Mariacristina Gallo, Giuseppe Fenza, Daniele Battista

Information Disorder: What about global security implications?

INFORMATION DISORDER: WHAT ABOUT GLOBAL SECURITY IMPLICATIONS?

Information sharing has become easy and fast in the digital era, where everyone can be connected. In addition, social networks spread rapidly, such that around 70% of the population has an active account. People use social media to communicate and also as information sources. As a result, traditional media has lost its role as a primary source of information; consequently, circulating news is not appropriately verified. On the other side, governments and organized non-state actors could exploit distorted information on national or foreign political sentiment, for example, to achieve strategic or geopolitical outcomes. It follows that the user is reached by a quantity of information that is often untrustworthy, which, guided by prejudices or emotions, could be rapidly shared. First, this paper introduces the most important aspects related to information disorder. Then, it proposes an interdisciplinary framework that, by integrating solutions from different areas, aims to combine educational objectives (in terms of Media literacy) with practical tips and Artificial intelligence-based tools. In summary, the aim is to give useful recommendations to involved actors, such as information agencies, policymakers, regulators, and technology companies, in terms of their possible contribution to the problem.

KEYWORDS *Information Disorder, Misinformation, Disinformation, Netwar, Cybersecurity.*

1. Information warfare

The spread of the Internet worldwide, combined with the use of smart devices with fast WiFi connections and extensive people's participation in social networks, caused traditional media to lose its role as a primary source of in-

Mariacristina Gallo, Dipartimento di Scienze Aziendali - Management & Innovation Systems, University of Salerno – 84084 Fisciano, email: mgallo@unisa.it, orcid: 0000-0002-5474-2697.

Giuseppe Fenza, Dipartimento di Scienze Aziendali - Management & Innovation Systems, University of Salerno – 84084 Fisciano, email: gfenza@unisa.it, orcid: 0000-0002-4736-0113.

Daniele Battista, Dipartimento di Studi Politici e Sociali, University of Salerno – 84084 Fisciano, email: dbattista@unisa.it.

ISSN 2785-0072

formation. Reports[1][2] demonstrate that last year, in Italy, 84% of the population was connected every day for, on average, 6 hours, and 68% were active users on social media, in line with the European population trend. Moreover, around 50%[3] of European adults declared to use social media as the primary source of information (mainly Facebook).

In this context, governments and organized non-state actors could work towards distorting national or foreign political sentiment, for example, to achieve strategic or geopolitical outcomes. In this sense, communication forms strongly affected military structures, doctrines and strategies, such that the netwar term was coined. Often, the objective is to manipulate public opinion through fake news, disinformation, fake accounts (fake amplifiers), etc. In line with Simon Sinek's assertion, «there are only two ways to influence human behavior: you can manipulate it or you can inspire it» (Sinek 2009), malicious actors leverage misinformation when inspiration is impracticable. Among relevant deceptive activities, there are: imposer websites spreading biased and misleading content; sock puppet accounts posing outrageous memes (for example, on Instagram); click farms manipulating social trends and their referral systems; a mass collection of personal data used to influence voters with tailored messages and advertisements; conspiracy communities trying to trick journalists into dealing with rumors or hoaxes. This phenomenon puts the average user in a disadvantageous situation in terms of information collection and understanding. Moreover, cognitive vulnerabilities, emotions, prejudices, and bias could influence false narrative recognition, and the illegal adoption of troll accounts alter the timing and dissemination process. It follows that misleading content spreads quicker than genuine ones (Lukito 2020).

Why does misleading content spread rapidly?

Misleading content leverages psychological mechanisms, mainly cognitive bias (Bakir and McStay 2018), that are also amply exploited by recommendation algorithms (i.e., News Feed) (Spohr 2017). Although profiling activities improve user experience and engagement, they also facilitate the dissemination of fake content. The cause is the risk of polarization associated with the high consensus concerning content (Chitra and Musco 2020).

Social media users can rarely spend time and energy verifying news. In particular, the human mind, guided by cognitive bias, tends to: (i) select the first suggested search result, (ii) glance only at news headlines, (iii) search for

[1] https://wearesocial.com/it/blog/2021/01/digital-2021-i-dati-globali.
[2] https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021.
[3] https://www.statista.com/statistics/718019/social-media-news-source.

images in posts and tweets to get a quicker idea of their content. Paying little attention to what reading or being reluctant to devote time to analyze it brings to hasty, incomplete, or downright wrong conclusions (Chadwick and Stanyer 2022). Cognitive biases can be considered shortcuts adopted by the human mind to reduce the cognitive load by, for example, framing facts. The most common biases are (Gelfert 2018; Luo *et al.* 2022):

- Truth bias: the inclination to accept all information as accurate.
- Confirmation bias: the inclination to accept only information (or evidence) confirming previous beliefs.

If, on the one hand, the news consumer tends to limit his/her cognitive load, on the other hand, the one who provides information should take advantage of it. The information provider, in fact, can leverage fallacies (i.e., logic breakdowns) to support its point of view, for example, in political campaigns through propaganda. Examples of frequent fallacies are (Musi and Reed 2022):

- Cherry picking: Consider only evidence supporting the proposed thesis, ignoring any that might refute it.
- Ad hominem: Lose focus on the controversial subject by contesting the interlocutor rather than his argument.
- Evading the burden of proof: Defend an argument by asking the interlocutor to prove the contrary.

In the literature, studies about the role of fallacies in influencing opinions are numerous (Bangalee and Bangalee 2021; Zompetti 2019).

What are the most affected subjects?

Misleading content feeds itself with emotional impact arguments or highly polarized topics (Humprecht 2019). In this sense, themes such as immigration, economy, health, and environment are fertile lands. In addition, the context aspects (e.g., specific events, circumstances, etc.) further increase the risk of disinformation. For example, during the first wave of the Covid-19 pandemic in Italy (i.e., from January to May 2020), 6%, on average, of related online news and posts were false or inaccurate (Agcom 2020).

A cross-country survey conducted at the European Union in July 2019 (Hameleers *et al.* 2021) demonstrated that citizens recognize more lies in immigration news, followed by the economy and the environment. They follow international politics, terrorism, welfare, and the Eu with relatively comparable levels of perceived misinformation. In particular, from the involved countries' point of view, Greece and Spain (i.e., the most affected by the Euro crisis) are the most susceptible to economic disinformation. At the same time, immigra-

tion is equally distributed across all countries. Germany, Denmark, Sweden, and the Netherlands mainly recognize environmental disinformation.

From Infodemia to Information disorder

Due to Internet diffusion, the user is reached by a quantity of information that is often unmanageable. The Internet favors the direct relationship between sources and the public, but this implies that the responsibility for determining the reliability of information lies directly with the user, not with the newspaper or the television network. Furthermore, users can instantly generate and share contents that can rapidly spread around the globe. This allows them to (voluntarily or involuntarily) participate in disinformation campaigns that can manipulate people with cognitive vulnerabilities. Disinformation campaigns can distort reality to polarize communities, generating confusion and disorder.

Strictly related to information disorder is the fake news problem. Fake news could refer to false or biased writings and truth omissions, often created and fine-tuned to generate emotional involvement or raise prejudices in order to incite people's reactions. In addition, it could have multiple political and financial goals passing through the so-called «cognitive warfare» (dir Avocat 2021). Fake news spans from classical information distortion (e.g., linking existing images or videos to false news) to the creation of DeepFakes. This latter exploits machine learning-based technology to produce or alter video/image content and present a distorted reality (Westerlund 2019). It follows that information disorder is a wider concept that includes:

1. Disinformation: false content consciously created and shared to cause harm. It is usually made for money or political motivations.
2. Misinformation: false content shared without awareness of its deceptiveness. Socio-psychological factors usually drive its sharing (e.g., people want to affirm their affiliation to a specific religion, race or ethnic group).
3. Malinformation: truthful information shared explicitly to cause disorder or damage someone.

Regarding the first two, we can do a further decomposition (going from low harmful intents to more ones):

1. satire or parody: although satire and parody can be considered art forms, the risk is that they are used strategically to circumvent fact-checkers and to distribute rumors and conspiracies;

2. false connection: a title, image or caption not in line with news content (e.g., click-bait or sensationalist titles);
3. misleading content: misleading use of information related to the Framing theory (Chong and Druckman 2007) in which pieces of evidence (e.g., statistics or images) are manipulated ad-hoc;
4. false context: authentic content (e.g., images) dangerously reformulated out of their original context;
5. imposter content: improper or fraudulent use of content created by others or private company logos;
6. manipulated content: original content managed to create information disorder;
7. fabricated content: completely false content of both textual and visual type.

Finally, although propaganda cannot be considered a cause of information disorders, it extensively adopts dis/mis-information to convey emotional messages and manipulate people's behaviors.

Characteristics

The main actors of mis/dis-information are:

- Agent. The person or group responsible for producing and distributing news. It can be official (e.g., intelligence services, political parties, newspapers) or unofficial (e.g., groups of citizens who have learned about an issue), and work individually or in groups. Motivations guiding agents can be financial (i.e., profiting from the information disorder through advertising), political (i.e., discrediting a political opponent or influencing public opinion), social (i.e., connecting with a particular group online or offline), psychological (i.e., looking for support or consideration).
- Message. The content of the misleading information. We can be interested in understanding shape and characteristics (e.g., mission, target, narrative type, media content, etc.)
- Interpreter. The target receives the message. In particular, we must understand how the interpreter assimilates it and what action he decides to take. Such decisions are related to the interpreter's experiences and socio-cultural status. In particular, the interpreter could become the new agent for the message by, for example, commenting or sharing it.

Phases and supporting technologies

An instance of information disorder undergoes a three-phase process often supported by specific technological tools:

- Creation. The message is created. Web analytics tools can infer information about individuals' preferences, tastes, purchasing behaviors and psychological aspects. Big data and Artificial intelligence allow us to understand circulating hot topics and the target audience.
- (Re)Production. The message is transformed into a multimedia product. In this phase, dedicated software facilitates the automatic generation and manipulation of content (e.g., video, image, etc.).
- Distribution. The message is distributed or made public. Such phase is supported by bot creation and management software, posting system (on social media), etc.

2. Information disorder threats

Socio-cultural aspects, cognitive bias and so on, together with recommendation mechanisms behind research engines and social media, are the leading cause of disinformation spread. Although profiling activities improve user experience and engagement, recommendation systems expose users to ideologically-aligned content, leaving them in a sort of filter bubble. It follows that users mostly communicate with like-minded others due to their continued exposure to only contents in line with their thoughts (Terren and Borge-Bravo 2021), generating the problem of echo chambers.

Some research from different subjects reveal tightly related effects of disinformation on democracy, especially for Western countries. Despite a greater political involvement due to social media discussed by Vaccari and Valeriani (Vaccari and Valeriani 2021), Palano describes the new society as a «bubble democracy» formed by self-referential and potentially polarized small niches (Palano, 2019). In this type of democracy, the political agenda has difficulty identifying a public argument. Often, citizens do not recognize their ideals in classical left or right political parties; instead, they are fragmented into small niches joined by very similar needs. As a result, political campaigns are forced to maximize niche satisfaction by, for example, personalizing proposals and generating the risk of fragmentation (Pariser 2011). On the other hand, the effect of fake news on political campaigns is well known, for example, in opponents' delegitimization (Kuehn and Salter 2020). Well-known incidents regar-

ded Russia's interference in the Us election in 2016 and disinformation attacks against French president Macron during the election campaign in 2017.

Effects of information disorder are also tangible in other aspects of society, related to citizens' life choices. For example, regarding public health threats, a randomized controlled trial in the Uk and the Usa about Covid-19 vaccines demonstrated that, in September 2020, misinformation induced a decline in the intent of 6.2% in the Uk and 6.4% in the Usa among people initially inclined to accept the vaccine (Loomba *et al.* 2021).

In terms of environmental protection and, particularly, climate crisis, disinformation campaigns aim to alter the public conception of the problem, discouraging citizens from supporting mitigation policies (Cook 2020).

Interesting insights also emerged from a recent Ipsos study[4] conducted on over 19.000 people in 27 countries: 48% affirm believed a story after revealed fake; however, 36% think that the term «fake news» is associated with stories that politicians or the media do not agree with. Moreover, 65% of people think other people live in a bubble, while only 34% think the same for themselves; 63% believe they can spot fake news, but only 41% think an average person does. In summary, the study confirms most just-revealed threats: people recognize the disinformation problem as a risk but are not always aware of their limits.

3. Open source and interdisciplinary based information disorder counter-measures

When speaking about fake news detection, the literature proposes many solutions mainly based on Machine and Deep learning classifiers able to distinguish between real and fake content (Manzoor *et al.* 2019). They principally study patterns in fake news of the training set and try to predict if the information could be fake. However, despite the high performance achieved by Artificial intelligence, it cannot prevent the creation of disinformation. It aims to reveal potential mis- and dis-information, but prevention and deterrence can be obtained only by combining it with the regulatory measures and efforts of education systems. Moreover, by definition, Artificial intelligence can fulfill a task when feasible for the human mind; this means we need to start with shared partial solutions.

[4] https://www.ipsos.com/en-uk/fake-news-filter-bubbles-and-post-truth-are-other-peoples-problems.

Many of the most important news agencies (e.g., Afp[5], The Washington Post[6], etc.) are sharing numerous tips aiming to suggest a rigorous fact-checking workflow for news checking. The recommended workflows (specific for each resource type) usually adopt Open source intelligence (Osint) fundamentals.

Open source intelligence is a crucial activity of governments, agencies, and corporations and represents the intelligence produced from publicly available information that is collected, exploited, and disseminated in a timely manner to an appropriate audience to address a specific intelligence requirement (Böhm and Lolagar 2021).

Osint activities, in the context of fake news detection, regard, for example, certifying the ownership of an image or discovering its first sharing. It uses devoted tools that scrape the Web or consult open data to give useful insights about contents.

This paper presents a solution as a multidisciplinary framework that melds results from different disciplines, such as intelligence analysis, data science, Artificial intelligence, social science, and so on, by leveraging open-source data.

As discussed later in this section, information disorder counter-measures firstly rely on more conscious use of new media that, by extending Media literacy, must reinforce awareness and critical thinking. The idea follows from difficulties in increasingly complex societies, defining law regulations when treating freedom of expression that raises a need to complement the regulatory role of law with the contribution of self-regulation (Aznar 2019).

Moreover, through Digital literacy, people should be trained in the use of methodologies and technologies supporting fact-checking and debunking practices. In this sense, new Artificial Intelligence evolutions should assist the overall mental process and become part of more well-structured solutions (e.g., tag-suggestion interfaces, bot account recognition, etc.). Thus, as suggested by recommendations, there is a need for a synergy of multiple solutions, and only one solution does not exist.

4. Media literacy

Persuading citizens to exercise sound skepticism during news reading could avoid a total rejection of information and, thus, a lack of trust in media (Pinkleton *et al.* 2012). Citizens should be able to evaluate information and then trust accurate one. In this sense, adequate Media literacy programs could

[5] https://www.afp.com.
[6] https://www.washingtonpost.com.

educate to evaluate information with a critical sensibility to discern possible deceptions and manipulations before accepting and distributing a message. On the other hand, such skills should also bring citizens to produce a message with a sense of responsibility and respect for themselves and others (Maloy *et al.* 2022).

Human thought gravitates towards prejudice, the need for generalization, common mistakes, self-deception, etc. (Paul and Elder 2019). Media literacy is a form of critical literacy that involves analytical, evaluation and essential skills of reflection. Particularly relevant is to strengthen critical and logical thinking. Critical thinking refers to essential abilities to:

- identify, analyze and effectively evaluate topics and statements of truth;
- discover and overcome personal preconceptions and prejudices;
- formulate and present compelling reasons to support conclusions;
- make sensible and intelligent decisions about what to believe and what to do.

In this sense, a critical thinker is able to identify narrative fallacies (and so, propaganda attempts) in people's arguments (Olariu 2022). It follows that a critical thinker is less susceptible to disinformation. The United nations educational, scientific and cultural organization (Unesco) defines a functional illiterate as a person unable to interpret texts and messages correctly, despite his/her educational level (Unesco 1978). He/She uncritically accepts elementary explanations in line with his prejudices, stereotypes, cognitive bias, and the opinion of friends (often in the social network), giving up verifying sources. From a global population survey, a severe percentage (i.e., about 28%) of Italians (between 16 and 65 years old) are functional illiterate, and about 25% have inadequate digital skills. It is the worst European score after Turkey (i.e., 46%), matched only by Spain (Oecd 2019).

Fact-checking workflow

In terms of information reliability verification, the fact-checking workflow leveraging open-source data is amply adopted by expert communities. In particular, beyond the checking to do, five pillars of verification are suggested (Urbani 2019):

- Check the provenance: is the examined one the original account, article or piece of content?
- Check the source: who created the account or article or captured the original piece of content?
- Check the date: when was the content created?

- Check the location: where was the account established, the website created, or the piece of content captured?
- Check the motivation: what causing the account establishment, the website created or the piece of content captured?

The more information is collected about each pillar, the stronger will be the resulting verification. The following sections give details about each aspect.

Provenance

The objective is to find the examined content in its original form. This way, it should be easier to understand who posted it, when, where, and why. The extracted insight could also be adopted to classify the content (Krishnan and Chen 2018).

Suggested techniques regard:

- Reverse image search. It consists in searching the same or similar content (image, in this case) in a large database (e.g., Google images). When an older version of the same image is found, we have evidence of misleading content.
- Reverse video search. Regarding videos, a frame is usually captured, and a reverse image search is made on it, as for images.
- Anonymous spaces research. Searching for media in more closed and anonymous spaces could be helpful when original content is not easily found with other described techniques. Examples are Reddit[7], 4chan[8], Discord[9], and, where possible, Twitter and Facebook.

Source

Everyone can re-post a piece of content written or captured by someone else on the Internet. Therefore, identifying the content «owners» can facilitate the verification process. Ones the first uploader is found, we should understand if the content is coherent with the authors' geographic position, other shared contents, its level of credibility (Sitaula *et al.* 2020), and so on. In particular, it could be interesting to investigate authors' social accounts, make reverse image searches of account images, search for shared posts in Google to understand if there are embedded contents, and so forth. In addition, checking if declared email addresses are associated with any user (for example, through Skype) could help determine the source's credibility. For example, it could be

[7] https://camas.github.io/reddit-search.
[8] https://4chansearch.com.
[9] https://disboard.org/search.

an automated account (i.e., a bot). In this sense, specific tools can be adopted, or paying attention to the number of daily posts and if there is a silence period associable with the night rest.

Date

Due to the diffusion of smartphones and the continuous connections of users, it usually assists in the quick share of pieces of content. However, we cannot assume facts are always accurate: the objective of this step consists in identifying the time in which the piece of content has been captured. The first suggestion, in this sense, regards accessing file metadata. For example, we can refer to the Exif (Exchangeable image file format) data for image files.

Location

Location identification has a similar issue related to date reconstruction: geo-tags could be wrong with respect to the location reported in the piece of content. For this purpose, it could be helpful to identify some details in the image or video and research through satellite images. Examples are detecting banners, signs, flags, squares, etc., to try associating a place with the image or video. In addition, spoken language and clothing could further help in identification. In this sense, it needs to pay close attention to the level of update of images and the latest events in the area (e.g., extreme weather conditions or war) that could drastically change the landscape.

Artificial intelligence solutions

Artificial intelligence has reached a powerful level of adoption in many areas. When discussing information disorder, the main solution consists of breaking down the problem into smaller sub-problems to be solved separately (Rubin 2022). Following this concept, the literature is rich in proposals that try to resolve specific disinformation aspects. In particular, one of the main objectives regards distinguishing between fake and trustworthy content (Reis *et al.* 2019) or classifying the source of information spread (Shahid *et al.* 2022). They consist of a binary classification problem aiming, for example, to detect deception, clickbait, satire, rumor, bots, etc.

Algorithms for deceptive detection assume that it is possible to identify imposters' speech through some linguistic signals. For example, the level of details (i.e., places, times, and descriptions of people and objects) is indicative of the level of truthfulness; conversely, scarce use of personal pronouns (e.g., «the house» instead of «our house»), negations and perceptual information, is a

symptom of deception (Rubin 2022). When working with social media posts, it could be helpful to consider footprints consisting of old messages and other networking activities. The objective is to identify contradictions or propensity to spread rumors.

Particularly relevant are computational fact-checking systems that reproduce humans in investigative reporting. It involves many other sub-tasks:

- Identifying suspected misleading claims, which is worth checking.
- Collecting evidence and identifying incorrect information in questionable content.
- Assessing source credibility.
- Deciding about the veracity of a statement.
- Provide motivations for the decision.
- Fact-checking can adopt Osint approaches to double-check the information on multiple sources or evaluate, in the case of social media, potential skepticism concerning a news article (Pérez-Rosas and Kleinberg 2017).

Recommendations

The Council of Europe (Wardle and Derakhshan, 2017) draws up a list of recommendations directed to technology companies, national governments, media organizations, civil society, education ministries, and Grant-making foundations. The most important suggestions are reported for each target as follows:

- Technology companies. Collaborate; foster sharing of data with researchers; give context and metadata information to help users in evaluating contents; avoid profit for purveyors; detect and repress automated accounts; invest in other languages (i.e., in addition to English); focus on multimedia contents; give more access to metadata; provide fact-checking and verification applications; research engines should make emerge trusted contents; work towards the minimization of filter bubbles impact.
- National governments. Commission research studies to understand tendencies in information disorder (e.g., topics, platforms, countries, etc.); avoid ads on fake websites; supervise Facebook ads; support local news agencies; incentivize advanced cyber-security training.
- Media organizations. Collaborate to avoid resource waste for investigating the same contents; debunk sources and contents; give information about conducted debunking process; train about infor-

mation disorder threats; improve the quality of headlines; do not disseminate fabricated content.
- Civil society. Inform the public about information disorder threats.
- Education ministries. Publicize digital and media literacy as well as forensic verification techniques.
- Grant-making foundations. Incentivize challenging research proposals; provide funding for smaller startups; support journalistic initiatives aiming to teach fact-checking and verification skills.

5. Conclusions

The emergent question of information disorder derives from (and directly affects) different aspects of modern societies. First, although fake news was known in the past, the diffusion of technologies and the Internet exacerbates it. Second, through modern devices, citizens take the place of official media producing and sharing (voluntarily or not) not conveniently verified content. Then, people's cognitive biases or vulnerabilities facilitate its spread. In fact, when conveyed with malicious intent, mis- and dis-information usually concern sensitive topics that immediately affect the reader's emotions, encouraging him/her to spread them further. Mentioned situation feeds citizens' polarization and undermines democracy.

Regarding disinformation counter-measures, the literature ranges from fact-checking (semi-)automatic methodologies based, for example, on Artificial intelligence, to approaches giving more awareness about the problem and available solutions. In particular, this paper presents a multidisciplinary approach, starting from official recommendations. It suggests sensitizing citizens and improving their critical thinking through Media literacy. Moreover, the proposal involves orchestrating available next-generation tools for improving final results. Since the Internet (and often social media) has become the primary source of information in Western society, search engines or recommender systems at the base of social media should consider the trustworthiness of contents (and their sources) before suggesting them.

References

Agcom (2020), *Osservatorio sulla disinformazione online. Speciale Coronavirus*, 3, https://www.agcom.it/osservatorio-sulla-disinformazione-online (last accessed on 27th December 2022).

Aznar, H. (2019), *Information Disorder and Self-Regulation in Europe: A Broader Non-Economistic Conception of Self-Regulation*, in «Social Sciences», 8(10), pp. 280-294.

Bakir, V. and Mcstay, A. (2018), *Fake News and the Economy of Emotions: Problems, Causes, Solutions*, in «Digital journalism», 6(2), pp. 154-175.

Bangalee, A. and Bangalee, V. (2021), *Fake News and Fallacies: Exploring Vaccine Hesitancy in South Africa*, in «South African Family Practice», 63(1), pp. 5345-5347.

Böhm, I. and Lolagar, S. (2021), *Open Source Intelligence: Introduction, Legal, and Ethical Considerations*, in «International Cybersecurity Law Review», 2(2), pp. 317-337.

Chadwick, A. and Stanyer, J. (2022), *Deception as a Bridging Concept in the Study of Disinformation, Misinformation, and Misperceptions: Toward a Holistic Framework*, in «Communication Theory», 32(1), pp. 1-24.

Chitra, U. and Musco, C. (2020), *Analyzing the Impact of Filter Bubbles on Social Network Polarization, paper presented at 13th International Conference on Web Search and Data Mining, Houston*, Texas, Usa, 3-7 February.

Chong, D. and Druckman, J.N. (2007), *Framing Theory*, in «Annual Review of Political Science», 10(1), pp. 103-126.

Cook, J. (2020), *Deconstructing Climate Science Denial*, in D. C. Holmes and L. M. Richardson (eds), *Research Handbook on Communicating Climate Change*, Uk, Edward Elgar Publishing, pp. 62-78.

Dit Avocat, A. B. (2021), *Cognitive Warfare: The Battlefield of Tomorrow?*, in A.B. Dit Avocat, A. Haxhixhemajli and M. Andruch (eds), *New Technologies, Future Conflicts, and Arms Control*, Prague, Center for Security Analyses and Prevention, pp. 60-65.

Gelfert, A. (2018), *Fake News: A Definition*, in «Informal Logic», 38(1), pp. 84-117.

Hameleers, M., Brosius, A. and De Vreese, C. H. (2021), *Where's the Fake News at? European News Consumers' Perceptions of Misinformation across Information Sources and Topics*, in «Harvard Kennedy School Misinformation Review», 2(3), pp. 1-10.

Humprecht, E. (2019), *Where «Fake News» Flourishes: a Comparison across Four Western Democracies*, in «Information, Communication & Society», 22(13), pp. 1973-1988.

Krishnan, S. and Chen, M. (2018), *Identifying Tweets with Fake News, paper presented at the 19th Ieee International Conference on Information Reuse and Integration (Iri)*, Salt Lake City, Utah, USA, 7-9 July.

Kuehn, K. and Salter, L. (2020), *Assessing Digital Threats to Democracy, and Workable Solutions: A Review of the Recent Literature*, in «International Journal of Communication», 14(22), pp. 2589-2610.

Loomba, S., De Figueiredo, A., Piatek, S., De Graaf, K., and Larson, H. J. (2021), *Measuring the Impact of Covid-19 Vaccine Misinformation on Vaccination Intent in the Uk and Usa*, in «Nature human behaviour», 5(3), pp. 337-348.

Lukito, J. (2020), *Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on three Us Social Media Platforms*, 2015 to 2017, in «Political Communication», 37(2), pp. 238-255.

Luo, M., Hancock, J. T. and Markowitz, D. M. (2022), *Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues,* in «Communication Research», 49(2), pp. 171-195.

Maloy, R., Butler, A. and Goodman, L. (2022), *Critical Media Literacy in Teacher Education: Discerning Truth Amidst a Crisis of Misinformation and Disinformation*, in «Journal of Technology and Teacher Education», 30(2), pp. 167-176.

Manzoor, S. I., Singla, J. and Nikita (2019), *Fake News Detection Using Machine Learning Approaches: A Systematic Review, paper presented at the 3rd international conference on trends in electronics and informatics (Icoei)*, Tirunelveli, India, 23-25 April.

Musi, E. and Reed, C. (2022), *From Fallacies to Semi-Fake News: Improving the Identification of Misinformation Triggers across Digital Media*, in «Discourse & Society», 33(3), pp. 349-370.

Oecd (2019), *Oecd Skills Studies Skills Matter Additional Results from the Survey of Adult Skills.* Paris: Éditions Ocde. Available at https://www.oecd.org/skills/piaac/publications/Skills_Matter_Additonal_Results_from_the_Survey_of_Adult_Skills_ENG.pdf (last accessed on 27th December 2022).

Olariu, O. (2022), *Critical Thinking as Dynamic Shield against Media Deception. Exploring Connections between the Analytical Mind and Detecting Disinformation Techniques and Logical Fallacies in Journalistic Production*, in «Logos Universality Mentality Education Novelty: Social Sciences», 11(1), pp. 29-57.

Palano, D. (2019), *The Truth in a Bubble the End of «Audience Democracy» and the Rise of «Bubble Democracy»*, in «Soft Power», 6(2), pp. 36-53.

Pariser, E. (2011), T*he Filter Bubble: What the Internet is Hiding from You.* London, Penguin.

Paul, R. and Elder, L. (2019), *The Miniature Guide to Critical Thinking Concepts and Tools.* London, Rowman & Littlefield.

Pérez-Rosas, V. and Kleinberg, B. (2017), *Automatic Detection of Fake News*, in «arXiv preprint», doi: https://doi.org/10.48550/arXiv.1708.07104.

Pinkleton, B., Austin, E.W., Zhou, Y., Willoughby, J.F. and Reiser, M (2012), *Perceptions of News Media, External Efficacy, and Public Affairs Apathy in Political Decision Making and Disaffection*, in «Journalism & mass communication quarterly», 89(1), pp. 23-39.

Puddephatt, A. (2011), *The Importance of Self-Regulation of the Media in Upholding Freedom of Expression Communication and Information*, Brasilia, Unesco Office Brasilia. Available at https://unesdoc.unesco.org/ark:/48223/pf0000191624 (last accessed on 27th December 2022.

Reis, J. Cs., Correia, A., Murai, F., Veloso, A. and Benevenuto, F. (2019), *Supervised Learning for Fake News Detection*, in «Ieee Intelligent Systems», 34(2), pp. 76-81.

Rubin, V.L. (2022), *Misinformation and Disinformation: Detecting Fakes with the Eye and Ai.* London, Springer Nature.

Shahid, W., Li, Y., Staples, D., Amin, G., Hakak, S. and Ghorbani, A. (2022), *Are You a Cyborg, Bot or Human? A Survey on Detecting Fake News Spreaders*, in «Ieee Access», 10, pp. 27069-27083.

Sinek, S. (2009), *Start with Why: How Great Leaders Inspire Everyone to Take Action.* London, Penguin.

Sitaula, N., Mohan, C. K., Grygiel, J., Zhou, X. and Zafarani, R. (2020), *Credibility-Based Fake News Detection*, in K. Shu, S. Wang, D. Lee and H. Liu (eds), *Disinformation, Misinformation, and Fake News in Social Media Emerging Research Challenges and Opportunities*, Switzerland, Springer Nature, pp. 163-182.

Spohr, D. (2017), *Fake News and Ideological Polarization: Filter Bubbles and Selective Exposure on Social Media*, in «Business Information Review», 34(3), pp. 150-160.

Terren, L. and Borge-Bravo, R. (2021), *Echo Chambers on Social Media: a Systematic Review of the Literature*, in «Review of Communication Research», 9, pp. 99-118.

Unesco (1978), *Records of the General Conference*, Paris, Unesco. Available at https://treaties.un.org/doc/source/docs/unesco_res_5_9.2_1-E.pdf (last accessed on 27th December 2022).

Urbani, S. (2019), *Verifying Online Information*, First Draft. Available at https://firstdraftnews.org/long-form-article/verifying-online-information (last accessed on 27th December 2022).

Vaccari, C. and Valeriani, A. (2021), *Outside the Bubble: Social Media and Political Participation in Western Democracies.* Oxford, Oxford University Press.

Wardle, C. and Derakhshan, H. (2017), *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking.* Strasbourg, Strasbourg Cedex. Available at http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf (last accessed on 27th December 2022).

Westerlund, M. (2019), *The Emergence of Deepfake Technology: A Review*, in «Technology Innovation Management Review», 9(11), pp. 39-52

Zompetti, J.P. (2019), *The Fallacy of Fake News: Exploring the Commonsensical Argument Appeals of Fake News Rhetoric through a Gramscian Lens*, in «Journal of Contemporary Rhetoric», 9, pp. 139-159.