CHAPTER 6

Identifying Disinformation That Targets Mass Audiences

Disinformation can be a component of operations conducted by intelligence services or specialized military units. Their activities may be directed at specific political or military decision-making centers or entire societies. Regardless of the object of disinformation, its goal is always to trigger changes in the consciousness of recipients and disturb their cognitive abilities. This can then lead to a change in attitudes and cause a specific social, economic, or political reaction (Aronhime et al., 2021). Disinformation activities directed at a mass audience may aim to:

- Sow doubts and manipulate public opinion.
- Influence social and political attitudes.
- Distract public debate.
- Polarize the political situation.
- Undermine the value system of a state or community.
- Weaken the cohesion of a state or group of states.
- Undermine trust in public institutions and the media.
- Spread ideologies and discredit the opposing ideologies.
- Inspire chaos, division, and conflicts.
- Destabilize a society, state, or community.
- Undermine the integrity of the government, constitutional principles, or political (decision-making) processes.

Despite the previously detailed characteristics that differentiate misinformation, disinformation, and malign information, these three classifications still have many things in common. All three are tools in the hands of a manipulator, and they can be used to elicit a certain response by a recipient while camouflaging the manipulator's real intentions.

Counteracting and combating disinformation are difficult in open societies where fundamental freedoms like freedom of speech are valued and protected. In such circumstances, it is difficult to punish or publicly stigmatize those responsible for creating, reproducing, and disseminating manipulated or false information. This is due to several reasons.

First, in some cases, disinformation can be created as a result of a person's insufficient intellectual or social competence. It can therefore be reproduced through inadvertent action and without malevolence. In democratic societies,

© ROBERT KUPIECKI ET AL, 2025 | DOI:10.1163/9789004715769_009
This is an open access chapter distributed under the terms of the CC-BY 4.0 license.

presumption of innocence is a fundamental component of the rule of law and freedom of expression. It has also been the foundation of efforts by the democratic community to de-penalize journalistic offenses (i.e., derivatives of misuse of information) and strongly present in the activities of the Organization for Security and Co-operation in Europe (OSCE).

Second, disinformation often has a basis in fact but relies on the author's interpretation, opinions, or comments to create a message that is inconsistent with reality. A dilemma therefore arises between the potential harm that false information may inflict on an individual or society and the right to freedom of expression. Both values are the subject of legal protection in modern democratic societies; but while it is often difficult to prove that harm was derived from such actions, the right to free expression remains highly valued and protected within the free world.

Third, disinformers who are caught spreading false or manipulated information often invoke their right to freedom of speech, which is guaranteed in Article 11 of the Charter of Fundamental Rights of the European Union. Despite the fact that EU institutions recognize that disinformation not only "undermines trust in institutions, traditional and digital media, and damages democracies by preventing citizens from making informed decisions" but also "impairs freedom of expression" (*Tackling online* 2018, p. 1), they have not put forth proposals for administrative and regulatory measures or an in-depth reform of applicable laws to address this threat. While the integrity of public debate is safeguarded by established institutions and the strength of civil society, this task is more difficult in the case of online platforms, which are seen as a "space for unlimited freedom of expression". To effectively counteract the phenomenon of disinformation, individual countries have independently introduced legal provisions aimed at minimizing its effect.

In 2017, the German parliament adopted the Network Enforcement Act (NetzDG) to improve law enforcement in matters relating to social media. According to the legislation, digital platforms must provide users with the ability to submit complaints and they must remove content that is illegal (within 24 hours) or contains false messages (within 7 days). Yet this does not address the fact that in the digital age disinformation can be reproduced countless times within 24 hours and its effects may not be reversed. Apart from the obvious weakness of such late intervention, some German political parties consider this law to be contrary to freedom of expression and the constitution.

In France, a similar law requires internet service providers to disclose information regarding entities funding the promotion of political content during an election period. It also prohibits the dissemination of inaccurate, misleading, or suspicious statements that could impact the fairness of elections, with the

exception of satirical content and programs. Moreover, the French regulatory authority has the power to reject a media license application if it determines that the organization in question poses a significant threat to values such as human dignity or freedom of thought and opinion, or if it jeopardizes essential national interests, including the proper functioning of public institutions. The act also imposes obligations on internet platform operators toward users and state authorities, such as:

- Providing tools for reporting false information.
- Ensuring the transparency of algorithms used.
- Promoting content from news agencies and audiovisual communication services.
- Disabling accounts promoting false information on a massive scale.
- Informing users about sponsored content of general interest and disclosing the amount of money received to promote it.
- Providing information on activities in the field of media and information education (Ogrodowczyk et al., 2020, pp. 24–28).

Disinformation targeting mass audiences refers to large-scale information warfare activities that aim to achieve a strategic goal through coordinated disinformation operations. It can also be defined as "a series of coordinated information operations conducted by a foreign entity aimed at influencing a specific group of recipients with the intention of achieving specific goals and benefits by the initiator" (Twetman et al., 2019, p. 5). A crucial aspect of this phenomenon is the unlawful use of information against society, the principles of democracy, and open public dialogue. Disinformation campaigns exploit weaknesses in media systems, cognitive biases, and public opinion-shaping processes. They are often conducted covertly, with the help of local proxies to make it challenging to identify the real, often foreign, perpetrator (Maurer, 2018, pp. 171–188). The objectives of such campaigns include eroding trust in public institutions, promoting social polarization, radicalizing public debate, and excluding certain social groups.

Disinformation campaigns often revolve around a "topic" that is easy to remember, evokes associations with the targeted entity, has significant social or political importance, and elicits strong emotions. They may also be based on issues used in several operations at the same time. Such campaigns often center around current political events (such as elections, legislative activities, public scandals, or other contentious issues) that provoke public objection. The choice of the subject of disinformation campaigns is not random but rather the result of operational diagnosis, which is akin to "market research" aimed at determining: (1) What topics are important to a given group of recipients? (2) Which of these topics evoke the greatest emotions? (3) How will the

manipulation of messages on specific topics affect the political and social situation in the target country?

Regardless of the subject of the disinformation campaign and its specific objectives, the overarching goal remains to instill doubt in the recipient, create a power-society discrepancy, and promote an intentionally crafted alternative way of interpreting reality. Due to the limited legal and physical means available for combatting disinformation within a democratic system of values, building social resilience is the primary defense mechanism. Its foundation is the individual skills of each recipient of information in recognizing falsehoods and manipulation, properly selecting reliable and credible sources, and debunking lies online.

1 Recognizing Disinformation

There are currently approximately 4.57 billion active internet users worldwide, each producing an average of 1.7 megabytes of data per second. This amounts to a staggering 2.5 trillion megabytes of data generated daily. With socioeconomic dynamics and technological acceleration expected to continue, these numbers are only set to increase, driven by faster computing powers, new applications like the Internet of Things, and the capabilities of quasihuman or superhuman machines fueled by artificial intelligence (Kupiecki, 2020a, pp. 472–497). These developments will create new opportunities for variously motivated entities operating in the domain of disinformation, who will attempt to create alternative realities to garner political or business benefits or harm individuals, nations, and the international community.

In the last two decades, as much as 90% of the world's data has been produced. This number is projected to increase to 463 exabytes by 2025 (Bulao, 2020). With so much information available, how can individuals and nations navigate it and evaluate its credibility? How can they distinguish real information from manipulated or false information? To answer these questions, it is essential to study the reliability of information sources and producers, particularly in a world where the monetized "click-through rate" is prioritized over reliable and credible information. The business model of news media outlets, which relies on advertising profits, often leads to the rapid publication of new content at the expense of quality and accuracy. This rush to publish can limit the time available to verify information, while sensationalized, emotionally manipulative headlines (clickbait) capture readers' attention and are used to generate ad revenue (Głowacka et al., 2019, p. 4).

In the modern world, true information (facts) coexists with falsehoods, and this makes it increasingly difficult for recipients to discern which is which. The ability to freely choose what to believe is a great privilege of freedom, but it requires basic knowledge and intellectual skills. Computer algorithms today can collect data on users' online activity, analyze their preferences and interests, and then send suggestions that align with their anticipated expectations or "needs". The current technological possibilities in the field of profiling offer a wide range of opportunities for social engineering, including direct and effective influence on the attitudes and political decisions of individuals and states. These practices raise serious ethical questions, as illustrated by the role of Cambridge Analytica (CA) in the 2016 U.S. presidential election, the Brexit referendum, and the secession referendum of Catalonia in 2017. CA obtained data from 50 million Facebook users without their awareness, which allowed the company to segment and target voters. Based on this, CA developed specific content and methods of distribution in social media to reach voters and influence their attitudes (Boldyreva et al., 2018, pp. 91–102).

In the 21st century, cyberspace has become a modern field for civil and military interactions, including information warfare, cyber-attacks, data and intellectual property theft, false identity usage, cyberespionage, tracking, and surveillance. New technologies offer wide possibilities for creating an alternative reality by means of mass production and distribution of manipulated or completely false information. Authoritarian states use these techniques systematically and over long periods to influence social attitudes and decisions made by other participants in international relations. These actions can also be used to provoke riots, social unrest, and even armed conflicts. It is therefore important to consider:

- How can we mitigate the disinformation threat?
- How do we distinguish real information from fake news?
- How can we assess the credibility of information sources and make appropriate selections?
- How can we immunize ourselves to our surroundings against massive disinformation attacks?
- How do we educate adults and youth on these topics from the earliest stages of education?

The group of targets for disinformation includes politicians, diplomats, soldiers, business people, experts, analysts, journalists, commentators, academics, and other social groups. Each of these groups must possess the necessary skills and tools to defend themselves against false information. The success of disinformation is measured by the degree to which the manipulated message

is recognized as 'true' by its recipients. This not only alters their perception of the phenomenon in question but also perpetuates the disinformation in a wider group of recipients. Anyone can therefore help mitigate the harmful impact of disinformation by checking the credibility of information and its source before sharing it, or by exposing information falsehoods if they have already been disseminated. Manipulated individuals – often enjoying social authority – who spread false information not only expand the scope of the destructive influence of disinformation but also legitimize its "truthfulness" among their followers. According to research conducted by the MIT Media Lab, real information takes about six times longer than fake information to reach 1,500 people on Twitter. Untrue news has nearly a 70% greater chance of being shared than news based on facts, particularly when it comes to politics (Metz, 2008). To counteract the phenomenon of mindless forwarding of unverified information, Twitter has introduced a mechanism that warns users and suggests reading linked articles before sharing. The "trap of authority" and the "trap of the attractiveness of fake news" additionally reinforce the functioning of recipients in "filter bubbles", which display personalized messages that may not always reflect reality.

Another concerning trend in information operations is the questioning of the credibility of scientific research and expert opinions by self-proclaimed pseudo-authorities. Bloggers, vloggers, and celebrities often lack relevant knowledge, skills, and qualifications yet they are very popular on social media platforms. With thousands of followers on Facebook, Instagram, YouTube, or Twitter, they can influence society to a greater extent than reliable and credible individuals and institutions. Targeting disinformation activities at such groups will be most effective in operations aimed at destroying the foundations of mutual social trust.

2 The RESIST Model: Recognition and Analysis of Disinformation

Recognizing and analyzing disinformation requires knowledge of the principles, methods, and techniques used to mislead the recipient. The British government communication service has developed a useful model called RESIST, which includes tools for identifying and analyzing disinformation, providing early warning, conducting proactive and reactive strategic communication, and verifying its effectiveness. This model was primarily created for public institutions and the private sector. It is one of many proposals on how to combat disinformation, hate speech, and harmful marketing campaigns (known as "black PR") by organizing special teams or departments responsible for monitoring, analyzing, and responding to such phenomena. The British government's initiative

aims to standardize the methods and means of combating disinformation and facilitate cooperation in situations of crisis or external aggression that employs hybrid techniques. The RESIST model can also serve as a guide to designing an organization's information or communication policy. Its flexible solutions allow for adapting to the individual needs and specificities of a particular entity.

The British model of recognition, analysis and response to disinformation (RESIST) contains the following elements:

a) Recognize. This provides an overview of the current information environment, specifically the vast amount of information available online and the difficulties it poses for individuals trying to navigate it effectively. It helps to explain the distinctions between misinformation, disinformation, and harmful information, as well as the potential negative effects they can have on those who consume them. As part of the diagnosis, a checklist is created that is used to determine the likelihood of the manipulation or falsification of information. It involves the following control questions:
   – What are the goals of disinformation?
   – What are the techniques of disinformation?
   – How are disinformation techniques combined to achieve an effect?

b) Early warning. This begins with a review of available tools that can be used to monitor the information space. The assessment enables organizations, groups, or individuals to prioritize their actions and identify areas vulnerable to disinformation. By identifying targets, audiences, disinformers, and risks, it helps to focus on monitoring key weaknesses and take necessary measures to mitigate the impact of disinformation. The relevant control questions are:
   – How can I prioritize digital monitoring?
   – How can I build an individual set of tools for digital monitoring?
   – How can I use digital monitoring to assess the facility's susceptibility to potential threats?

c) Situational insight. This explains how disinformers shape the information environment and emphasizes the importance of situational context analysis. The analysis can be conducted systematically through regular reports (daily, weekly, or monthly) or in response to emerging threats and issues. It focuses on the following control question:
   – What is the situational context of disinformation activities and how can it be used to support a rapid response?

d) Impact analysis. By analyzing the methods and techniques used in disinformation campaigns, it is possible to not only understand the meaning of specific disinformation operations but also predict likely future campaigns and their impact on recipients. The control questions are:

  - What is the likely purpose of the disinformation?
  - What is the likely impact of the disinformation?
  - What is the likely extent of the disinformation?
  - How should disinformation be prioritized?

e) Strategic communication. This contains a set of key methods and tools for proactive, active, and reactive communication. Public and private sector entities should consider these tools when developing and implementing their communication policies and strategies. It is necessary to identify effective channels to reach target groups and use "friendly voices" to increase the credibility and reach of the entity's communication activities based on various situational scenarios. The control questions are:
  - What should the public response to disinformation look like?
  - What is the approval process like?
  - What are the available response options?

f) Track outcomes. This step enables users to evaluate the effectiveness of their own strategic communication. It starts with the following control questions:
  - How should information on the disinformation campaign be recorded and shared ?
  - How can I evaluate my own actions and understand the conclusions drawn? (*RESIST Disinformation,* 2020, p. 4; *RESIST 2*, 2021, pp. 6–7)

In a simplified version, the RESIST model can also be used by individuals or analytical networks whose abilities to identify, analyze, and debunk false information contribute to the system of social resistance to disinformation. The knowledge of basic methods, tactics, techniques, and mechanisms of disinformation can improve their capabilities. Such knowledge makes recipients better prepared to recognize manipulations of narratives, facts, or contexts. The basic components of disinformation can be summarized with the acronym FIRST:

- Fabrication. This is the manipulation of a message's content through the use of false text, documents, or pictures.
- Identity. This concerns concealing or stealing a person's identity to use it on fake social media accounts.
- Rhetoric. This is the use of argumentation based on false information or offensive content in messages, for example, by trolls commenting on social media posts.
- Symbolism. This is the malign use of symbols to strengthen a communication message by comparing a politician to a controversial historical figure, for instance.
- Technology. This involves the use of technological advantages, like bots, which can automatically produce false messages on a mass scale (*RESIST Disinformation*, 2020, p. 9).

Although the above elements provide a general understanding of how false information is created and disseminated, their purpose is primarily to raise awareness among recipients about the need to verify information consciously. Unlike professional, reliable, and credible press offices, information published on the internet is often not subject to a comprehensive verification and approval process. In principle, each user can create their own "media" (such as a website, blog, vlog, or portal) and publish any content they want. The quality and credibility of such content must be evaluated independently by the recipient, who may not always have sufficient knowledge about the subject matter. Technical capabilities to manipulate content, modify photos, impersonate others, commit identity theft, or create fake accounts are widely available. However, this does not mean that all users employ them with bad intentions. Maintaining a well-understood skepticism and criticism toward unknown sources of information can reduce susceptibility to disinformation. A "classical methodology" for conscious verification of information was characterized by the French researcher specialized in this phenomenon, Vladimir Volkoff, who distinguished the following information manipulation methods:

- Negation or inversion of facts.
- Combining truth and lies. This is used in a situation where the basic circumstances are generally known but the details have not been made public.
- Modification of a motive. This refers to a tactic used by disinformers where they change an element whose value was only known to them before a certain event. Once the situation is recognized, the disinformers change their motive of action to something that is socially acceptable and compliant with the norms of their environment.
- Modification of circumstances. This refers to a change in circumstances that are difficult to confirm unequivocally, particularly at an emotional level, such as judgments, feelings, and relationships. This method introduces chaos into the assessment of the situation, making it harder to determine the truth.
- Blurring. This has the same effect as the previous method but consists of flooding the main information with a large number of irrelevant facts.
- Camouflage. This is the opposite of blurring and involves breaking down the main information into such small details that important elements are lost.
- Interpretation. This is used when the facts are indisputable. They are then interpreted in such a way as to achieve the desired effect.
- Generalization. This tactic constructs a universal principle based on an individual example.
- Illustration. This is the use of an individual event as a "legitimate" illustration of a wider social phenomenon.

– Unequal representation. This is the manipulation of the quality and popularity of an information source, which can influence the difference in perception of information between an opinion-forming and popular source versus a little-known and niche source.
– Equal representation. This method is used in the final phase of a disinformation campaign, when the majority of the target group is already convinced of the theses promoted by the agents of influence. Its primary aim is to consolidate the already widely accepted opinion and close the topic (Volkoff, 1999a).

Despite the passage of time, the disinformation methods described by Volkoff continue to be used. Their universal nature is based on the cognitive limitations of human beings. Manipulating the content of a message falsifies the image of reality that reaches the recipient, influencing perception, shaping opinion, and determining actions. Technological capabilities for collecting big data about internet users facilitate the selection of target groups for disinformation campaigns. Combined with the ability to mass disseminate false content using computer algorithms, disinformation can be considered a unique "weapon of mass destruction of human minds in the 21st century". New technologies have expanded the possible methods, techniques, and mechanisms of disinformation, which include:

– Astroturfing. This involves presenting top-down agitation campaigns as civic initiatives or falsely attributing a given message to other entities to authenticate them.
– The bandwagon effect. This is a cognitive effect which reinforces a specific opinion or belief because it is shared by others. Social media users are more likely to share articles that have been shared by many others, regardless of their content.
– False connotation. This a situation in which the title, lead, photos, or graphics used do not correspond to the content of the message.
– False context. This is a situation in which content is based on facts but is placed in a manipulated information context.
– Filter bubble. Filtering algorithms use personalized access to information' based on a user's search history and social media activity. This puts the user in a situation where they are more likely to see content that corresponds to their previous online activity.
– Leaks. This is the deliberate distribution of information obtained illegally, which includes, for instance, the publishing of classified documents or the theft and publication of private or business correspondence written by persons holding state positions.

- Malign rhetoric. This is the use of slanderous and false accusations used to disrupt public debate.
- Manipulation. Modifying the content of information can be used to alter its meaning.
- Misappropriation. This involves falsely attributing an argument, statement, or position to someone.
- Satire and parody. This involves making fun of people (e.g., using memes) or perpetrating narratives or opinions with the aim of undermining their importance.
- Sock puppets. This is the creation of a fictional debate between two (or more) entities using new technologies. For example, this can be done by creating fake social media accounts and conducting discussions between them.
- Trolling. This involves the deliberate provocation and aggravation of discussions on social media by placing controversial, offensive, or emotional comments to provoke outrage in recipients and draw them into a discussion (*RESIST Disinformation,* 2020, pp. 21–22; Brodnig, 2017).

The scale of the problem around disinformation is best illustrated by the amount of false information circulated about the coronavirus pandemic. According to a study by Carnegie Mellon University (CMU), 82% of the 50 most popular Twitter accounts duplicating fake news about COVID-19 were bots (Huang & Carley, 2020). The consequence of this situation is the development of an "infodemia", as described in the first part of this book, resulting from the creation and reproduction of huge amounts of misinformation, disinformation, and conspiracy theories that questioned medical facts or assigned the "invention" of the virus to Western states or pharmaceutical companies for their political and commercial goals (Constantinou et al., 2021, p. 4).

In Poland, for example, approximately 50% of the population agrees with conspiracy theories about the coronavirus, with between 43% and 56% of respondents aged 18–74 believing in them, depending on the theory (Duplaga, 2020). Based on COVID-19 disinformation, numerous anti-vaccine movements emerged that questioned the lethality of the virus and the effectiveness of the vaccines. The lack of social immunity to such content translated into an insufficient level of vaccination of the population, leading to lower collective immunity to the virus. As a result, the number of victims of the pandemic increased to over 106,000 in Poland and over 5.7 million globally (as of February 2022). It is important to note that not every case of death or coronavirus infection was the result of omission or fear compounded by disinformation. It undoubtedly contributed to these statistics, however.

It may seem that individual users are helpless when faced with new technologies in the battle against disinformation. However, knowing the methods, techniques, and mechanisms of disinformation is very useful in identifying and analyzing materials with manipulated content. An example of this is a short study of source material containing a false historical narrative regarding the introduction of martial law in Poland in 1981. On December 13, 2020, the 39th anniversary of this tragic event, the Russian disinformation outlet *Sputnik Polska* published an article aimed at depreciating the importance of these events and trying to change the perception of Poles.

To do so, the author employed the following techniques:

– Satire and parody. Ironizing about the alleged "cruelty of the communist leader General Wojciech Jaruzelski", whose greatest crime in the author's opinion was "the lack of a morning TV program for children", the author attempted to diminish the significance of the general's actions. Additionally, the author mentioned her "gratitude to Jaruzelski" for not having to attend school for three weeks.
– False context. In the article, the author deprecated the brutality of the methods used by the communist repression apparatus, ignoring the facts of 40 deaths during martial law with 60 additional people wounded and over 10,000 interned.
– Providing data without a source. The author cites alleged public opinion polls suggesting that Poles have recently changed their attitudes toward martial law to be more positive, but no source is referenced. The hyperlink provided leads to a text about the conflict in Nagorno-Karabakh instead of a relevant source on the topic.
– Justification. The author justifies the decision to introduce martial law by suggesting that, otherwise, the Soviet army would have intervened in Poland according to the "Brezhnev Doctrine." While historical knowledge confirms the risk of such an intervention, it does not justify the brutality of the repressive apparatus of the Polish communists, who had begun preparations to pacify Solidarity well before December 13, 1981.

Russian disinformation operations often employ historical manipulations (Juurvee et al., 2020; Domańska & Rogoża, 2021; Legucka & Kupiecki, 2022), particularly regarding the Second World War, which is referred to as the "Great Patriotic War" in Russia. During the 80th anniversary celebrations, Russian authorities, including Vladimir Putin, suggested that Poland was responsible for provoking the Third Reich to attack, depreciating the significance of the secret Molotov-Ribbentrop Pact and pointing to the "Munich Agreement" of 1938 as the main cause of the conflict. By denying the Red Army's complicity in the aggression against Poland on September 17, 1939, the myth of the "liberator"

and "defender of the Slavic nations" is perpetuated. Attempts to combat this false narrative are met with accusations of "Russophobia." By shaping an alternative interpretation of history, the Kremlin influences not only external recipients but also the historical awareness and sensitivity of Russians themselves.

According to research by the Center for Polish-Russian Dialogue and Understanding and the Levada Center, most Russians view the Red Army's intervention in Poland as "brotherly help" (47%) or "defense of their own territory" (48%). Almost half (43%) believe that the Nazis are responsible for the Katyn massacre, which involved the murder of around 22,000 Polish citizens, including army and police officers, by the Soviets in the spring of 1940. Only 26% of Russians are aware of the Stalinist apparatus of repression's responsibility for this act (*Obraz Polski,* 2020, pp. 22–23).

The simple identification of disinformation may be considered an adequate defense mechanism if the recipient detects it, recognizes it as a threat to the cognitive process, and rejects it, rendering it permanently unreliable. Recognizing false or manipulated information is not a straightforward task, however; it requires theoretical knowledge and practical skills in critical thinking and fact-checking. These competencies, bolstered by open-source intelligence (OSINT) techniques, provide a broad range of possibilities for every disinformation analyst.