective Exposure to Misinformation: Evidence from the Consumption of Fake News during the 2016 US Presidential Campaign*, 9.01.2018, [on-line:] <https://about.fb.com/wp-content/uploads/2018/01/fake-news-2016.pdf>.
- Hughes Ch., "It's Time to Break Up Facebook", *The New York Times*, 9.05.2019, [on-line:] <https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html>.
- "Introducing Home and Feeds on Facebook", *Meta*, 21.07.2022, [on-line:] <https://about.fb.com/news/2022/07/home-and-feeds-on-facebook/>.

- “Jak tworzymy i wykorzystujemy listy zniewag typowych dla danego rynku”, *Transparency Center*, 12.08.2022, [on-line:] <https://transparency.fb.com/pl-pl/enforcement/taking-action/how-we-create-and-use-market-slurs>.
- Jungherr A., Schroeder R., “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>.
- “Kalendarium – sankcje UE wobec Rosji w sprawie Ukrainy”, *Strona Rady Europejskiej i Rady Unii Europejskiej*, [on-line:] <https://www.consilium.europa.eu/pl/policies/sanctions/restrictive-measures-against-russia-over-ukraine/history-restrictive-measures-against-russia-over-ukraine/>.
- Kemp S., “Digital 2022: Global Overview Report”, *DataReportal*, 26.01.2022, [on-line:] <https://datareportal.com/reports/digital-2022-global-overview-report>.
- Kissinger H., *Porządek światowy*, transl. M. Antosiewicz, Wydawnictwo Czarne, Wołowiec 2017.
- Lima C., “A Whistleblower’s Power: Key Takeaways from the Facebook Papers”, *Washington Post*, 26.10.2021, [on-line:] https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/?itid=co_facebookunderfire_1.
- McKay S., Tenove C., “Disinformation as a Threat to Deliberative Democracy”, *Political Research Quarterly*, vol. 74, no. 3 (2021), <https://doi.org/10.1177/1065912920938143>.
- McSweeney S., “Our Ongoing Approach to the War in Ukraine”, *Twitter*, [on-line:] https://blog.twitter.com/en_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine.
- Merrill J.B., Oremus W., “Five Points of Anger, One for a ‘Like’: How Facebook’s Formula Fostered Rage and Misinformation”, *The Washington Post*, 26.10.2021, [on-line:] https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/?utm_campaign=wp_main&utm_medium=social&utm_source=twitter.
- Meta Ads Library*, [on-line:] https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=PL&media_type=all.
- “Meta’s Ongoing Efforts Regarding Russia’s Invasion of Ukraine”, *Meta*, 26.20.2022, [on-line:] <https://about.fb.com/news/2022/02/metass-ongoing-efforts-regarding-russias-invasion-of-ukraine/#latest>.
- Meta Transparency Center*, [on-line:] <https://transparency.fb.com/pl-pl/>.
- “Mincifra zvernulasâ z listom do Meta šodo moderacii ukrâins’kogo kontentu” [“Мінцифра звернулася з листом до Meta щодо модерації українського контенту”], *The Digital*, 27.07.2022, [on-line:] <https://thedigital.gov.ua/news/mintsifra-zvernulasya-z-listom-do-meta-shchodo-moderatsii-ukrainskogo-kontentu>.
- NATO StratCom Coe, “Social Media Manipulation 2021/2022: Assessing the Ability of Social Media Companies to Combat Platform Manipulation”, *NATO Strategic Communications Centre of Excellence*, 27.04.2022, [on-line:] <https://stratcomcoe.org/publications/social-media-manipulation-20212022-assessing-the-ability-of-social-media-companies-to-combat-platform-manipulation/242>.
- “Online Safety Bill to Return as Soon as Possible”, *BBC*, 20.09.2022, [on-line:] <https://www.bbc.com/news/technology-62908598>.

- "Oversight Board Q1 2022 Transparency Report", *Oversight Board*, 8.2022, [on-line:] <https://oversightboard.com/news/572895201133203-oversight-board-publishes-transparency-report-for-first-quarter-of-2022/>.
- Połowaniuk M., "Rosyjska inwazja to pierwsza 'wojna TikTokowa' Ukraina pokazuje, jak wygrać bitwę w sieci", *Spider's Web*, 28.02.2022, [on-line:] <https://spidersweb.pl/2022/02/jak-wygrac-wojne-w-internecie.html>.
- Profile of Volodymyr Zelensky on Twitter, [on-line:] <https://twitter.com/ZelenskyyUa>.
- "Pro-Navalny Protests in Russia", Oversight Board, [on-line:] <https://www.oversightboard.com/decision/FB-6YHRXHZR/>.
- Putnam R., *Samotna gra w kręgle. Upadek i odrodzenie wspólnot lokalnych w Stanach Zjednoczonych*, transl. P. Sadura, S. Szymański, Wydawnictwo Akademickie i Profesjonalne, Warszawa 2008.
- Rocha Y.M. et al., "The Impact of Fake News on Social Media and Its Influence on Health during the COVID-19 Pandemic: A Systematic Review", *Journal of Public Health* (2021), <https://doi.org/10.1007/s10389-021-01658-z>.
- "Russia Passes Legislation Banning 'Disrespect' of Authorities and 'Fake News'", *The Moscow Times*, 7.03.2019, [on-line:] <https://www.themoscowtimes.com/2019/03/07/russia-passes-legislation-banning-disrespect-of-authorities-and-fake-news-a64742>.
- Saurwein F., Spencer-Smith C., "Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe", *Digital Journalism*, vol. 8 no. 6 (2020), pp. 820–841, <https://doi.org/10.1080/21670811.2020.1765401>.
- Scott M., "Facebook Did Little to Moderate Posts in the World's Most Violent Countries", *Politico*, 25.10.2021, [on-line:] <https://www.politico.com/news/2021/10/25/facebook-moderate-posts-violent-countries-517050>.
- "Strategia Bezpieczeństwa Narodowego Rzeczypospolitej Polskiej 2020", GOV, Warszawa 2020, [on-line:] https://www.bbn.gov.pl/ftp/dokumenty/Strategia_Bezpieczenstwa_Narodowego_RP_2020.pdf.
- Sun Tzu, *Sztuka wojny*, transl. J. Zawadzki, Hachette, Warszawa 2009.
- Tandoc Jr. E.C., Lim Z.W., Ling R., "Defining 'Fake News': A Typology of Scholarly Definitions", *Digital Journalism*, vol. 6, no. 2 (2018), pp. 137–153, <https://doi.org/10.1080/21670811.2017.1360143>.
- "Twitter będzie blokował reklamy polityczne. Dorsey kpi ze stanowiska Facebooka", *Wirtualne Media*, 31.10.2019, [on-line:] <https://www.wirtualnemedial.pl/artykul/twitter-bedzie-blokowal-reklamy-polityczne-dorsey-kpi-ze-stanowiska-facebook>.
- "Twitter Transparency Center", [on-line:] <https://transparency.twitter.com/>.
- Wasiuta O., "Geneza pojęcia i zmiany podejścia do wojny hybrydowej w zachodnim dyskursie politycznym i wojskowym", *Przegląd Geopolityczny* 2016, no. 17, pp. 26–40.

Information security is one the key aspects of modern security and its importance has been significantly increasing in contemporary international relations. This publication presents the results of studies on several key aspects related to this issue. The publication contains results of research on considerations related to information security and its implementation, as well as research on social media, analysed through the lens of the object and subject of disinformation activities.



<https://akademicka.pl>

