6. AI and the weaponization of information: Hybrid threats against trust between citizens and democratic institutions

INTRODUCTION

Although there is nothing necessarily new about propaganda, the affordances of social networking technologies – algorithms, automation, and big data – change the scale, scope, and precision of how information is transmitted in the digital age. (Bradshaw & Howard, 2019, p.11)

The control of information and telecommunication infrastructure, with the ability to respond to cyberattacks and to ensure cybersecurity, offers real power and is already one of the most significant political, economic and technological issues of the twenty-first century. (Ghernaouti, 2013, p.ix)

Disinformation is not new. Roman emperors and the British and German commanders in the two World Wars harnessed this power against their adversaries (Bittman, 1981). Joseph Goebbels, the Nazis’ propaganda expert, modeled their propaganda efforts around the “Big Lie,” the idea that repetition makes information truer (Ramakrishna, 2018). Jean Oberle (2017) reported that the Germans did not use bombs on Paris but rather false news: the Germans were pushing the former “Petainists” to act. During the Cold War, the USSR developed the concept of Reflexive Control theory (Thomas, 2004) to describe their approach to disinformation. A milestone in modern disinformation history is the Cold War, when the tensions between the two blocs led states to set up professional organizations whose main role was to produce and broadcast deception (Bittman, 1981).

According to Facebook, disinformation is intentional, often strategic in the sense that it targets specific demographics, and embeds false stories and coordinated efforts from real and fake accounts to engage the audience (Bennett &
Livingston, 2018). Facebook adopted the following operational definition of disinformation:

Disinformation – Inaccurate or manipulated information content that is spread intentionally. This can include false news, or it can involve more subtle methods such as false flag operations, feeding inaccurate quotes or stories to innocent intermediaries, or knowingly amplifying biased or misleading information. Disinformation is distinct from misinformation, which is the inadvertent or unintentional spread of inaccurate information without malicious intent. (Weedon, Nuland, & Stamos, 2017, p.5)

Traditionally, an electoral campaign fulfills its role of informing and raising the awareness of very broad segments of society (Sidjanski, 1979). Thornton (2015) refers to the notion of “information warfare” to describe a context where information is the weapon and the minds of citizens the new “battlefield” (Cavelty and Mauer, 2008). They consist of operations to polarize civil society, sow chaos in the population of another state, and thereby weaken the opponent. Many tools and tactics exist, including AI-powered social bots to influence online conversations. Although varying in resources and capabilities, many governments’ armed forces and intelligence agencies “have developed aggressive external operations” (Deibert & Pauly, 2019, p.83).

The previous chapter introduced the notion of fake news and discussed disinformation in the context of national politics. This chapter focuses on disinformation campaigns from an international relations perspective, meaning how disinformation is weaponized by some states to gain power, destabilize and weaken other states (Golovchenko, Hartmann, & Adler-Nissen, 2018). These operations target the trust that citizens have placed in their institutions (governments, representative mechanisms, the press), and in the social dialogue (leading to polarization of society). Without trustworthy information, citizens (more so than civil society at large) are more vulnerable to manipulation and disengagement from democratic processes. The grand strategy behind the operations takes time to assess, and therefore to identify all the operations that are in fact either stemming from the same group of actors or have the same grand strategy.

The factors that make this strategy so powerful are that this type of “warfare” is continuously ongoing and hard to detect. It is complicated to identify its source, particularly as more often than not it is waged from several sources simultaneously. And finally, such a warfare strategy penetrates all levels of society at a very low cost. Even if the audience does not necessarily believe in the planted information, the abundance of unvetted information of itself leads to a persistent distrust of public information and the media. (Spruds et al., 2016, p.8)
This strategy is operationalized through tactics and tools that include new forms of deception and image-manipulation activities (Molander, Riddile, Wilson, & Williamson, 1996) and a weaponization of social media platforms. AI and more precisely Machine Learning Algorithms (MLAs) of social media platforms play a key role in the diffusion of disinformation campaigns. These tactics are about

(...) influencing the target audience’s values and belief system, their perceptions, emotions, motives, reasoning, and ideally, their behaviour. It is (...) aimed at maintaining the support of the loyal; convincing the uncommitted and undermining the opposition. This is achieved through influencing people’s perception of what is going on and, in turn, influencing their online and offline behaviour by playing on emotional and logical arguments drawn from conversations and history, and by tapping into an existing narrative. (Nissen, 2015, p.84)

This chapter examines the main characteristics of disinformation campaigns in the context of a geopolitical power play on cyberspace. It argues that the liberal democratic model is under attack by authoritarian regimes, which use disinformation campaigns to threaten the citizen–government relation. The role of AI in this power play is crucial: it is used both to spread disinformation (intentionally through automated bots or unintentionally through MLA of social media platforms) and to defend against disinformation (the main solution to face the information avalanche produced and consumed globally). This chapter also highlights the difficulty of governments of liberal democracies to ensure that citizens have access to information that is “complete, objective, reliable, relevant, easy to find and to understand” (OECD, 2001, p.1). Their dependence on big tech companies to police the online information environment is problematic since it grants private companies the role to censor content. It also highlights the challenge that government and public entities face when enforcing law and the protection of citizens.

This chapter first discusses the ongoing geopolitical power play on cyberspace with various disinformation strategies and tactics. It then examines disinformation operations to harm trust in democratic institutions and the news ecosystem. It also discusses specific disinformation campaigns that were conducted prior to and during the Covid-19 pandemic in Europe. Finally, it presents non-technological responses to disinformation campaigns.

GEOPOLITICAL POWER PLAY ON CYBERSPACE

The well-known paper “Cyberwar is Coming!” by the two RAND Corporation scientists, John Arquilla and David Ronfeldt (1993), argued that the information revolution was altering not only how conflicts take place (e.g. parties involved, terrain, technologies, etc.), but also the nature of conflicts, spurring
AI and the weaponization of information

a need for new military structures, doctrines and strategies. The use of disinformation as part of the tactics to weaken other states is not new: intelligence services have always led espionage and reconnaissance activities to support kinetic military operations. What is different today is the domain in which disinformation is spread. This domain is man-made, meaning that it can be easily altered and manipulated.

In the mid-2000s, the topic of cybersecurity was increasingly discussed by the media and policy makers, following cyberattacks such as the 2007 attack against Estonia. This cyberattack followed the relocation of a statue commemorating the engagement of the Russian army in the Second World War to the suburbs of the city. The choice of day to relocate this statue was not random, since it was 1 May, the same day Russia honors its participation in the Second World War. The cyberattack that followed targeted websites and online services of private and public organizations, including banking and government services, as well as newspapers and broadcast media. This attack didn’t cause any human harm but millions of dollars of losses (McGuinness, 2017). Although Russia denied any involvement, Estonia pointed at its neighbor and international experts agreed (Van Puyvelde & Brantly, 2019).

This attack put cybersecurity on the agenda of a larger number of countries, since they realized that increased connectivity meant increased vulnerability as well, as pointed out by the US joint forces years later: “The prosperity and security of our nation are significantly enhanced by our use of cyberspace, yet these same developments have led to increased exposure of vulnerabilities and a critical dependence on cyberspace, for the US in general and the joint force in particular” (Joint Publication, 2018a, p.1–2). This also signaled the realization that cyberconflict can be ongoing in times of inter-state peace. States conduct offensive operations in cyberspace, whether it is to steal information, destroy capabilities or infrastructure, or disinform populations, when there is no conflict between them. Sheldon (2011) characterizes cyberspace both as a space where conflict can take place, as well as means to advance their interests: “Cyberspace is the domain in which cyber operations take place; cyberpower is the sum of strategic effects generated by cyber operations in and from cyberspace” (p.96).

Cyberspace is poorly regulated. A large-scale disinformation campaign to interfere with the internal affairs of another state, including electoral processes, does not constitute an act of war that could justify a kinetic military response. The diplomatic response given during the last days of the Obama administration to the Russian interference (CNN, n.d.) is a good illustration of this challenge: “After discovering the existence, if not the full scope, of Russia’s election interference efforts in late-2016, the Obama Administration struggled to determine the appropriate response. Frozen by ‘paralysis of analysis,’ hamstrung by constraints both real and perceived, Obama officials..."
debated courses of action without truly taking one,” said committee chairman Senator Richard Burr, a North Carolina Republican (Cohen & Herb, 2020).

Disinformation campaigns take place primarily on social media platforms designed and maintained by companies with headquarters in the USA, and most disinformation content comes from outside Europe. Consequently, this means that disinformation campaigns are influencing European citizens and states but on foreign media and with foreign content. This double territoriality challenge adds to the qualification challenge discussed previously. These elements challenge traditional military strategies and call for developing new military strategy specifically dedicated to disinformation. Moreover, a specific tactical warning system is required to distinguish between disinformation attacks and other activities such as espionage or accidents.

In his famous paper, Thomas Rid (2012) argued that “any act of war has to have the potential to be lethal; it has to be instrumental; and it has to be political” (p.5), which consequently means that cyberwar will never take place: “Cyber war has never happened in the past. Cyber war does not take place in the present. And it is highly unlikely that cyber war will occur in the future” (Rid, 2012, p.6). In 2013, NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) sponsored a research project that led to the publication of the Tallinn Manual, which addressed the most severe cyber operations, and where the authors argue that a cyberattack can only be considered an armed attack if its impact leads to injury, death, or destruction (Schmitt, 2013, p.106). Nevertheless, the existence of such conflict requires European pluralist democracies to not only adopt measures to combat disinformation campaigns, but also to perceive these cyber operations for what they are: cyberattacks against the integrity of a state, an interference in the internal affairs of a state, and an attempt to weaken European pluralist democracies. We will refer to the notion of cyberconflict that reflects the reality of the ongoing offensive and defensive operations, and yet avoids the challenging question of qualifying these operations.

The cyberspace layer model developed by the US military is helpful to distinguish between different types of targets, tactics and actors. This model identified three interrelated layers with specific actors, attacks, and technologies. The physical layer corresponds to the physical IT devices and infrastructure (e.g. computing devices, storage devices, network devices, and wired and wireless links). The logic network corresponds to the logical connections between network nodes. Finally the cyber-persona consists of users, whether human or automated, as well as the content created and their behaviour (Joint Publication, 2018b). Disinformation campaigns target the top layer of cyberspace (cyber-persona layer). Because governments, companies, and civil society all use these platforms, they can be directly or indirectly affected by these efforts. Since space and time fail to exist in cyberspace, “targets
in the continental United States are just as vulnerable as in-theater targets” (Molander, Riddle, Wilson, & Williamson, 1996, p.xiii).

However, as mentioned previously, these operations can also be part of a broader strategy that includes other forms of cyberattacks that target the two other layers. In this context, cyber operations should be understood in relation to other domains of warfare and human activity (Van Puyvelde & Brantly, 2019). It is particularly the case for hybrid conflicts, where online and offline offensive operations are simultaneously conducted:

Disinformation campaigns, in particular by third countries, are often part of hybrid warfare, involving cyber-attacks and hacking of networks. Evidence shows that foreign state actors are increasingly deploying disinformation strategies to influence societal debates, create divisions and interfere in democratic decision-making. These strategies target not only Member States but also partner countries in the Eastern Neighbourhood as well as in the Southern Neighbourhood, Middle East and Africa. (EU, 2018g, p.3)

Cyberconflicts have specific characteristics that affect how and why disinformation campaigns are led in cyberspace. Lindsay (2013) recognizes three of them: First, critical economic and military infrastructure is highly vulnerable to cyberattacks, making developed states a prime target. Second, offense has become easier while defense is growing harder in cyberspace. This is due to anonymity and the fact that few organizations share information when they are attacked, that a cyber-weapon can be used, sold, and re-used numerous times before it is identified by the maker of the vulnerable technology, and that it can be patched, which makes cyberattacks fairly risk-free and accessible even without technology expertise. Third, traditional deterrence does not work in cyberspace: the attribution issue undermines deterrence. For these reasons, cyber-tools are often considered by cybersecurity scholars as “the weapon of the week” in the sense that they empower “weaker” states (i.e. with lower penetration rates and fewer military and economic resources), more so than “stronger” states (i.e. developed, military resourceful, and highly connected states) (Van Puyvelde & Brantly, 2019). In other words, connectivity should be considered simultaneously as an asset and a liability. It is precisely this dual characteristic of connectivity as an asset and a liability that makes European citizens vulnerable to disinformation campaigns.

**Disinformation Strategies**

State-sponsored disinformation campaigns pursue grand strategies where information is used to reach political and military objectives (Thornton, 2015). As mentioned previously, strategy has to do with the direction of an organization (Johnson, 2017). Clausewitz argued that “[t]he strategist must
therefore define an aim for the entire operational side of the war that will be in accordance with its purpose” (Clausewitz, 2003, p.177). To illustrate a State-sponsored disinformation, this section focuses on Russia. However, it does not imply that it is the only state conducting disinformation operations abroad. As briefly mentioned in a previous chapter, the use of information and disinformation to manipulate populations is probably as old as civilization. The choice of Russia is linked to its long history of information weaponization, which leads to a greater accessibility of the sources and analysis.

Russian and Western scholars tend to use concepts such as “hybrid conflict,” “new generation warfare,” “the Gerasimov Doctrine,” “cross-domain coercion,” and “gray zone tactics” among others (Chivvis, 2017; Adamsky, 2015; Morris et al., 2019; Galeotti, 2018; Kofman, 2016). They aim to describe the Russian understanding that modern warfare must be conducted through armed violence as well as non-military tactics (Chekinov & Bogdanov, 2015a, p.34; Chekinov & Bogdanov, 2015b; Gerasimov, 2013; Burenok, 2018, pp.61–66).

The employment of non-military measures in warfare is not a new debate among Russian military elite. For instance, a 1920 Russian military manual stated that: “Political sentiment of the population in an enemy’s rear plays a big role in an opponent’s successful activities; because of this it’s extremely important to generate sentiments among populations against the enemy and use them to organize people’s uprisings and partisan detachments in the enemy’s rear” (Shil’bakh & Sventsitskiy, 1927).

Yet, it is only since the early 2000s and the Ukraine crisis that senior Russian leaders and military theorists formed a consensus on this new conceptualization of warfare (Lilly & Cheravitch, 2020), where the line between war and peace cannot be so clearly distinguished anymore, and where the weaponization of information and cyberattacks can be as effective as violent measures (Jonsson, 2019). In 2011, Russia’s Ministry of Defense provided a clear description of its intention to weaponize information in the context of conflicts:

(…) inflicting damage to information systems, processes and resources, critical and other structures, undermining the political, economic and social systems, a massive psychological manipulation of the population to destabilize the state and society, as well as coercing the state to take decisions for the benefit of the opposing force. (Ministry of Foreign Affairs of the Russian Federation, 2011)

The 2014 Russian Military Doctrine (Russian Federation Security Council, 2014) identifies a series of geopolitical threats and the new methods that the West is deploying against the country. According to this doctrine, this new context is forcing Russia to adopt a new strategy that consists of military and non-military measures, and new and non-traditional methods (Darczewska,
Another policy document that refers to information security is the Russian Strategy for Counteracting Extremism: internet and online forums are identified as spheres of great importance for Russian security, since they can be utilized to foster ethnic, religious, and national hatred and organize terrorist activities. Most policy documents present Russia as a defensive actor (Spruds et al., 2016) and only refer to Russia’s efforts to fight “for the demilitarisation of […] the global information network, because it cannot permit the country and its surrounding areas to come under American ‘quasi-occupation’” (Darczewska, 2015). By positioning the country and its actions as a defense against the threats coming from the US, NATO and other allies, it allows Russian authorities to justify intervention in the information space of its own population. Also, the quasi-absence of any mention of offensive uses of information and psychological persuasion abroad comes from the fact that their value lies in their covert nature (Spruds et al., 2016), which renders the attribution even more challenging.

In their 2019 report “Warring Songs: Information Operations in the Digital Age,” Krasodomski-Jones, Smith, Jones, Judson, and Miller identified four strategic aims of disinformation campaigns conducted abroad:

• influence sympathetic changes in citizen behavior and perception,
• reduce the participation of one part of the population in the decision-making process,
• decrease the quality of their communications environment,
• diminish the quality of information available to citizens.

These strategic aims correspond to two grand strategies: the first two target the trust citizens have in their institutions, whereas the last two focus on the trust citizens have in news gatekeepers. A single disinformation campaign often combines several of these strategies and tactics at once. These two grand strategies and the tactics used to support them are presented in the following two subsections and will be illustrated in the third section of this chapter by disinformation operations during the 2020 Covid-19 pandemic. The examples stem from two studies conducted by the Oxford Internet Institute for the Project on Computational Propaganda.1

Disinformation Operators

Disinformation campaigns have specific characteristics. Molander, Riddile, Wilson, and Williamson (1996) identified seven key features. First, their low cost of entry allows a large number of actors to engage; an internet connection and a laptop are enough to sabotage or conduct malevolent activities. Second,
disinformation campaigns blur the lines between geographical spaces, public and private motives, warfare, and criminal conduct (Molander, Riddile, Wilson, & Williamson, 1996). They are blurry in terms of time, space, and grand strategy. Their duration is spread over a long period and is difficult to clearly define. The geographical space where they occur is not clearly delimited: they can take place cross-border by supporting opposing parties on both sides of a border for example, or by supporting the extreme political parties in several countries. And yet, although they blur the lines, they are connected to all the other domains of military operation (Gartzke, 2013), business, and society as mentioned previously.

The identification of the source of the actions is a challenge in cyberspace. This is the well-known issue of attribution and the anonymous nature of cyberattacks (Libicki, 2009). Without specific identification, hidden behind multiple fake identities, the high level of anonymity allows disinformation operators to spread messages in the dark. Yet, Bennett and Livingston (2018) identified four categories of actors who are the most susceptible of producing and diffusing false news:

- news outlets promoting radical right-wing agenda, anti-immigrant and globalist conspiracies;
- disinformation operations from foreign states targeting elections and governments;
- political parties and movements supplying party updates punctuated with “nostalgic” nationalist content, such as the Austrian Freedom Party;
- for-profit content producers that use false news to attract web traffic and make a profit from the attention economy.

Reporters Without Borders (2019) adds another category, which in fact consists of the first two identified by Bennett and Livingston (2018), and further defines “state-backed media outlets as organisations that are either directly funded by the state or are editorially managed by their governments.” Reporters Without Borders considers China, Iran, Russia, and Turkey as the countries where governments are most likely to obstruct news reporting. This control of information is not limited to internal audiences, but also to international ones, where they use their national media outlets to diffuse disinformation. The state-backed media outlets considered in this chapter are the ones with a global outreach and which target their communication on Europe.

The largest English-speaking state-backed media outlets from China, Iran, Russia, and Turkey include:

- China: China Global Television Network (CGTN)*, China Daily, China Plus, the People’s Daily, China Radio International (CRI)*, and Xinhua News Agency*
AI and the weaponization of information

- Russia: RT* (formerly Russia Today) and Sputnik*
- Iran: Mehr News and Press TV*
- Turkey: Anadolu and TRT network.* (*Some of these outlets also have services in other European languages, such as German, French, and Spanish.)

To make the situation even more complex, the for-profit false news producers often take on partisan aspects (Bennett & Livingston, 2018), since these allow them to better target the message and substantially increase web traffic. This partisan for-profit content is then picked up by bots, which can be part of systematic disinformation strategies, by either national or foreign actors.

The apolitical disinformation entrepreneurs (Bennett & Livingston, 2018), create fake stories that circulate on social media platforms, and thanks to the algorithms, and the sensationalist content and good targeting, become viral and bring traffic to the political websites they own and manage, the objective being to make a profit from the advertising that is displayed on their website. In the last US Presidential campaign, some false news stories, such as the one in which the Pope endorses Donald Trump, became viral and brought substantial traffic to these websites. Over 100 of these websites were located in Macedonia (Silverman & Alexander, 2016).

Disinformation Tactics to Erode Trust in Democratic Institutions

The first strategic aim as identified by Krasodomski-Jones, Smith, Jones, Judson, and Miller (2019) concerns affecting the link between political figures and citizens, and more precisely how citizens perceive political leadership. The intent is to increase public support for a political party or a political leader. These strategies are executed through two main tactics: false amplification (bots and fake accounts) and imposter content (Wardle, 2017) (astroturfing, impersonation). They include the amplification or fabrication of critiques (Brown, Parrish, & Speri, 2017) and trends (including conspiracy theories), the impersonation of public (Harding, 2018) and political figures (Kelly, Truong, Shahbaz, Earp, & White, 2017) and political opponents (Reporters Without Borders, 2018), and faking support from grassroots organizations (astroturfing) (Kelly, Truong, Shahbaz, Earp, & White, 2017) to show support from different sources. Although part of the support may be artificially created, these measures can also trigger real support if the audience does not realize that part of the support is fabricated.

The second strategic aim of disinformation campaigns as identified by Krasodomski-Jones, Smith, Jones, Judson, and Miller (2019) regards political participation, and more particularly reducing the participation of citizens in electoral processes in order to support political opponents. This aim is quite
broad since it includes strategies to undermine trust in democracy and electoral processes, foster polarization, and suppress voices.

In terms of tactics, it includes interfering with the political process, concentrating on/utilizing harassment and intimidation to keep some voices out of the information spaces (defamation, e.g. white trolls against journalists in Turkey; Kelly, Truong, Shahbaz, Earp, & White, 2017), and exploiting legislative systems that are not up-to-date with the large array of uses of digital technologies (dark advertising).

Two narratives support this first grand strategy of global leadership. The first one showcases the authoritarian regimes’ successful response to Covid-19, and the second one depicts how weak European pluralist democracies are.

The first narrative highlighted the successful management of the crisis by China, Turkey, and Iran. CRI in German depicted China as the leading force that drove the global economic recovery (CRI, 2020d). TRT in Spanish showcased the Turkish healthcare system as one of the best in the world to combat Covid-19 (TRT, 2020b), and alleged its international collaboration in the development of a new vaccine (TRT, 2020d), including a new radiation system being tested in the US (TRT, 2020e). HispanTV showed Iran’s support provided to Kyrgyzstan (HispanTV, 2020a) and praised Cuban doctors for allegedly taking care of over 26,000 Covid-19 patients in the world (HispanTV, 2020b). RT in German reported on an Italian businessman who changed the EU flag for the Russian one “to thank Russia for sending 12 planes [while] the EU closes everything down” (RT, 2020a).

The second narrative focused on the weak response of Western democracies to Covid-19. Sputnik and RT in French and German mentioned stories about the lockdown and civil unrest in France (RT, 2020b), Italy (RT, 2020g), Germany (Sputnik, 2020a; RT, 2020f), and Poland (RT, 2020d), including the violent protests involving the yellow vests (RT, 2020e). They also framed some “real stories” differently, including how healthcare workers in Belgium turned their backs on the Prime Minister who arrived at their hospital (RT, 2020b, 2020c), and the increase in the gap between the rich and poor in Germany (RT, 2020h). TRT in English presented homeless people in Europe (TRT, 2020a), and highlighted the difficult situation of refugees living in camps in Greece (TRT, 2020c). TRT also depicted France as a country where “discrimination is rampant” and constitutes a “societal sickness” (Ramadani, 2020). RT in English highlighted the fact that one of the main pillars of European liberalism – free movement of people – was being torn apart (Dockery, 2020).

Chinese media intended to undermine the credibility of US political leadership (CRI, 2020a, 2020e). Xinhua in French claimed that the US were diffusing a political virus (Xinhua News Agency, 2020b), and CGTN in French argued that the US staged a “Hollywood-style” show to distract from its disastrous Covid-19 management (CGTN, 2020a). CRI in German and
Spanish stated that US leadership was racist and self-serving and was bringing their country to a fall (CRI, 2020b, 2020c). CRI in German further highlighted how income inequalities among the US population (CRI, 2020g) and alleged incompetence of the US political leadership exemplified the failure of the US democratic model (CRI, 2020f, 2020g). HispanTV alleged that the Russian foreign minister initiated talks with other state representatives to clarify U.S. military and microbial activities in various regions, including near Russia's borders (HispanTV, 2020c).

**Disinformation Tactics to Erode Trust in the News Ecosystem**

The third strategic aim of disinformation campaigns as identified by Krasodomski-Jones, Smith, Jones, Judson, and Miller (2019) is to target the integrity of the communications environment itself: when compromised, anti-government protests cannot be coordinated and discourses are undermined (Shiffrin, 2014). The aim is to disrupt communication channels and create a digital environment that citizens no longer trust. In terms of tactics, it includes the abuse of content moderation, playing both sides to foster anger and confusion, fabricating and diffusing scare stories, shocking or graphic content, and dominating online discourse with hashtag poisoning and spam diffusion (Kelly, Truong, Shahbaz, Earp, & White, 2017).

The last strategic aim of disinformation campaigns as identified by Krasodomski-Jones, Smith, Jones, Judson, and Miller (2019) relates to the quality of the information citizens can access; more precisely the aim is to create information chaos, where it is not clear any longer what is true and what is false. Consequently, facts lose their value, and debate reaches an epistemic paralysis, a post-truth, or weaponized relativism (*The Guardian*, 2015). The strategies associated with this aim focus on undermining trust in media and digital media, and affecting the content produced (fabricating, suppressing, or promoting content).

To support this broad aim, a large array of tactics is available, including false news as will be discussed in the next subsection. In addition, tactics also include the exploitation and manipulation of algorithms, suppressing access to some content, and diffusing conspiracy theories. According to Nisbet and Kamenchuk (2019), another tactic is called information gaslighting. Gaslighting comes from the British play entitled *Gas Light* and its later 1940 and 1944 film adaptations, where systematically a husband psychologically manipulates his wife. In the context of disinformation, information gaslighting describes the fast production of false and contradictory information. As Adkins (2019) argues, citizens can no longer distinguish between reality and fantasy.
when exposed to the accumulation of disinformation. In other words, information gaslighting alters the target’s perception of reality:

Gaslighting exploits weaknesses in the human mind and has a debilitating effect on the victim’s ability to think rationally and to function independently of the gaslighter. It can take many forms. In all instances, however, it involves the clever manipulation of “reality” by a predator that undermines the victim’s independent mental functioning for the gaslighter’s own political, financial, or psychological motives. (Welch, 2008, p.6)

Information gaslighting is this form of disinformation campaign that targets the perception capacity of citizens: in a situation of chaos, where no one knows what is correct and what is incorrect, citizens are more easily manipulatable and vulnerable to false news. Information gaslighting confounds citizens and distracts them from what is really happening offline (Nisbet & Kamenchuk, 2019).

Another tactic used during disinformation campaigns, identified by Nisbet and Kamenchuk (2019), has to do with the incidental exposure of citizens to the large variety of forms of false news. It refers to the fact that users of social media platforms can be exposed “by accident” to false news, even when they are not the main target, for instance when false news is discussed and disseminated within a network. This is particularly prevalent on social media platforms, where echo chambers do not favor the exposure to other sources of information and pluralist views, which could contradict the false information disseminated.

Two narratives support the second grand strategy: sowing confusion. This strategy aims at diffusing conspiracy theories about Covid-19, in particular about its origin and about some of the remedies. Disinformation efforts first diffused content to cast doubt on the origin of the virus, intending to make citizens in the world doubt the official version of EU authorities and EU Member States, and European press. For instance, CRI in Spanish showed cases of Covid-19 that could not have their origin in Wuhan, such as a New Jersey mayor who was supposedly infected a long time before the epidemic started in China.

Second, the disinformation efforts intended to make citizens believe that the virus was of military origin and coming from US military bases. Iran’s Press TV claimed that the virus came from a “biowarfare” lab based in the US, and HispanTV claimed that the virus came from a US laboratory. CGTN stated in an editorial that the US military may have brought the coronavirus to Wuhan (Fuhua, 2020). CRI in German called for the US to provide an explanation about its biological laboratories in the world with military purposes (CRI, 2020a), while CGTN in Spanish wondered if the US 200+ military biolog-
ical laboratories were preparing new biological weapons and lethal viruses (CGTN, 2020c).

Sputnik in German hinted that the US was leading threatening experiments in epidemiology in Georgia, Kazakhstan, and Armenia and reported on bloggers who claimed that the US was testing bioweapons (Sputnik, 2020b). RT and Sputnik in English diffused content produced by other outlets, in particular Iranian outlets (Nimmo et al., 2020a). Russia disinformation efforts also engaged with other conspiracy theories to gain more engagement. For instance, RT in German cited an Italian politician who had asked for the arrest of Bill Gates for crimes against humanity for the role of the Bill and Melinda Gates Foundation during the Covid-19 pandemic (RT, 2020g). Fact-checkers identified a large volume of conspiracy theories distributed in the Western Balkans about the “man-made” characteristics of the virus and some “miracle cures.”

As discussed, disinformation campaigns should be considered part of a grand strategy to challenge the common understanding of the benefits, relevance, and resilience of European liberal democracies, and in doing so, contribute to a global geopolitical power play. The following subsection will present some disinformation operations.

DISINFORMATION OPERATIONS

This section focuses on disinformation operations led by two countries active in disinformation activities in Europe: China and Russia. Although they are not the only foreign sources of disinformation in Europe, these two countries have a long history of information control both domestically and internationally, which provides scholarship a larger array of data to illustrate disinformation operations. In addition, disinformation not only comes from outside Europe but also from inside as illustrated below, as is the case with some far-right movements. This section illustrates disinformation in Europe with concrete examples2 of its use to weaken EU institutions and its Member States.

Changing Narratives during Covid-19

In December 2019, Wuhan, the provincial capital of Hubei in the People’s Republic of China became the focus of a new virus: coronavirus. But it was not until 31 December that the authorities in this city published an official statement about the virus, and only in January officially admitted that the virus was transmissible to humans. The authorities placed the whole area under quarantine with the confinement of nearly 60 million people and the building of temporary hospitals to meet the needs of a massive number of sick people.
Chinese President Xi Jinping assured the population that the authorities were doing everything they could to contain the virus.

Among the first victims was a doctor (Li Wenliang) who died on 7 February 2020. He was a whistle blower and alerted the authorities to the virus. Many of his patients were infected with Covid-19. Other doctors sounded the alarm at the end of December. They circulated information in private groups on WeChat, one of the Chinese social media platforms. This equivalent of WhatsApp in China is highly controlled and monitored by the Chinese central authorities. Hence, if they circulated this information in private groups, these doctors probably knew that they could be spied on (Sautedet, 2020). When this information began to filter out, Doctor Li Wenliang was arrested by the local authorities for having, according to Beijing, spread illegal information. His death caused outrage on Chinese social media platforms (Bondaz, 2020) where millions of citizens expressed their anger and did not hesitate to denounce the false information that the Chinese government communicated. Several videos also show Chinese citizens in front of windows shouting “it is all fake” (L’OBS, 2020).

The first reaction of the Chinese authorities was to allow the explosion of anger on social media platforms. Until the end of the month of January, there seemed to be a form of tolerance in the Chinese propaganda apparatus toward news items that were quite critical of the authorities’ Covid-19 crisis response (Bondaz, 2020).

However, at the beginning of February, Chinese authorities changed their approach and again started controlling the flow of information very closely. They decided to change the narrative of what happened and their crisis management response. This was done through increased censorship in two stages. The first step of this effort was to eliminate any information that negatively portrayed the role of the public authorities. The second step was to produce and diffuse new content that depicted a positive government response to the crisis, such as the construction of hospitals in just a few days. Doctor Li Wenliang, who had been accused of fomenting a state conspiracy and harming social stability, was now recognized by the same public authorities as a hero (Sautedet, 2020). He was presented as a member of the Communist Party and a martyr. He was being used by the central authorities to stage this denial of responsibility (Bondaz, 2020).

Moreover, disinformation campaigns addressed the origin of the virus. On social media platforms, rumors about the virus coming from outside China were not eliminated, while others about the virus coming from experimental laboratories in Wuhan were systematically eliminated. Starting in mid-February, the Chinese authorities expanded their disinformation campaign to other countries, including Europe. New stories appearing in Chinese media outlets such as Global Times, and relayed by Chinese diplomats abroad, aligned with
the argument that the origin of the virus was not China (Sautedet, 2020). For example, the Chinese media Global Times implied that cases of Covid-19 had been identified in Italy as early as November, referring to a doctor, who in fact had spoken of atypical pneumopathy. He confirmed that he was opposed to this rumor and that these cases had nothing to do with Covid-19 (Bondaz, 2020). Another example of content that was spread abroad through social media platforms was the attempt by the spokesman of the Chinese Ministry of Foreign Affairs to relay rumors about an American origin of the virus, using as an argument that American soldiers had taken part in the military games in Wuhan in autumn 2019 and brought the virus to China. The objective of this information gaslighting form of disinformation campaign was clear: the Chinese authorities were trying to sow doubt and minimize their responsibility.

In this narrative abroad, the central Chinese authorities also included China’s health diplomacy, with a message to developing countries: China is an essential partner and this partner is capable of helping them when Europeans and Americans are unable to do so. In Europe, the mask diplomacy is part of the effort to replace the old narrative of mis-management of the crisis with capability and support. Images and declarations about China supplying protective equipment and detection kits flooded social media platforms. Chinese media showcased the support of China to European countries affected by Covid-19, such as the distribution of masks and respirators to Italy (Xinhua News Agency, 2020a), Spain (CGTN, 2020c), and the UK. CGTN published an article entitled “China announces to help 82 countries fight COVID-19” (CGTN, 2020b) highlighting the wide scope of China’s international aid. The Chinese People’s Daily newspaper (2020) celebrated “academicians from the Chinese Academy of Sciences and Chinese Academy of Engineering [who] have become known as ‘warriors in white’.”

The aim of this disinformation campaign was on the one hand to influence international perceptions of China, and on the other hand to challenge the social imaginary of European liberal democracies. The Chinese authorities deleted content that negatively depicts their management of the crisis, and produced new content aligned with the narrative that China is handling Covid-19 better than the Western democracies. The Covid-19 crisis and its response were an opportunity to show the successes of the Chinese political system, and the superiority of the Chinese system over European liberal democratic systems. In doing so, they hamper the social imaginary of European liberal democracies. A similar objective is pursued by Russian disinformation campaigns.

**Russian “Secondary Infektion” Disinformation Campaign**

The Oxford University computational propaganda project already back in 2017 identified Russian interference in electoral processes in different European
countries, including the United Kingdom, Germany, the Netherlands, Norway, and Sweden. The Russian hacker group called “Fancy Bear,” the same one that leaked the emails from the Clinton campaign in 2016, conducted cyberattacks against political leaders and governmental agencies in Germany, while other hackers associated with Russian intelligence gained access to data from the Bundestag.

Russian disinformation campaigns operated through troll factories, hackers, and bots to sow chaos and exploit the vulnerabilities of elections and the public sphere in democratic states (Pomerantsev, 2014) where information flows with few public gatekeepers and many communication channels. In 2013, with an annual budget of USD 10 million, the well-known Internet Research Agency in St. Petersburg, Russia, employed about 600 people (Bugorkova, 2015). Operators were assigned a specific audience, and goals with a precise number of followers to attract. For instance, they were expected to manage at least 10 Twitter handles and tweet 50 times a day on each; manage six Facebook accounts and publish at least three posts daily; and post around 50 articles per day (Bugorkova, 2015). The US Special Prosecutor Robert Mueller further documented the extent of these operations in 2018.

The resulting confusion makes the efforts of states more difficult, since they not only have to ascertain/understand/identify the reality of the field, but then share that reality with the public and businesses. In other words, disinformation campaigns both change what individuals see of reality, and they obstruct the view of states. This change is reflected in the Russian military’s understanding of the emergence of a “new generation of warfare” (voina novogo pokoleniya), and is well-illustrated by the use of information during the Russian military annexation of Crimea and Ukraine (Thornton, 2015).

Creating confusion is the first tactic of disinformation operators, whether it is in times of crisis or in areas of conflict. By spreading confusion, the state at the origin of the disinformation campaign increases the level of risk for its opponents. Another tactic used by Russia is to alter its image abroad. The Russian narrative of unpredictable leadership is a key element of Russia’s disinformation campaigns, as it feeds three other objectives. First, it triggers uncertainty about the real situation on the ground as well as Russia’s intentions. Second, it supports dissension within and among other states. Third, it contributes to the perception of a strong Russia (Thornton, 2015).

During the Paris terrorist attacks for instance, governments took time to sort through the false alerts and the real call for help and description of what was happening. The Paris crisis exhibited a large base of social media users tweeting and posting about the attack in a concentrated metropolitan area. This produced an “information cascade,” whereby platforms such as Twitter and Facebook were inundated with posts of dubious credibility, thus complicating action on the ground (Melissen & Caesar-Gordon, 2016).
In Ukraine, contradictory information about movements of Russian troops near the eastern border of Ukraine was published before and during the conflict. This effort resulted in buying time in the initial stages of the conflict by thickening the fog of war (Wirtz, 2015). The Russian government also supported bloggers and individuals to broadcast pro-Russian narratives on social media networks (Dougherty, 2014) and sometimes simulate anti-Russian news sources to disseminate false information about the ongoing conflict. “Foreign politicians talk about Russia’s interference in elections and referendums around the world. In fact, the matter is even more serious: Russia interferes in your brains, we change your conscience, and there is nothing you can do about it” (Vladislav Surkov, Adviser to Russian president Vladimir Putin, in Maza, 2019).

Information as a tool of power and control has a long history in the world, and particularly in Russia, where it is a systemic phenomenon: it has become part of Russian strategic culture (Darczewska, 2015). Russia and prior to that the USSR have long engaged in information control, manipulation and weaponization of information as discussed previously. Recent and well-documented examples include disinformation in the form of fabricated content about the plane crash MH17 in Ukraine, the 2016 EU–Ukraine Association Agreement referendum, the Crimea annexation, and more broadly the Russia–Ukraine conflict. Other well-known cases include the 2016 Brexit, 2017 French presidential election (Dreux-Vachon, 2017), and the 2017 Catalonian independence referendum. However, other smaller European countries were also affected in recent years by disinformation stemming from Russia, including Sweden (Nimmo et al., 2020b) and Czechia (Syrovátka, 2019). Russian disinformation efforts are reported to have produced 2,500 pieces of content in seven languages over 300 online platforms since 2014 (Nimmo et al., 2020b). In 2019 Facebook announced “16 accounts, four pages, and one Instagram account as part of a small network emanating from Russia” (Gleicher, 2019b). But the operation was in fact much larger and part of Moscow’s decades-long strategic engagement to sow chaos and weaken Western democracies (The Associated Press, 2020).

To illustrate, Yevgeny Primakov, former Director of the Russian Foreign Intelligence Service, recognized that its services supported the diffusion of the narrative claiming that the US Government created the AIDS virus in the 1980s (Kello, 2017). This disinformation campaign was given the name “Operation InfeKtion” by historian Thomas Boghardt (2009), although the real code name was in fact identified later as Operation “DENVER” (Selvage, 2019). The objective of this campaign was to undercut the United States’ credibility, promote anti-Americanism, and generate friction between the US and its allies. The narrative about AIDS supported the view that US military bases were the origin of the spread of the virus abroad (US Department of State,
Artificial intelligence and democracy

1987). It started with an anonymous letter sent to an Indian journal supporting this scientific claim, and was followed in 1985 by broader diffusion with the help of allied secret service agencies such as the Bulgarian Committee for State Security:

We are conducting a series of [active] measures in connection with the appearance in recent years in the USA of a new and dangerous disease, “Acquired Immune Deficiency Syndrome – AIDS”..., and its subsequent, large-scale spread to other countries, including those in Western Europe. The goal of these measures is to create a favorable opinion for us abroad that this disease is the result of secret experiments with a new type of biological weapon by the secret services of the USA and the Pentagon that spun out of control. (KGB, 1985)

A false scientific report was then diffused at the summit meeting of the Non-Aligned Movement in 1986 entitled, “AIDS: USA home-made evil, NOT out of AFRICA” (Selvage and Nehring, 2019).

But this narrative and its pseudo-scientific claims were soon denounced by Western and Soviet virologists, and by the Western press, further to letters sent to newspaper editors and journalists by US Embassy officials. Their argument was that it was not scientifically possible at the time to create such a complex virus (US Department of State, 1987). As a result of this international denunciation, the Russian authorities decided to abandon this narrative in 1987 (Andrew & Mitrokhin, 2005).

In 2019, after Facebook deleted a large number of accounts associated with Russian disinformation efforts, researchers at Atlantic Council’s Digital Forensic Research Lab (DFRLab) used the term “Secondary Infektion” to describe a new wave of false news that applied the same strategy to create false content: first create a fake account on a forum to plant a false story, often with the image of a counterfeit document to support the claim. Then create a set of fake accounts on various social media platforms to diffuse this story broadly and in different languages. The false news is used in the internet forum as the source to support the claim on other platforms (DFRLab, 2019). What differs from the first Infektion campaign, however, is that the second wave supports several stories.

Contrary to the disinformation strategy adopted during the 2016 US Presidential campaign, the Secondary Infektion aimed to better hide its identity (DFRLab, 2019): By creating a single-use burner account, publishing false news, and then abandoning it to create a new one and publish another false news story or another version of the same story, the disinformation operators not only covered their tracks better, but they also diminished their impact, since they had no time to develop an audience and outreach (The Associated Press, 2020). This second generation of disinformation campaigns continued using platforms such as Reddit, Medium, Twitter, Quora, Facebook, and
You Tube, but also increasingly used blogging forums to diffuse false news and politically explosive stories. It first diffused images of falsely “leaked” documents on blogging forums, and then spread the news on social media platforms (Nimmo et al., 2020b).

In terms of content, the disinformation operators pursued a high drama low impact strategy. Their objective was to generate an emotional response from conspiracy-minded internet communities and make the story viral. One of the stories identified by DFRLab (2019), for instance, claimed that Spanish intelligence unveiled a plot to assassinate Boris Johnson in 2018.

Although the scale of the operation is vast – a large number of channels, it promotes only nine main narratives as identified by the 2020 Graphika report (Nimmo et al., 2020b):

• Russia is the victim (USA and NATO allies are belligerent actors, Turkey is a destabilizing state, Muslims are aggressive invaders, world sporting events are Russophobic, Russia is the victim of Western plots, critics against the Russian government stem from morally corrupt, alcoholic, or mentally unstable individuals);
• Western democracies are weak (Europe is weak and divided, Western democratic elections are rigged);
• Ukraine is unreliable and a failed state.

This second wave of disinformation campaigns did not mainly focus on election interference, as often argued in the press. The objective was more about the traditional geopolitical power strategy: Divide to better conquer. It aimed to intensify divisions between Western countries, such as Poland against Germany, Germany against the USA, Germany against the UK, and everyone against Ukraine (Nimmo et al., 2020b).

Moreover, this second wave of disinformation made intensive use of counterfeit documents to support its claims, including false communications (tweets, letters, and blog posts) from political leaders of Western democracies, such as US Secretary of State Mike Pompeo, representatives of the German, British, and Ukrainian governments, and former national leaders including Carl Bildt (Sweden). The disinformation operators also counterfeited content from nonprofit organizations, ranging from the World Anti-Doping Agency (WADA) and the Organization for Security and Co-operation in Europe (OSCE) to the Committee to Protect Journalists (CPJ) and the environmental group Greenpeace (Nimmo et al., 2020b). Although this tactic is not unique, the volume, consistency, and persistence are. In that sense, it is aligned with previous Russian active measures (Rid, 2020).
This disinformation campaign has unique features, including a very limited impact in terms of engagement metrics when compared to past efforts:

Almost none of the operation’s posts across six years of activity achieved any measurable engagement, in terms of shares, likes and positive reactions across platforms. This may indicate that the operators were not interested in engagement metrics – for example, if they were driven by production quotas rather than engagement targets – or that they were using some other form of metrics not visible to outside observers. The lasting mismatch between effort expended and apparent impact gained is yet another mystery about this operation. (Nimmo et al., 2020b)

In other words, disinformation operations do not necessarily become viral. This means that it is crucial not to inflate their potential impact on political outcomes and polarization. However, their impact is difficult to assess as it can have longer-term and secondary effects, including reduced trust in public institutions, and lack of interest in political representative processes.

Russia’s disinformation efforts are now well-documented. However, it is important to avoid inflating its power more than it actually is (Győri & Krekó, 2019). The recent disinformation environment is more complex than before, with tactics and procedures different than the ones developed by the Internet Research Agency and Russia’s GRU military intelligence. Moreover, Iran is also now an active disinformation actor on Western social media platforms (Nimmo et al., 2020b), China has become more aggressive abroad (Twitter, 2019), and Western political parties also run their own disinformation campaigns (Gleicher, 2019a).

At the same time, if the actors and techniques of disinformation have evolved, the response has also improved with more researchers involved, tech companies adopting new measures, states adopting new legislations, and citizens more aware of this issue: “The repeated exposure of Secondary Infektion’s operations by platforms, journalists, and researchers may have triggered the steep drop in output observed in July 2019 and January 2020. If this model can be continued and reinforced, our collective defenses will be in a significantly better state than in 2016” (Nimmo et al., 2020b).

As discussed in this subsection, EU institutions and European Member States are the targets of large disinformation operations, whose objective is to challenge the common understanding of the benefits, relevance, and resilience of European liberal democracies. In this context, EU institutions and Member States responded gradually and with different means to this threat. The 2019 EU Parliamentary elections were particularly under scrutiny.

From January to May, online platforms have taken action against inauthentic behaviour to limit the scope of spam and disinformation globally. Google reported to have globally removed more than 3.39 million YouTube channels and 8,600 channels
As discussed, disinformation campaigns from China and Russia aim to challenge European liberal democracies. By doing so, they aim to hinder how liberal democracies and their institutions are perceived by their citizens and other countries in the world. Hence, disinformation campaigns are part of a global power play to reduce the influence and the role of liberal democracies and democratic values in the world. The following section presents the main efforts of EU institutions to combat disinformation campaigns in this context.

RESPONSE TO DISINFORMATION CAMPAIGNS

With the growth of false news and disinformation campaigns, fact-checking has become one of the objectives of online platforms, the press, and Western governments. ReportersLab identified about 160 fact-checking organizations in the world (Lim, 2019). In Europe, some organizations such as EUFactcheck.eu³ or EUvsDisinfo.eu⁴ are prime examples of the efforts of the press and EU institutions to combat the spread of false news. This point is discussed in more detail in a subsequent chapter of this book. Facing growing disapproval from their users, and a new set of regulations to force them to take down false content, online platforms turned to AI to increase their content moderation capacity and as much as possible automate it. Part of the moderation is still done by human operators based in developing countries. Automated fact-checking (AFC) technologies pursue three objectives: “to spot false or questionable claims circulating online and in other media; to authoritatively verify claims or stories that are in doubt, or to facilitate their verification by journalists and members of the public; and to deliver corrections instantaneously, across different media” (Graves, 2018, p.2). Although promising, this technology faces several challenges, since fact-checking requires judgment and sensitivity to context, which fully automated fact-checking systems cannot do. AFC are particularly challenged by conversational sources, such as discussion on social media platforms, where users use pronouns and refer back to earlier points, or use words with double meanings. So far, AFC technologies prove useful predominantly “to assist fact-checkers to identify and investigate claims, and to deliver their conclusions as effectively as possible” (Graves, 2018, p.1).

In 2015, the European Council “stressed the need to challenge Russia’s ongoing disinformation campaigns and invited the High Representative, in cooperation with Member States and EU institutions, to prepare by June an
action plan on strategic communication. The establishment of a communication team is a first step in this regard.” In response, the East Stratcom Task Force was created as part of the European External Action Service (EEAS). This task force has identified and catalogued over 9,000 examples of pro-Kremlin disinformation. The website EUvsDISINFO that the East Stratcom Task Force then created raises awareness about these “messages in the international information space that are identified as providing a partial, distorted, or false depiction of reality and spread key pro-Kremlin messages.”

The years 2016 and 2017 marked a change in the international context. Former UK Prime Minister Theresa May claimed that Russia was “weaponizing information” and the US acted on Russian interference in the US Presidential election by imposing a number of sanctions against diplomats and individuals associated with the Internet Research Agency. The 2019 Report by Special Counsel Robert Mueller stated that “The Russian government interfered in the 2016 presidential election in sweeping and systematic fashion. Evidence of Russian government operations began to surface in mid-2016.” Further, “The campaign evolved from a generalized program designed in 2014 and 2015 to undermine the U.S. electoral system, to a targeted operation that by early 2016 favored candidate Trump and disparaged candidate Clinton” (Mueller, 2019).

In 2016, the Joint Framework on countering hybrid threats “encouraged a whole-of-government approach, with 22 areas for action, to help counter hybrid threats and foster the resilience of the EU and the Member States,” and recommended a series of actions “ranging from bolstering EU’s intelligence analysis capacity to strengthening protection of critical infrastructure and cybersecurity to fighting radicalisation and violent extremism” (EU, 2016). It also led to the creation of the Hybrid Fusion Cell as the focal point for all EU institutions for the analysis of hybrid threats. It was followed a year later by the establishment in Helsinki of the European Centre of Excellence for Countering Hybrid Threats.

In June 2017, the European Parliament adopted a resolution on “Online platforms and the Digital Single Market” that “Stresses the importance of taking action against the dissemination of fake news” and “Calls on the Commission to analyze in depth the current situation and legal framework with regard to fake news, and to verify the possibility of legislative intervention to limit the dissemination and spreading of fake content” (EU Parliament, 2017). At the end of 2017–beginning of 2018, the EU Commission launched a public consultation on fake news and online disinformation with two questionnaires: “one for the citizens and one for legal persons and journalists reflecting their professional experience of fake news and online disinformation” (EU, 2018e). In a 2018 Eurobarometer survey, 73% of people interviewed expressed their
concern about the impact of false news for the upcoming European Parliament elections (EU, 2018h).

This consultation showed that “Intentional disinformation aimed at influencing elections and immigration policies were the two top categories considered likely to cause harm to society, according to respondents to a public consultation conducted by the Commission” (EU, 2018e).

A majority of respondents to the public consultation considered that educating and empowering users to better access and use online information and informing users when content is generated or spread by a bot are measures online platforms can take that would have a strong impact on preventing the spread of disinformation. (EU, 2018e)

In January 2018, the year before the EU Parliamentary elections, then EU Commissioner for the Digital Economy and Society Mariya Gabriel convened the High-Level Expert Group (HLEG) to advise the European Commission on disinformation campaigns (2018), and their recommendations were based on five pillars:

1. Augmenting transparency of online news;
2. Encouraging media and information literacy;
3. Developing new tools to empower citizens and journalists to tackle disinformation;
4. Protecting the diversity and sustainability of the news media ecosystem;
5. Promoting continued research on the impact of disinformation (HLEG, 2018).

In April, the EU Commission took on these recommendations and the results of the public consultations to develop a European approach to tackle online disinformation:

A well-functioning, free, and pluralistic information ecosystem, based on high professional standards, is indispensable to a healthy democratic debate. The Commission is attentive to the threats posed by disinformation for our open and democratic societies. This Communication presents a comprehensive approach that aims at responding to those serious threats by promoting digital ecosystems based on transparency and privileging high-quality information, empowering citizens against disinformation, and protecting our democracies and policy-making processes. (EU, 2018e)

The Communication recommended the creation of an EU-wide independent network of fact-checkers and initiatives to enhance the quality of journalism and augment the digital media literacy of citizens (Bentzen, 2019). This Communication also called for the organization of a Multi-Stakeholder Forum
on Disinformation (2018) to be composed of “representatives of online platforms, the advertising industry and advertisers, as well as academics, media and civil society organisations” (EU, 2018c).

This Forum led to the creation of an EU-wide Code of Practice to improve the explicability of information selection by algorithms and the accessibility to reliable news:

The Code of Practice on disinformation is the first worldwide self-regulatory set of standards to fight disinformation voluntarily signed by platforms, leading social networks, advertisers and advertising industry in October 2018. Signatories are Facebook, Twitter, Mozilla, Google and associations and members of the advertising industry. Microsoft subscribed to the Code of Practice in May 2019. TikTok joined the code in June 2020. (EU, 2018b)

This EU Code of Practice calls for “deleting fake accounts, labelling messaging activities by ‘bots’ and cooperating with fact-checkers and researchers to detect disinformation and make fact-checked content more visible” (Bentzen, 2019).

In June 2018, the European Council called on the EU to “protect the Union’s democratic systems and combat disinformation, including in the context of the upcoming European election” (EU, 2018c). The same month, the European Commission and the Vice-President of the Commission/High Representative of the Union for Foreign Affairs and Security Policy (HR) wrote a joint communication on boosting resilience against hybrid threats, emphasizing strategic communications as a priority (EU, 2018d).

In September 2018, on the occasion of his State of the Union Address, former EU President Jean-Claude Juncker set out a series of new measures to ensure free and fair 2019 European Parliament elections, including enhanced transparency of online political advertisements, and possible sanctions for the illegal use of personal data (EU, 2018a). The same year, the EU Commission published a Joint Communication on increasing resilience and bolstering capabilities to address hybrid threats in 2018 “to address disinformation emanating from inside and outside the EU and to deter hostile disinformation production and hybrid interference by foreign governments” (EU, 2018d).

In December 2018, the EU Commission developed an “Action Plan against Disinformation” with four pillars:

1. Additional funding, specialized staff, and data analysis tools provided to the Strategic Communication Task Forces, the EU Hybrid Fusion Cell, and the EU delegations in neighborhood countries;
2. Creation of the Rapid Alert System (RAS), a platform where EU Member States and EU institutions can share insights on disinformation in real time and coordinate responses;
3. Call to the online platform companies to effectively implement the commitments they agreed to when signing the EU-wide Code of Practice on Disinformation (signed on 26 September 2018);

4. Promotion of media literacy, creation of targeted awareness campaigns about disinformation, and support to national teams of independent fact-checkers and researchers (Bentzen, 2019).

The Commission launched the Social Observatory for Disinformation and Social Media Analysis (SOMA) facilitating networking, knowledge exchange and development of best practices among independent fact checkers. A first group of 14 European fact-checking organisations have access to SOMA, which is also launching multidisciplinary centres for research on disinformation. The Connecting Europe Facility will also provide funding (EUR 2.5 million) for a new digital service infrastructure aimed at networking fact checkers and researchers (EU, 2019).

As shown, the EU institutions and EU Member State representatives took the threat of disinformation seriously prior to the 2019 elections. The European Commissioner for Security Julian King argued in early 2019 that European elections were “Europe’s most hackable election” (Becker, 2019). Although the impacts of disinformation campaigns on electoral outcomes are still debated among experts and scholars (Tucker et al., 2018), their existence is proven by a large array of evidence, as is their purpose to influence public opinion and interfere with the outcome of democratic elections. The study conducted by Avaaz and the Institute for Strategic Dialogue showed that far-right political groups used Facebook to spread disinformation in Germany, the UK, France, Italy, Poland, and Spain prior to the elections (Institute for Strategic Dialogue & Avaaz, 2019).

The 2019 European elections illustrate well the efforts of foreign states to interfere in the democratic processes of other states (Syrovátká, 2019). The cybersecurity company SafeGuard published a study in 2019 that received substantial media attention since it claimed that “half of the citizens of EU Member states have come into contact with disinformation from Russian sources” (Boffey, 2019). The study was based on the analysis of the audience of over 6,000 bots and semi-automated accounts on Twitter, Facebook, and YouTube that were apparently connected to Russian disinformation efforts. However, the list of accounts was never published by SafeGuard, meaning that the attribution was not confirmed by other experts and scholars. Since attribution is the main challenge to assessing the impact of disinformation campaigns, this study must be considered with caution (Syrovátká, 2019).

After the elections, the EU highlighted that “The preliminary analysis shows that it contributed to expose disinformation attempts and to preserve the integrity of the elections, while protecting freedom of expression” and “Malicious
sources, both within and outside the EU, are constantly using new tactics, opting increasingly for smaller-scale local operations that are less likely to be detected and exposed. However, the objective remains the same: dividing our society and undermining the trust of citizens in democratic processes and institutions.” (EU, 2019, p.9).

CONCLUDING REMARKS

This chapter explored contemporary forms of disinformation, focusing on disinformation campaigns through social media platforms in Europe. AI is both a defense against and a weapon to support disinformation campaigns. On the one hand, AI is offered as the main solution of filtering out false news. On the other hand, AI is used to spread false news either through the MLA of social media platforms (Hao, 2021) or automated tools such as bots.

Disinformation campaigns target the trust established between citizen governments (Barela & Duberry, 2021), as well as their trust in the information ecosystem:

> [d]isinformation erodes trust in institutions and in digital and traditional media, and harms our democracies by hampering the ability of citizens to take informed decisions. Disinformation also often supports radical and extremist ideas and activities. It impairs freedom of expression, a fundamental right enshrined in the Charter of Fundamental Rights of the European Union (Charter). (EU, 2018f, p.1)

Disinformation campaigns aim to hinder how liberal democracies and their institutions are perceived by their citizens and other countries in the world. Hence, disinformation campaigns are part of a global power play to reduce the influence and role of liberal democracies and democratic values in the world. Disinformation campaigns must also be understood from this global perspective, where authoritarian regimes push the narrative on social media platforms that liberal democracies are not relevant and resilient, and focus on the challenges they present rather than on the benefits they offer to their citizens. AI is at the center of this battlefield both as an enabler of disinformation diffusion by controlling content distribution (i.e. MLA of online platforms favoring juicy content), and as a potential opportunity to mitigate their diffusion (i.e. automated fact-checking).

NOTES

1. Two memos summarize their latest findings at the time of writing this book:
Targeting French, German and Spanish-Speaking Social Media Users.” Data Memo 2020.3. comprop.oii.ox.ac.uk [Accessed 22 September 2021].
2. The examples are not exhaustive but provide a real-life illustration.

REFERENCES


AI and the weaponization of information


AI and the weaponization of information

03/11/1020600/facebook-responsible-ai-misinformation/ [accessed 22 September 2021].


TRT. (2020b, 19 May). Turquía Está a la Cabeza De Las Naciones Que Han Superado Este Proceso Con El Mínimo Daño Posible [Turkey Leads Among Nations That Have Overcome This Process With the Least Possible Damage]. https://www.trt.net.tr/espanol/vida-y-salud/2020/05/19/turquia-esta-a-la-cabeza-de-las-naciones-que-han-superado-este-proceso-con-el-minimo-dano-posible-1419911 [Accessed 22 September 2021].


