The Copernicus Journal of Political Studies 2023/2024
No. 1 2023/2024, pp. 35–59
ISSN 2299-4335
DOI: http://dx.doi.org/10.12775/CJPS.2023/2024.003
www.apcz.umk.pl/czasopisma/index.php/CJPS

Rafał Paradowski
Uniwersytet im. Jana Długosza w Częstochowie
rafalparadowski@hotmail.com

TRENDS AND MECHANISMS OF RUSSIAN DISINFORMATION CONCERNING POLAND: THE ANALYSIS OF INFORMATION OPERATIONS ON TWITTER IN 2009–2020

ABSTRACT

With the rise of social media, disinformation has become an important instrument in information warfare at the state level. The article aimed to determine the dominant directions of the Russian Federation's disinformation and propaganda activities on the Internet concerning Poland and its citizens. The study used content made available as part of Twitter's *Information Operations* project. It focused on the dynamics of the communication, the range of topics discussed and the emotions they evoke in recipients. From four databases containing almost 10 million tweets from 2009–2020, a sample related to Poland was extracted and analysed using data mining techniques. The findings showed that the discourse promoted by Russian services focused on undermining public trust, stoking social tensions and feeding anxiety. The impact of messages included both institutional (EU countries, NATO members) and interpersonal relationships (ethnic groups, representatives of professions, individuals). The research also illustrated the functioning of some disinformation mechanisms in practice. Its results may constitute a reference point for future studies of online hostile communication activities.

Keywords

disinformation, information warfare, national security, foreign intelligence, social data science

Introduction

With the rise of social media, disinformation has become a well-established tool of information warfare at the state level. This weapon is used to influence democratic election campaigns, moderate the actions of other countries in the international arena, and shape discourses regarding domestic affairs (Colliver et al., 2018; Giles, 2015; Lanoszka, 2019; Watanabe, 2018; Zannettou, Caulfield, De Cristofaro, et al., 2019). The deliberate spread of false or biased content is carried out through many media channels, but the role of social media is unique. Their main advantage is the ability to quickly reach large audiences with the desired demographic characteristics. Moreover, they allow imitating civil disobedience attitudes and grassroots movements (Keller et al., 2017, 2019). Recently, the use of disinformation practices on social media has been proven primarily to the Russian Federation (Bail et al., 2020; Bodine-Baron et al., 2018; Giles, 2015; Unver, 2019), but also to Iran, Venezuela, Turkey, North Korea or China (Conger, 2019; Dolan, 2022; Wilson, 2022).

The general aim of this study was to determine the dominant directions and goals of disinformation and propaganda activities of the Russian Federation on the Internet that concerned Poland and its citizens. It focused on defining the false and harmful image of the country created for a broad English-speaking audience on a sample social media platform. Therefore, the study used English-language content published on Twitter and made available by the administrators of this service as part of the public project the *Information Operations* (Gadde & Roth, 2018; Roth, 2019; Roth & Gadde, 2022). The study had an exploratory character and utilized the existing knowledge in the field of internet disinformation. In particular, attention was paid to the dynamics of communication, the range of topics covered and the emotions that could potentially be evoked in the audience. The first hypothesis stated that the main objective of Russia's Twitter information war against Poland was to create a sense of general threat and distrust in Polish society with regard to relations with democratic institutions, other states and communities. This could have led to social, political and economic destabilization at the domestic level as well as to distortion of relations with Poland's allies. The second assumption was that the topics discussed changed depending on the geopolitical situation and were determined by the internal affairs of the affected country.

Disinformation, misinformation and propaganda

The phenomena of information manipulation are not uniform. We find at least three key terms whose meanings largely coincide with the issues discussed in this paper: disinformation, misinformation and propaganda. Their similarity lies in the fact that "all three concern false or misleading messages spread under the guise of informative content, whether in the form of elite communication, online messages, advertising, or published articles" (Guess & Lyons, 2020, p. 10). However, the first of these is the key term of this study. According to the European Commission, disinformation is "verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm. Public harm includes threats to democratic processes as well as to public goods such as Union citizens' health, environment or security" (European Commission, 2018, p. 2). Such a broad denotation may include also systemically generated propaganda messages with a variety of attitudes towards the same objects (i.e. propaganda as a form of disinformation) and other complementary communication practices conducted to mislead the public. A similar interpretation of disinformation is presented by the Polish National Security Bureau (*Biuro Bezpieczeństwa Narodowego*), which emphasizes that this phenomenon exists today in many forms, including propaganda, ideological sabotage or specific actions tailored to particular target groups (Wrzosek, 2019, p. 9). It should be added that disinformation (along with misinformation and malinformation) can be considered a type of information disorder or information pollution (Wardle & Derakhshan, 2017, pp. 5–6). However, Giles (2015, pp. 12–13) points out that all communication activities carried out by state agents should be considered very broadly, as an element of information warfare.

Wardle proposes a typology of seven disinformation activities that can be used in political activities (originally called "types of mis- and dis-information"). The first two are satire or parody and false connections (e.g. misleading headlines, captions, clickbait). Hostile information operations conducted by another state or its agents mainly apply to types 3–7, i.e. misleading content, false context, imposter content, manipulated content and fabricated content (Wardle & Derakhshan, 2017, pp. 16–17). They reflect a wide spectrum: from content classified as true but used inappropriately or presented in a manipulated context, to completely false messages.

The most important feature of disinformation activities is their intentionality to advance political goals (Bennett & Livingston, 2018, p. 124). The recipient

is offered apparent, worthless or harmful knowledge to achieve the intended effect, i.e. to influence them to make faulty decisions in favor of the disinformer (Bielawski & Ziółkowska, 2018, p. 94). Another characteristic of disinformation is the use of automated and aggressive ICT techniques to disseminate or amplify content in online communication, such as bots and artificial intelligence (Ajir & Vailliant, 2018, pp. 75–76; Wrzosek, 2019, p. 7).

In the literature, there is no clear demarcation of the scope of the concepts of disinformation and propaganda. Guess and Lyons (2020) point out that the definitions of both phenomena overlap and are used interchangeably. According to Lock and Ludolph (2019), the concept of propaganda has a longer history than disinformation and can be used as a very general category covering all institutionalized persuasive activities carried out by public and private entities, with varying intensity and many methods, not necessarily with malicious intentions. In turn, Gorwa (2017) proposes using the term "computational propaganda" to describe all organized disinformation and manipulation campaigns appearing in today's media, especially on the Internet. In addition, propaganda may be considered as a separate means of exerting communicative influence, and disinformation could be treated as complementary to it on a par with intelligence activities (Bielawski & Ziółkowska, 2018, p. 90; *Krajobraz bezpieczeństwa polskiego internetu raport roczny 2019*, 2020, pp. 7–9). In the last context, the main purpose of propaganda is to create a positive image of the entity that is its author. However, the cited examples of activities of Russian intelligence services suggest that the occasional spreading of favorable content about the sender-state is only one of the techniques used in an integrated and long-term disinformation campaign. The same applies to misinformation, which is inherently unintentional or inadvertent. In the discussed context, accidental misleading of the recipient in some cases may also implement a top-down policy of information manipulation. This phenomenon is usually based on the behavior of "ordinary" social media users who disseminate malicious content just because the message corresponds with their views, is attractive or shocking (Chitra & Musco, 2019; Grzywińska & Batorski, 2016; Zannettou, Caulfield, Setzer, et al., 2019). This is done with the significant involvement of filtering algorithms operating in the background of social media applications – their mechanisms can be intentionally (mis)used by hostile state agents (Kreft, 2019; Wardle & Derakhshan, 2017, pp. 20–24).

Summarizing the above considerations, this study assumes that all information operations conducted by the Russian Federation on social media will be referred to as disinformation. This approach covers various online communication

activities and allows them to be treated as an element of information warfare (Ajir & Vailliant, 2018, pp. 75–81).

Selected mechanisms of online disinformation

Disinformation content has various textual and graphic forms, and often relies on symbols and single associations. It utilizes national and cultural stereotypes, refers to emotions and elements of common wisdom, and sometimes pretends to look like expert knowledge. In the context of entire communities, false or biased content tends to weaken the constitutive elements of attacked groups: norms, values, objectives and interests, communication structure and identity. The audience's ability to reach a collective consensus and maintain social order can be reduced, which is achieved by stoking debates on controversial topics, feeding fears, building conspiracy theories, depreciating existing authorities, etc. (Bastos & Farkas, 2019; Colliver et al., 2018; Gruzd & Mai, 2020; Johnson et al., 2017; Keller et al., 2019; Watanabe, 2018). Also, the mechanism of cognitive dissonance has a special place in the list of manipulation techniques. It is based on rationalizing and justifying behavior considered inappropriate by the potential recipient, and attacking the user with varied and contradictory messages (Bielawski & Ziółkowska, 2018, p. 94).

Disinformation can be a key tool in information warfare. This activity encompasses a wide variety of multidimensional information processing practices, with the primary aim of influencing the adversary's knowledge, emotions and behavior to gain an advantage over him (Lelonek, 2018, pp. 69–70). According to the theory of information warfare, actions related to the transmission of fabricated content are directed at two categories of targets. The first is the mental resources, e.g. ways of perceiving reality, values, motivations and beliefs, attacked at the level of collective and individual consciousness. The second category of targets includes elements of a social nature, including rules for the transmission of information or decision-making that are not vulnerable to conventional attacks like physical assets (Lelonek, 2018, pp. 71–72). It should be emphasized that both sets of aims can be realized against military and civilian populations (Batorowska et al., 2019, p. 166).

The ultimate effect of disinformation, according to the theory of information warfare, is to subjugate a state from the inside, with little or no military force. This means ideological conquest, i.e. a situation in which the opponent completely changes their perception of reality, and the system of values is replaced by hostile elements (Batorowska et al., 2019, p. 145). Achieving this goal

is possible, but it is a long-term and complex process. Therefore, an information campaign with such an aim must be planned as a sustained, strategic initiative (e.g. Russian actions against Ukraine). However, it is much easier and more effective to aim communication weapons at destabilization of the social, political or economic order, and at the same time to stimulate those attitudes of people which are consistent with the aggressor's goals. The realism of this scenario was confirmed in a report summarizing several years of disinformation activities of Russian intelligence in social networks against the United States (DiResta et al., 2019, p. 99).

The secret to the effectiveness of modern disinformation activities using social media lies in the ease of generating new content and the availability of numerous tools for its dissemination. This means that the benefits of information warfare are reaped by the party that has initiated a given discourse (Batorowska et al., 2019, p. 225).

Attention should also be paid to other features of the online space that are particularly conducive to the distribution of false and harmful content. One of them is the mentioned activity of content-filtering algorithms (Kreft, 2019). Further features include a very low degree of social control during online interactions, user anonymity and general availability of social media applications. This results in users' expressiveness, ease of interaction and thoughtless redistribution of content without critical evaluation. Moreover, the Internet favors niche attitudes, which become more visible among the general information recipients (Chandio & Sah, 2020). Social media also create favorable conditions for the radicalization of political views (Bail et al., 2018, 2020; Furman & Tunç, 2019; Zannettou, Caulfield, Setzer, et al., 2019).

The Russian model of information manipulation

The Russian Federation is a source of exemplary and innovative disinformation activities on social media. The essential feature of that manipulation system is not its secrecy, but the degree of complexity and specialization. The various elements are complementary and hierarchical while ensuring maximum diversification of distribution channels. Experts call this model the "Russian Disinformation Chain" (Bodine-Baron et al., 2018, pp. 7–11). The top of this structure consists of the head of state, their closest associates, and senior army and intelligence officials. The second level is made up of media organs and proxies, including TV stations, radio stations, and online news services, e.g. Russia Today, Sputnik, Baltic Media Alliance, local Russian media and other entities like the Internet

Research Agency (IRA). Next come amplification channels, which include all social media, bots (scripts or programs that perform automatic, pre-programmed tasks on the web), unaffiliated or random websites, and occasionally American and European media. The final components of the system are the consumers of the content. This system resembles a top-down managed media corporation designed to maintain an imperial discourse and ensure information superiority over the West (Gac, 2020, p. 92; Unver, 2019).

Russian disinformation operations also make use of official profiles of their own or friendly diplomatic offices, politicians, activists, and other public figures and institutions operating beyond the borders of the initiator of the misleading content. The recipient is given an impression that the message is widespread and presumedly credible, which may encourage them to further disseminate this content, in a completely organic manner. One study (Bail et al., 2020, p. 245) showed that 19% of the analyzed group of Americans had interacted with camouflaged accounts of the Russian Internet Research Agency on Twitter, and 11.3% of respondents had directly engaged in discourse with so-called trolls.

Valuable insights into the diverse disinformation activities were provided by the report of the United States Senate Select Committee on Intelligence (Mueller, 2019). It demonstrated that the Russian Internet Research Agency maintained not only automated accounts on Twitter but also individualized profiles that interacted freely with other users. As a result, a network of authentic people with similar views had been created. The context for these activities was the 2016 US presidential campaign. Misappropriated profiles of American activists, groups and thematic services were also used in the information battle. These accounts were used to top-down disseminate prepared disinformation content or start new discussions around selected political and social topics (e.g. immigrants or gun ownership). Such actions made it possible to undermine the trust of American citizens in the electoral process, politicians and the government. Similar rhetoric was used immediately before the 2018 Swedish general election, when 2,000 messages suggesting electoral fraud appeared on Twitter (Colliver et al., 2018, p. 6). It can be noted that a significant part of Russian disinformation content is widely used by far-right media to discredit their political opponents (Bennett & Livingston, 2018, pp. 132–133).

Another example of identified Russian disinformation activities was the campaign for Britain's exit from the European Union (Krasodomski-Jones et al., 2018; Russia Report, 2020). Its aim was to destabilize all countries of the community by promoting Brexit. EU membership was described in the language of supposed economic losses, and demographic and security threats to the country.

Analogous actions against the UK also took place in 2017, with a focus on highlighting the threat of Islam, terrorism and immigration (Krasodomski-Jones et al., 2018, pp. 10–14).

The information campaigns targeted at Lithuania, Latvia and Estonia were dominated by creating an atmosphere of instability and fear by presenting a negative image of the EU and NATO (Gac, 2020, pp. 92–95). Historical stereotypes were used, glorifying Soviet Russia and accusing the Baltic states of collaboration with the Third Reich. A policy of polarizing society along ethnic issues was also implemented. Similar actions were targeted towards Poland, although at a different intensity level, with a slightly varied context and proportions of the themes raised. In this case, disinformation has served to build tension in Polish-Ukrainian relations, depreciate the country's defense capabilities, stimulate a sense of insecurity among citizens, and weaken ties with other NATO and EU member states (Kmiecik, 2019, p. 100; *Krajobraz bezpieczeństwa polskiego internetu raport roczny 2019*, 2020, pp. 41–43). The specificity of Russian influence in Poland also includes: fueling distrust towards Israel and the Jewish community in the world; playing with the image of Russia and its place in Polish security policy; undermining the credibility of public services and institutions (e.g. the Catholic Church); stoking religious dissonance; and alluding to pan-Slavic ideas (Wrzosek, 2019, p. 9).

The above examples constitute a small part of the online disinformation activities carried out by the Russian Federation before the invasion of Ukraine in 2022. They are diverse in terms of the topics discussed, methods of communication used and level of involvement. Moreover, it can be assumed that they are subject to adjustments over time, depending on needs and the current geopolitical situation. Evidence of the flexibility and specificity of the indicated activities seems to be provided by the analysis of the infosphere in Turkey. In this case, Russian information operations on the Internet and traditional media were aimed at legitimizing the government and promoting cooperation with Russia, while at the same time they were also ready to conduct, if needed, disinformation activities e.g. referring to the opposition (Unver, 2019).

Twitter's Information Operation project

Twitter, a microblogging service and social network, was launched in 2006. It allows registered users to publish brief real-time text and multimedia messages ("tweets") and interact with each other. The role of Twitter in the creation of social communication is reflected by the commonly used neologism *twittersphere*

as it stands for opinion-forming network in the context of social and political issues. Journalists regard the service as a source of information, so tweets are often quoted in traditional visual media (von Nordheim et al., 2018).

The service is particularly popular in the US (69.3 million active users), Japan (50.9 million), India (17.5 million), the UK (16.45 million), and Brazil (16.2 million) – estimates as of January 2021. This adds up to almost 190 million monetizable daily active users worldwide, as calculated for Q3 2020 (Tankovska, 2021).

It can be assumed that the service's administrators had not considered information manipulation a serious threat before 2016. A turning point in Twitter's security policy was the report related to the possible influence of Russia's Internet Research Agency on the outcome of the US presidential election in November 2016 (Assessing Russian Activities and Intentions in Recent US Elections, 2017). In September 2017, Twitter presented its first in-depth analysis of Russian disinformation activities ("Update: Russian Interference", 2017). In a parallel action, the European Commission also drew attention to the problem of the use of social media by foreign states to manipulate citizens. The work of EU officials produced in 2018: the EU Code of Practice on Disinformation, which was signed by representatives of the Internet and advertising industry, including Twitter Inc. ("2018 Code of Practice on Disinformation", 2022).

Since 2018, the site's administrators systematically published sets of tweets attributed to state-linked information operations originating from Russia, Iran, Bangladesh, Venezuela, Spain, China, UAE, Egypt, Saudi Arabia, Ecuador, Turkey, Cuba, Ghana/Nigeria, Serbia, Honduras, Indonesia, Thailand, Armenia, Tanzania, Mexico and Uganda (the order of countries follows the chronology of published sets). Content authored by the GRU (Russian Chief Intelligence Office) and IRA (Russian Internet Research Agency) has been marked separately. The project was carried out until the first quarter of 2022 under the name "Information Operations". Data for this period is still public. It primarily covers those entries that are closely related to US international politics and domestic affairs, and to a lesser extent those related to the EU. Other suspicious tweets were made available only to authorized entities as part of the so-called Twitter Moderation Research Consortium (Gadde & Roth, 2018; Moderation Research – Twitter Transparency Center, n.d.).

Sample approaches in Twitter research

The studies using Twitter data are quite interdisciplinary. They are founded mainly on knowledge and tools introduced by computer science and linguistics

with the theoretical framework of other disciplines referring to particular research projects. The most common are investigations concerning civic attitudes and electoral campaigns (Chauhan et al., 2021; Jain & Kumar, 2017), social communication and political marketing practices (Adamik-Szysiak, 2014; Belford et al., 2016; Gorwa, 2017; Johnson et al., 2017; Smith, 2019), consumer attitudes and advertising evaluation (Asur & Huberman, 2013; Kisiołek, 2018).

Researchers have repeatedly attempted to identify disinformation practices in social networks. So far, no unified scientific approach has emerged. Most authors recognize the need for continuous adaptation of techniques and analytical models. For example, Gorwa (2017, p. 26) pointed out the need for a better understanding of the mechanisms of accounts' activity and the techniques of political influence, as well as the lack of knowledge about the real impact of disinformation on the audience. Keller et al. (2019, p. 567) highlighted that behind every astroturfing information campaign, there is a pattern of actions in line with bureaucratically imposed policies. Rogers and Niederer (2020, pp. 51–53) pointed out the particular importance of fake news in the manipulation of social media users and noticed great variation in the techniques of creating and disseminating such content. Krasodomski-Jones et al. (2018, pp. 2, 15–16) found that the information attack phase can be preceded by a preparatory period, which is characterized by a surge of user activity publishing apolitical content or unreadable spam. The same authors also observed that fake accounts can, while attacking the primary target, generate messages that also hit other countries as collateral impact. However, Bail et al. (2020) argued for the ineffectiveness of Russian intelligence disinformation campaigns in relation to the 2016 US presidential election, which was explained by the limited impact of polarizing messages, which mainly reached already radicalized users within specific information bubbles. Colliver et al. (2018) applied a complex model to monitor multiple social media. They demonstrated that accounts involved in disinformation activities can form networks of cross-border associations and be explicitly inspired by a variety of foreign actors, including official media and political organizations.

Twitter data research is mainly based on text analysis. The most widely used techniques are: counting the word frequency, sentiment analysis and topic modeling. The first is the simplest, most standardized and fastest, while the others require the implementation of algorithms at the level of manual or automatic language processing. Both sentiment analysis and topic modeling allow the use of supervised or unsupervised learning methods to label or categorize respectively (Deho et al., 2018; Kharde & Sonawane, 2016; Michalak, 2017).

Research method

Tweets were downloaded from the public Twitter database within the *Information Operations* project (https://transparency.twitter.com/en/reports/information-operations.html). Firstly, the data was converted from CSV format into spreadsheet files, which facilitated its subsequent querying. Four out of seven datasets linked by the site administrators to the Russian Federation were selected, which had the highest occurrences of keywords: "Poland" and "Polish" (Table 1). During filtering, tweets that were not written in English or contained words correlating only incidentally with the requested terms were removed (e.g. "polishing", "polandmary"). Thus, out of around 9.78 million entries, only 1609 lines were left, containing meta-information about particular users and specific tweets. The study did not include content written in Polish because it was almost absent.

Table 1. Volume of tweets in databases

	2021-02 Russia IRA	2021-02 Russia GRU	2019-01 Russia	2018-10 Russia	TOTAL
	n				N
Raw tweets	68,914	26,684	≈0.92*10^6	8.88*10^6	≈9.78*10^6
Selected tweets	181	11	422	995	1,609

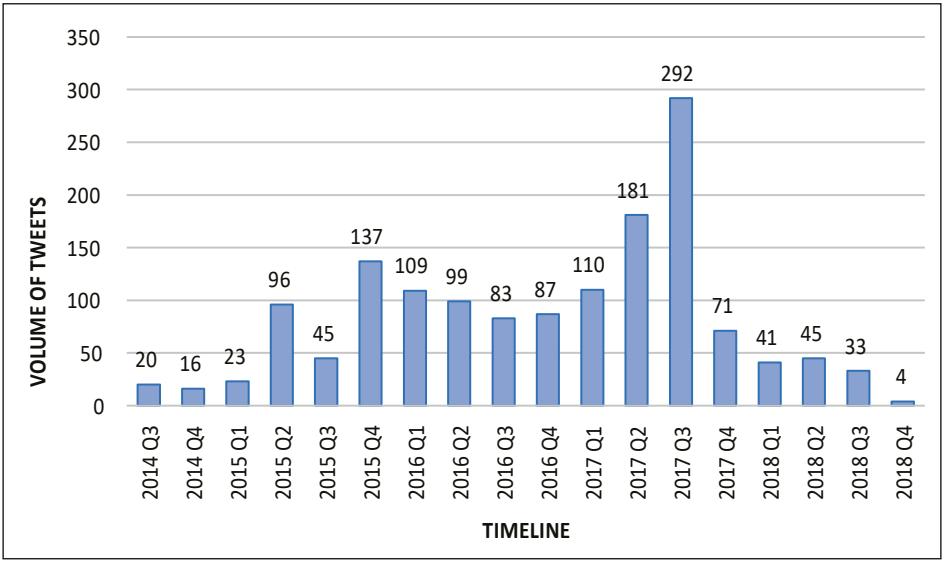

Fig. 1. Timeline of tweets
Note: other periods were omitted in the chart.

Tweets covered dates from 31 August 2009 to 21 October 2020, with a significant concentration in the years 2015–2017, which accounted for as much as 82.85% of the selected posts in the entire timeline (Fig. 1.).

Table 2. Distribution of tweets and their authors

Volume of tweets			User group characteristics			Mean tweets per user group
quartile	n	% cum.	n	followers	following	
1st	464	28.84	4	30,366	14,564	116
2nd	348	50.47	6	38,336	19,580	58
3rd	417	76.38	31	812,985	415,410	13.45
4th	380	100.00	237	n/a	n/a	1.60
TOTAL	1,609	–	278	–	–	–

Content relating to Poland was mainly published by a group of specialized users, forming the vocal minority – 17.3% of all accounts (41 from 237 profiles) generated 76.38% of analysed tweets. This small group was observed by a large audience of 881,687 users. It can be assumed that even around 50% of this figure might have been constituted by genuine Twitter users, according to the followers–following ratio.

The study used a variety of complementary data processing and analysis techniques:

1. **Topic modeling**. Each tweet was assigned to one of 15 thematic categories. Their list was developed through a two-phase coding, which began by grouping posts according to the NASK typology of "Russian influence matrices" (Wrzosek, 2019, pp. 8–9). The pattern was based mainly on hashtags and nouns occurring in the text. The following were labeled as excluded (and removed from further calculations):
   - sports information,
   - mentions of Donald Trump without relevance to the work or with an incomprehensible meaning, e.g. "Do you vote for #Trump in 2020 ????? #Warsaw #TrumpinPoland #TRUMPwPOLSCE Vote and Retweet!!"
   - tweets containing words related to Poland without a clear context, of an occasional or entertaining nature, e.g. "Meet the Amazing Cat from Poland that Takes Care of Sick Animals".

2. **Semantic analysis**. After topic modeling, each tweet was assigned to one of seven emotion subgroups (two for positive, five for negative) or labeled as undefined. This coding was also carried out in two phases: trial and proper.
3. **Distribution of most frequent hashtags**. Volumetric analysis of tweets was done in linguistic software KWIC Concordance (http://dep.chs.nihon-u.ac.jp/english_lang/tukamoto/kwic_e.html).
4. **In-depth analysis of the most active users**. Metadata about the activity of selected profiles was combined with the results of topic modeling and sentiment analysis.

	Topic modeling (thematic categories)	Semantic analysis (emotions)	Two human coders (HC) or machine learning (AI)
Trial phase	1. Trial coding of each entry. Coders independently determine the number of thematic categories and emotions in reference to NASK typology of "Russian influence matrices"		HC
	2. Comparing results. Identifying potential errors and differences.		
	3. Creating a training dataset (221 tweets) for supervised machine learning. The patterns were based on coders' consistent labeling results. Performing AI labelling on all tweets using Monkey Learn service (https://monkeylearn.com).		AI
	4. Analysing results obtained from machine coding and comparing them with labeling schemes used by human coders.		HC
	5. Establishing uniform manual coding patterns.		
Final phase	1. Manual coding using 14 thematic categories (plus "other" or "excluded") and 7 emotion categories (plus "undefined").		HC
	2. Comparing the results of both coders.		
	3. Discussing inconsistencies and reaching agreement on final codes for them. The researcher had the deciding vote.		
Common rules	1. Each tweet was classified into only one category considered dominant or most obvious.		
	2. Coders were allowed to search for additional tweets using referred hashtags or key terms to widen the interpreting context. Short tweets often did not contain enough information to accurately classify the content.		

Fig. 2. Coding scheme

Results: themes and emotions

Within the discourse related to Poland, 15 thematic categories were distinguished, of which five were dominant (Table 3.):

1. "Refugees from the Middle East and Africa" (17.8% of analysed tweets with 99.0% negative emotions) who are supposed to represent a hostile religion and culture – a threat to the foundations of Polish society. This theme has often served as a tool for criticizing Western European states and building discord between Poland and its allies.
2. "Poland's domestic affairs" (17.5% of entries with 59.7% negative emotions and only 10.0% positive). Almost half of these tweets were intended to cause anxiety and consternation. Increased criticism of Polish authorities emerged in the second half of 2015 and continued through 2016, correlating with the change of government (Fig. 3.). Users often commented on protests by the opposition and feminist circles. Ambivalent information of an economic nature also appeared. Overall, those entries created an image of Poland as an unstable, internally divisive and poorly managed country.
3. "Polish–US relations" (16.5% of tweets with 28.6% negative and 48.9% positive emotions). The volume of messages in this category was low for most of the time, with an average of 7.5 tweets per quarter, w/o excluded (Fig. 4.). Their tone was rather unfavourable for Poland until the end of 2016. Around July 2017 (President Trump's visit to Poland), there was a sharp increase in the number of posts with positive emotions dominating. The increase in popularity of this thread was largely due to the specifics of the database, which was mostly supposed to consist of alleged US-based internet users.
4. "Relations with European allies" (12.4% of tweets with 88.9% negative and 4.0% positive emotions). Most of the tweets in this group were intended to create distrust of Poland or its allies in Europe. Individual states were presented as unreliable partners in international politics, which created an atmosphere of discouragement and distrust for further relations, also at the interpersonal level.
5. "Polish-Russian relations" (6.8% of tweets with 66.0% negative and 28.5% positive emotions). In general, the content and its repercussions varied strongly depending on the time of publication. There was limited scope for interpretation of this category due to the discourse being stretched over several years and the relatively small number of tweets per quarter within the studied sample. Nevertheless, it was possible to distinguish

Table 3. Distribution of main themes and emotions of tweets

Main theme			Dominant emotions/character							
			positive		negative				other	
			fondness friendship admiration	satisfaction praise	hostility antipathy fear	distrust disappointment dishonesty	anxiety consternation worry	dissonance puzzlement	neutral informative	undefined
category	n	%	%							
Refugees (from the Middle East and Africa)	287	17.8		0.3	87.5	3.8	6.3	1.4		0.7
Poland's domestic affairs	281	17.5	0.7	9.3	0.7	2.5	44.8	11.7	7.8	22.4
Polish-US relations	266	16.5	37.6	11.3	0.4	4.5	9.4	14.3	4.5	18.0
Relations with European allies	199	12.4	2.5	1.5	12.1	46.2	21.6	9.0	3.0	4.0
Polish-Russian relations	109	6.8	19.3	9.2	18.3	13.8	28.4	5.5	0.9	4.6
Attitude towards Poles	67	4.2	13.4	6.0	56.7	6.0	4.5	4.5		9.0
US Army in Poland	64	4.0		9.4		7.8	15.6	18.8	21.9	26.6
World War II	45	2.8			2.2		13.3	37.8	17.8	28.9
Western media	31	1.9			3.2	64.5	22.6	9.7		
Poland's energy issues	17	1.1		5.9	5.9	5.9	64.7	5.9	5.9	5.9
Christian / Slavic culture	15	0.9	40.0	46.7			6.7			6.7
Other	13	0.8		7.7			7.7	7.7	7.7	69.2
Polish-Ukrainian relations	13	0.8	15.4		23.1	15.4		30.8		15.4
Relations with the UK	12	0.7	16.7			8.3	66.7			8.3
Polish-Israeli relations	11	0.7		9.1		54.5	9.1	18.2		9.1
EXCLUDED	179	11.1								
TOTAL / MEAN	1609	100.0	10.3	6.3	23.9	12.3	20.3	9.9	4.5	12.4

two characteristic periods: 2009–2012 and 2016–2017 (Fig. 5). In the first, entries on bilateral relations were probably marked by the Smolensk air disaster in March 2010. Immediately after this event, a warming up of messages from the Russian side was evident, dominated by sympathy towards Poland with the hope of improving the image of the Eastern superpower. The second characteristic period began shortly after the Law and Justice party, sceptical towards Russia, came into power in Poland. It may be significant that in 2016–2017 there was not a single positive message in the context of relations between the two countries.

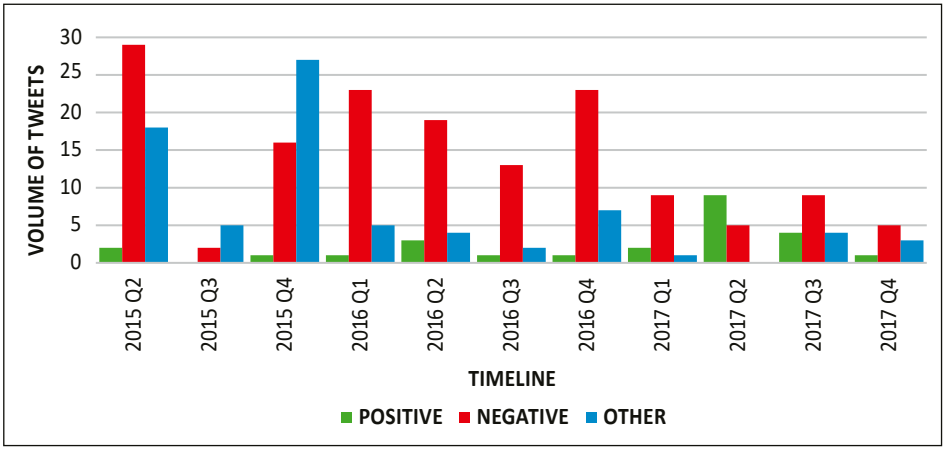

Fig. 3. Distribution of dominant emotions in tweets labeled as "Poland's domestic affairs"
Note: periods with marginal volume or no tweets were omitted

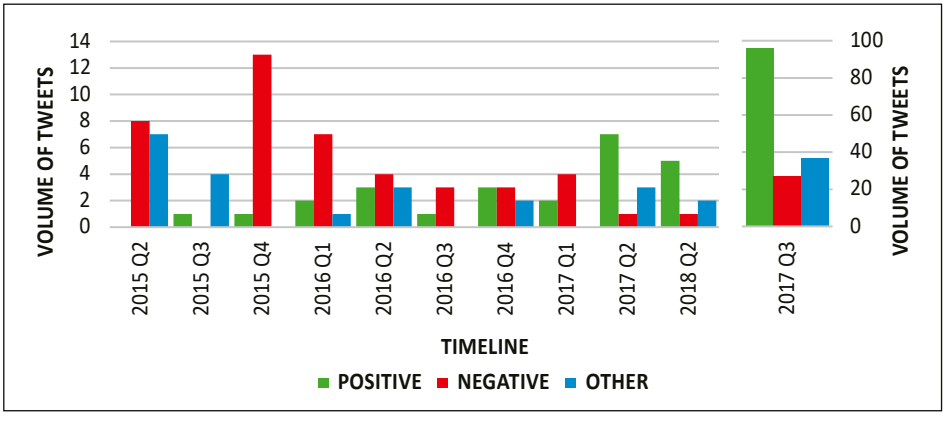

Fig. 4. Distribution of dominant emotions in tweets labeled as "Polish-US relations"
Note: periods with marginal volume or no tweets were omitted)

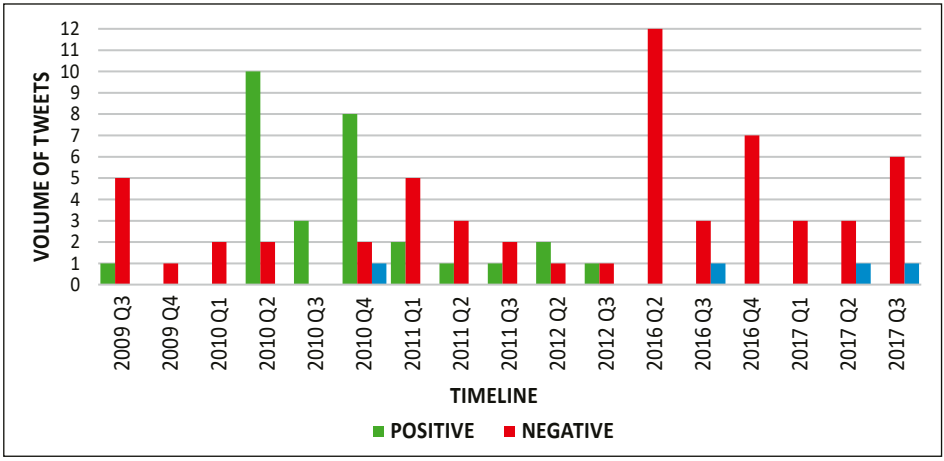

Fig. 5. Distribution of emotions in tweets labeled as "Polish-Russian relations
Note: periods with marginal volume or no tweets were omitted

Results: popular hashtags

Distribution of most frequently used hashtags did not show significant communication patterns (Table 4). Only the widespread use of simple, generic tags is noteworthy. Their presence could facilitate the dissemination of content within the inconspicuously neutral discourses determined by the indicated hashtags. It is likely that including data from outside the studied sample would have produced better results within this approach.

Table 4. Top 10 hashtags

hashtags	n	%
#poland	176	11.40
#news	167	10.82
#world	133	8.61
#russia	39	2.53
#politics	28	1.81
#life	24	1.55
#business	23	1.49
#warsaw	21	1.36
#maga	18	1.17
#environment	17	1.10
TOTAL	1544	100.00
Unique hashtags	460	–
Hashtags per tweet	0.96	–

Results: top tweeters

An in-depth analysis of the eight most frequently tweeting users, who represented over 45% of discourse volume, showed significant behavior patterns (Table 5). Each profile conducted a multi-themed discourse, focusing on one or two major and a few minor topics related to Poland. At the same time, they generated heated discussions concerning non-Polish issues (e.g. domestic US or UK problems), which was not covered by the study. Each of the analyzed themes was commented on similarly over a fixed period, with accompanying emotions being almost unequivocal. In the case of content referring to Polish-US and Polish-Russian relations, tweet repercussions varied strongly depending on the time of publication, as detailed in section 5.1. Many accounts were created shortly (a few weeks or months) before the analyzed activity. The increased generation of entries studied lasted from several months to two years. Moreover, two of these accounts (2nd and 3rd user) had twinned characters: they were founded on the same day, produced similar content, were active at the same period and had an analogous network of followers and following.

Table 5. Top users' activity

	Top user			
	1st	**2nd**	**3rd**	**4th**
tweets (w/o excluded)	130	113	105	93
main theme (emotions)	48% PL domestic affairs (pos. 0%, neg. 60%) 22% relations with European allies (pos. 0%, neg. 86%)	48% PL-US relations (pos. 65%, neg. 11%) 22% refugees (pos. 0%, neg. 100%)	40% PL-US relations (pos. 71%, neg. 10%) 31% refugees (pos. 0%, neg. 100%)	61% PL-Russian relations (pos. 51%, neg. 46%) next theme >10%
dominant time of activity	2015Q2 – 2016Q4 (100%)	2017Q1 – 2017Q4 (100%)	2017Q1 – 2017Q4 (100%)	2009Q3 – 2011Q3 (70%)
profile creation date	2014-12-27	2016-06-15	2016-06-15	2009-07-02
followers	13358	2748	2718	11542
following	13851	265	264	184

	Top user			
	5th	6th	7th	8th
tweets (w/o excluded)	74	62	60	36
main theme (emotions)	62% refugees (pos. 0%, neg. 100%) next theme >15%	58% refugees (pos. 0%, neg. 100%) next theme >12%	58% refugees (pos. 3%, neg. 93%) next theme >15%	31% PL domestic affairs (pos. 9%, neg. 73%) 28% PL-US relations (pos. 10%, neg. 50%)
dominant time of activity	2017Q2 – 2017Q3 (99%)	2015Q3 – 2017Q1 (97%)	2018Q1 – 2018Q3 (98%)	2015Q3 – 2016Q4 (94%)
profile creation date	2017-03-25	2015-01-15	2015-08-13	2014-12-30
followers	7524	8654	4960	12357
following	4500	1446	4396	8501

Discussion

The study provided a closer look at the structure of the disinformation discourse directed against Poland and conducted by entities linked to the Russian Federation. The subject of analysis was a Polish-related sample of English-language tweets made available under Twitter security policy.

Topic modeling allowed the identification of the most popular themes: refugees from the Middle East and Africa (17.8%), Poland's domestic affairs (17.5%), relations with the US (16.5%), with European allies (12.4%) and with Russia (6.8%). Research proved that all types of disinformation themes identified in the literature were present in relation to Poland before the current war in Ukraine. In addition, there were also new issues that could not be classified within the previously known categories. These included content concerning general attitudes towards Poles (4.2% of valid entries) and criticism of Western media (1.9%). At the same time, several weighty and controversial themes appeared only rarely, i.e. Poland's energy issues and Polish-Ukrainian and Polish-Israeli relations. However, it became clear that specific topics were raised to achieve a certain impact on public opinion. Their intensity and meaning were adapted to the current political situation and needs.

The sentiment analysis showed that the broadcast was mostly based on undermining public trust, stoking social tensions and creating an atmosphere of anxiety towards EU and NATO members as well as in interpersonal relations. Surprisingly dangerous was the ease with which the messages and accompanying sentiments moved from the institutional to the personal level, aiming at generalized representatives of ethnic or social groups (e.g. refugees, Polish immigrant workers, Western journalists). Interestingly, there was not a single mention of racial disparities in the sample. Overall, negative emotions occurred in 66.4% of tweets from the sample, and those clearly positive appeared only in 16.6%.

The false or biased content studied was potentially able to influence the social norms, values, goals and interests, the structure of communication and the identity of the genuine users on Twitter. At least two types of actors could be affected by this mechanism: people who were more or less passive recipients (e.g. Americans) and communities who were direct targets (Poles). Russian information operations on Twitter in the context of Poland could create a sense of distrust and threat within Polish society and in Poland's relations with other states, institutions or communities. The consequences could be partial social, political and economic destabilization. Moreover, the tweets examined could have a negative impact on the image of Poland and its citizens in the world.

The analysis of the most frequently used hashtags did not show significant communication patterns, except for one – the widespread use of simple, generic tags. They could be used to disseminate desired content within neutral discourses.

The study also illustrated the functioning of some disinformation mechanisms in practice. Analysis of most frequently tweeting users, representing 45% of the sample's volume (i.e. the most active instances of the vocal minority), showed that each of them conducted a well-managed multi-themed discourse that followed a preconceived strategy. The dynamics of their activities also had many common features, such as being created several weeks or months before the attack and raising specific issues in tweets with similar intensity and emotions.

Conclusions

The study was an attempt to use simple data mining techniques to analyze information warfare on Twitter. It confirmed previous assumptions and the main observations of the scientific community related to the state-driven Russian disinformation activity. However, it did not explore the effectiveness of the impact

of these communications on the audience, nor did it explore complex psychological mechanisms of indoctrination, propaganda or disinformation. A general explanation of the influence of discourse was given with reference to basic sociological knowledge.

The database created for the analysis was small compared to the source material, but it represented only the content that directly referred to Poland and Poles. The consequence of its size was the limited interpretability when the calculations were performed on increasingly smaller subgroups of data. Furthermore, the representativeness of the findings can be a contentious matter, due to the research being based on a snippet of Twitter communication as seen through the eyes of English-speaking audiences.

REFERENCES

2018 Code of Practice on Disinformation. (2022). Retrieved from https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation

Adamik-Szysiak, M. (2014). Twitter in Communication Strategies of the Leaders of the Polish Political Parties. Kwartalnik Naukowy OAP UW 'e-Politikon', 9, 109–131.

Ajir, M., & Vailliant, B. (2018). Russian Information Warfare: Implications for Deterrence Theory. Strategic Studies Quarterly, 12(3), 70–89.

Assessing Russian Activities and Intentions in Recent US Elections. (2017). USA Office of the Director of National Intelligence. Retrieved from https://www.dni.gov/files/documents/ICA_2017_01.pdf

Asur, S., & Huberman, B. A. (2013). Predicting the Future with Social Media. Applied Energy, 112, 1536–1543. https://doi.org/10.1016/j.apenergy.2013.03.027

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 115(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bail, C. A., Guay, B., Maloney, E., Combs, A., Hillygus, D. S., Merhout, F., Freelon, D., & Volfovsky, A. (2020). Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. Proceedings of the National Academy of Sciences, *117*(1), 243–250. https://doi.org/10.1073/pnas.1906420116

Bastos, M., & Farkas, J. (2019). "Donald Trump Is My President!": The Internet Research Agency Propaganda Machine. Social Media + Society, *5*, 205630511986546. https://doi.org/10.1177/2056305119865466

Batorowska, H., Klepka, R., & Wasiuta, O. (2019). Media jako instrument wpływu informacyjnego i manipulacji społeczeństwem (1st ed.). Wydawnictwo 'Libron'.

Belford, M., Greene, D., & Cross, J. P. (2016). Tweeting Europe: A text-analytic approach to unveiling the content of political actors' Twitter activities in the European Parliament. 6th Annual General Conference of the European Political Science Association (EPSA'16), 44.

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

Bielawski, R., & Ziółkowska, A. (2018). Media społecznościowe, a kształtowanie bezpieczeństwa państwa. Wydawnictwo Naukowe Tygiel.

Bodine-Baron, E., Helmus, T., Radin, A., & Treyger, E. (2018). Countering Russian Social Media Influence. RAND Corporation. https://doi.org/10.7249/RR2740

Chandio, M. M., & Sah, M. (2020). Brexit Twitter Sentiment Analysis: Changing Opinions About Brexit and UK Politicians. In L. C. Jain, S.-L. Peng, B. Alhadidi, & S. Pal (Eds.), Intelligent Computing Paradigm and Cutting-edge Technologies (Vol. 9, pp. 1–11). Springer International Publishing. https://doi.org/10.1007/978-3-030-38501-9_1

Chauhan, P., Sharma, N., & Sikka, G. (2021). The emergence of social media data and sentiment analysis in election prediction. Journal of Ambient Intelligence and Humanized Computing, 12, 2601–2627. https://doi.org/10.1007/s12652-020-02423-y

Chitra, U., & Musco, C. (2019). Understanding Filter Bubbles and Polarization in Social Networks. arXiv:1906.08772 [Physics]. Retrieved from http://arxiv.org/abs/1906.08772

Colliver, C., Pomerantsev, P., Applebaum, A., & Birdwell, J. (2018). Smearing Sweden. International Influence Campaigns in the 2018 Swedish Election. Retrieved from https://www.lse.ac.uk/iga/assets/documents/arena/2018/Sweden-Report-October-2018.pdf

Conger, K. (2019, August 19). Facebook and Twitter Say China Is Spreading Disinformation in Hong Kong. The New York Times. Retrieved from https://www.nytimes.com/2019/08/19/technology/hong-kong-protests-china-disinformation-facebook-twitter.html

Deho, O. B., Agangiba, W. A., Aryeh, F. L., & Ansah, J. A. (2018). Sentiment Analysis with Word Embedding. 2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST), 1–4. https://doi.org/10.1109/ICASTECH.2018.8506717

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., & Matney, R. (2019). The Tactics & Tropes of the Internet Research Agency. New Knowledge. Retrieved from http://arks.princeton.edu/ark:/88435/dsp01fb494c31z

Dolan, C. J. (2022). Hybrid Warfare in the Western Balkans: How Structural Vulnerability Attracts Maligned Powers and Hostile Influence. SEEU Review, 17(1), 3–25. https://doi.org/10.2478/seeur-2022-0018

European Commission. (2018). *Action Plan against Disinformation*. Retrieved from https://commission.europa.eu/publications/action-plan-disinformation-commission-contribution-european-council-13-14-december-2018_en

Furman, I., & Tunç, A. (2019). The End of the Habermassian Ideal? Political Communication on Twitter During the 2017 Turkish Constitutional Referendum. *Policy & Internet*, *12*. https://doi.org/10.1002/poi3.218

Gac, M. (2020). Zagrożenia płynące ze strony Federacji Rosyjskiej dla państw regionu Morza Bałtyckiego. *De Securitate et Defensione. O Bezpieczeństwie i Obronności*, *7*(2). https://doi.org/10.34739/dsd.2020.02.06

Gadde, V., & Roth, Y. (2018). Enabling further research of information operations on Twitter. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html

Giles, K. (2015). Handbook of Russian Information Warfare. NATO Defense College.

Gorwa, R. (2017). Computational Propaganda in Poland: False Amplifiers and the Digital Public Sphere. University of Oxford.

Gruzd, A., & Mai, P. (2020). Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. Big Data & Society, 7, 205395172093840. https://doi.org/10.1177/2053951720938405

Grzywińska, I., & Batorski, D. (2016). How the Emergence of Social Networking Sites Challenges Agenda-setting Theory. Konteksty Społeczne, 4(1 (7)), 19–32.

Guess, A. M., & Lyons, B. A. (2020). Misinformation, Disinformation, and Online Propaganda. In N. Persily & J. A. Tucker (Eds.), Social Media and Democracy (1st ed., pp. 10–33). Cambridge University Press. https://doi.org/10.1017/9781108890960.003

Jain, V. K., & Kumar, S. (2017). Towards Prediction of Election Outcomes Using Social Media. International Journal of Intelligent Systems and Applications, 9(12), 20–28. https://doi.org/10.5815/ijisa.2017.12.03

Johnson, K., Jin, D., & Goldwasser, D. (2017). Modeling of Political Discourse Framing on Twitter. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 556–559.

Keller, F., Schoch, D., Stier, S., & Yang, J. (2017). How to Manipulate Social Media. Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), 5. Retrieved from https://www.researchgate.net/publication/317290047_How_to_Manipulate_Social_Media_Analyzing_Political_Astroturfing_Using_Ground_Truth_Data_from_South_Korea

Keller, F., Schoch, D., Stier, S., & Yang, J. (2019). Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/a5gk6

Kharde, V. A., & Sonawane, S. S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. International Journal of Computer Applications, 139(11), 5–15. https://doi.org/10.5120/ijca2016908625

Kisiołek, A. (2018). Analiza wpisów na portalu Twitter z wykorzystaniem narzędzi big data zawartych w pakiecie R. Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, 362, 306–317.

Kmiecik, P. (2019). Bezpieczeństwo informacyjne Rzeczypospolitej w dobie Fake News—Przykłady wykorzystania mediów cyfrowych w szerzeniu dezinformacji. Bezpieczeństwo Obronność Socjologia, 11/12, 82–102.

Krajobraz bezpieczeństwa polskiego internetu raport roczny 2019. (2020). NASK PIB/CERT.

Krasodomski-Jones, A., Miller, C., Bartlett, J., Smith, J., Chauvet, A., & Jones, E. (2018). Russian Influence Operations on Twitter. DEMOS. Retrieved from https://demos.co.uk/project/russian-influence-operations-on-twitter/

Kreft, J. (2019). Władza algorytmów: U źródeł potęgi Google i Facebooka. Wydawnictwo Uniwersytetu Jagiellonskiego.

Lanoszka, A. (2019). Disinformation in International Politics. European Journal of International Security, 4(2), 1–22. https://doi.org/10.1017/eis.2019.6

Lelonek, A. (2018). Wojna informacyjna, operacje informacyjne i psychologiczne—Pojęcia, metody i zastosowanie. Fundacja «Centrum Badań Polska-Ukraina», 24.

Lock, I., & Ludolph, R. (2019). Organizational propaganda on the Internet: A systematic review. Public Relations Inquiry, 9, 2046147X1987084. https://doi.org/10.1177/2046147X19870844

Michalak, J. (2017). Social media jako kopalnia informacji—Wybrane obszary wykorzystania danych na przykładzie portalu Twitter. Acta Universitatis Nicolai Copernici Zarządzanie, 44(3), 107. https://doi.org/10.12775/AUNC_ZARZ.2017.040

Moderation Research—Twitter Transparency Center. (n.d.). Retrieved from https://transparency.twitter.com/en/reports/moderation-research.html

Mueller, R. S. (2019). Report on the Investigation into Russian Interference in the 2016 Presidential Election Vol. I. U.S. Department of Justice.

Rogers, R., & Niederer, S. (Eds.). (2020). The Politics of Social Media Manipulation. Amsterdam University Press. https://doi.org/10.2307/j.ctv1b0fvs5

Roth, Y. (2019, Jun 13). Information operations on Twitter: Principles, process, and disclosure. Retrieved from https://blog.x.com/en_us/topics/company/2019/information-ops-on-twitter

Roth, Y., & Gadde, V. (2022, Aug 24). Expanding access beyond information operations. Retrieved from https://blog.x.com/en_us/topics/company/2021/-expanding-access-beyond-information-operations-

Russia Report (HC 632). (2020). Intelligence and Security Committee of UK Parliament.

Smith, J. (2019). The outrage election. A CASM investigation conducted with BBC Click. DEMOS.

Tankovska, H. (2021). Countries with the most Twitter users 2021. Retrieved from https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Unver, A. (2019). Russian Disinformation Ecosystem in Turkey (SSRN Scholarly Paper 3534770). Retrieved from https://papers.ssrn.com/abstract=3534770

Update: Russian interference in the 2016 US presidential election. (2017, Sep 28). Retrieved from https://blog.x.com/en_us/topics/company/2017/Update-Russian-Interference-in-2016--Election-Bots-and-Misinformation

von Nordheim, G., Boczek, K., & Koppers, L. (2018). Sourcing the Sources: An analysis of the use of Twitter and Facebook as a journalistic source over 10 years in The New York Times, The Guardian, and Süddeutsche Zeitung. Digital Journalism, 6(7), 807–828. https://doi.org/10.1080/21670811.2018.1490658

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking (Vol. 27). Council of Europe Strasbourg. Retrieved from http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf

Watanabe, K. (2018). Conspiracist propaganda: How Russia promotes anti-establishment sentiment online? Retrieved from https://blog.koheiw.net/?page_id=59

Wilson, C. (2022). Under the Radar: Analyzing Recent Twitter Information Operations to Improve Detection and Removal of Malicious Actors, Part 1 (SSRN Scholarly Paper 4389821). https://doi.org/10.2139/ssrn.4389821

Wrzosek, M. (Ed.). (2019). Zjawisko dezinformacji w dobie rewolucji cyfrowej: Państwo, społeczeństwo, polityka, biznes. NASK Państwowy Instytut Badawczy. retrieved from https://cyberpolicy.nask.pl/raport-zjawisko-dezinformacji-w-dobie-rewolucji-cyfrowej-panstwo-spoleczenstwo-polityka-biznes/

Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019). Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. https://doi.org/10.1145/3308560.3316495

Zannettou, S., Caulfield, T., Setzer, W., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019). Who Let The Trolls Out?: Towards Understanding State-Sponsored Trolls. https://doi.org/10.1145/3292522.3326016