AUSTRIA INSTITUT FÜR EUROPA- UND SICHERHEITSPOLITIK

AIES-STUDIES

Arthur de Liedekerke & Michael Zinkanell

Deceive and Disrupt: Disinformation as an emerging cybersecurity challenge

Nr. 13
June 2020
Vienna

About the Authors

Arthur de Liedekerke *joined CERT-EU, the Computer Emergency Response Team for the European Union institutions, bodies and agencies, in 2018 where he is currently working as an External Affairs Officer. He previously worked in the European Parliament, advising two senior MEPs on foreign affairs and security issues. He has collaborated with a number of think tanks and strategic intelligence companies, in the United States, the United Kingdom and Belgium. He holds two masters' degrees – in international relations and geopolitics – from the University of Maastricht and King's College London and was a Visiting Policy Fellow at the Copenhagen Center for Social Data Science, University of Copenhagen.*

Michael Zinkanell *is working at the Austrian Institute for European and Security Studies (AIES) since January 2019. In his position as Research Fellow, he is focusing on cybersecurity and disinformation campaigns, European foreign and security policy, and geopolitical trends. Previously, Michael Zinkanell worked at the Austrian Ministry of Defence, analysing the conflicts in Syria and Iraq. He holds a Master's degree in International Development from the University of Vienna and a Bachelor's degree in Peace and Conflict Studies from Malmö University. During his academic and professional career, he conducted fieldwork during research trips to Uganda, Turkey, Iraq, Japan, and China.*

AIES STUDIES Nr. 13

June 2020

Vienna, Austria

Published by © Austrian Institute for European- and Security policy / AIES
Dr. Langweg 3, 2410 Hainburg/Donau, Austria
Tivoligasse 73a, 1120 Vienna, Austria
Tel.: +43 1 3583080 | Email: office@aies.at | Website: www.aies.at

AUSTRIAN INSTITUTE FOR EUROPEAN AND SECURITY POLICY

STUDIES

DECEIVE AND DISRUPT:

DISINFORMATION AS AN EMERGING CYBERSECURITY CHALLENGE

ARTHUR DE LIEDEKERKE | MICHAEL ZINKANELL

Abstract

Today's technological connectivity has opened up both extraordinary opportunities and unprecedented risks. Recent events have served as an acute reminder that digital disinformation has evolved from being a marginal nuisance to a potent tool to target people, organisations, or entire societies. Worryingly, disinformation campaigns appear to be the latest weapon in the arsenal of hostile actors carrying out cyberattacks.

Demonstrating the growing overlap between disinformation and cybersecurity is the central focus of this study. The authors notably call on the cybersecurity community to recognise disinformation as an emerging challenge and to offer solutions to address it.

This analysis, which is based on comprehensive secondary source research and first-hand expert interviews, starts by presenting an overview of the typology and nature of disinformation, followed with a novel tactical approach to online influence operations. At this stage, contextual depth is provided with relevant examples of disinformation, comparing malicious campaigns and narratives of several state actors. Lastly, the authors provide a future outlook and concrete recommendations, arguing that it is vital to implement integrated political approaches that treat the nexus of disinformation and cybersecurity holistically.

ACKNOWLEDGEMENTS

The authors would like to thank all the interviewees who contributed their time and valuable expertise as well as the reviewers who provided valuable suggestions and comments to improve this paper.

In particular, Arthur de Liedekerke would like to express his gratitude to Prof. Rebecca Adler-Nissen and Dr. Kristin Eggeling as well as the Copenhagen Centre for Social Data Science (SODAS) at Copenhagen University for hosting him for the duration of his Visiting Policy Fellowship and enabling him to work on this project.

TABLE OF CONTENTS

INTRODUCTION AND CONTEXTUALISATION

"Because audiences worldwide rely on the Internet and social media as primary sources of news and information, they have emerged as an ideal vector of information attack."

— Rand Waltzman, The weaponization of information, 2017

Disinformation is not new. Far from it. Understood as "false, inaccurate or misleading information designed, presented and promoted to intentionally cause public harm or profit"[1], it is a practice that has been tried and tested for centuries. But newfound concerns over the relationship between disinformation and its corrosive, real-world impact emerged following major elections events in the United States, the United Kingdom, and France in 2016 and 2017.

What has changed is its ubiquity and its power to shape debates and nudge public opinion at an unprecedented scale. There are many reasons behind this shift, but two intertwined causes deserve particular attention: "the digital transformation of news from offline to online distribution and the rise of social media platforms as news distribution channels."[2]

First, the speed of connectivity afforded by the internet and portable devices (laptops and mobiles) has drastically modified the ease with which (dis)information is propagated. The "increases in the volume and velocity of information (…)" rendered possible by the advent of our digital age and its quasi-instantaneous communication technologies "(…) have created a louder and more chaotic information environment."[3] Online content has the potential to reach a wide and geographically distributed audience irrespective of boundaries.

In parallel, "the role of the internet as a source of news has altered the way in which news moves from creators to consumers."[4] Low barriers to entry of online environments and social media have spurred the 'disintermediation' of our institutions and legacy media, which used to act as gatekeepers. States and traditional media organisations have seen their respective regulatory and information dissemination roles eroded, their claims to authority undercut.

Thus, we moved from an area of one-dimensional, predominantly state-owned, media to a reality that encompasses a much more diverse interconnection between actors of information. Today we observe a variety of new and old players in a multi-dimensional environment beyond the traditional state-monopoly of (dis)information. The audience is empowered to interact and plays an active role as new protagonists, facilitating the decoupling of (dis)information as an instrument exclusive to national entities. This development is compounded by the internet, involving an unprecedented set of opportunities to spread data via a multitude of platforms. Consequently, geographical and institutional boundaries are blurred as state, state-backed, and non-state actors blend together in the borderless world wide web, facilitating the 'globalisation of disinformation'.

Individuals can now access news and connect with one another directly, without the hierarchical structures that historically organised the top-down interaction between producers and consumers of information. End-users proactively seek out and publish their own stories on Twitter, Reddit and Facebook. Platform power and sweeping advertising operations have put considerable pressure on the traditional business model of journalists (print and online) who are forced to publish at ever-growing speed, sometimes at the detriment of in-depth fact-checking. With this, features of the conventional news cycle, such as quality assurance, are fading.

Human (troll farms or factories) or digital tools (bots) supercharge this phenomenon by gaming the algorithms that tech giants and social media platforms rely on. They are able to generate impressive amounts of activity to artificially amplify the popularity of their message, distort the tone or steer the direction of a conversation. Following efforts to increase user interaction, applied algorithms have been proven to aggravate social polarisation, fuel the growth of extremists' groups on social media, and contribute to the proliferation of racism as well as conspiracy theories.[5] Thus, "disinformation – and other forms of low quality, hyper-partisan, or conspiratorial content – can readily find a home on social media",[6] where it perniciously blends with the daily deluge of content produced online.

These evolutions have changed the face of disinformation. In turn, this is having an impact on cybersecurity. Digital disinformation has evolved from being a marginal nuisance into a potent tool to target people, organisations, or entire societies. Calls to view it through a cybersecurity lens are growing, not least because the pairing of multifaceted campaigns of disinformation and traditional cyberattacks is expected to continue in coming years.[7]

PAPER STRUCTURE

This paper is based on a comprehensive secondary source research and first-hand interviews with key stakeholders in the field. It relies on an extensive review of available English-language literature (news articles, independent reports and academic studies) as well as insights and observations gleaned from the authors' operational environment. These findings are supplemented by interviews with over twenty academics, representatives of international organisations, think tank researchers and practitioners from leading cybersecurity companies.

It begins with a concise overview of the typology of disinformation, describing in more detail the characteristics of state-abetted digital disinformation. The following section portrays what is arguably a relatively novel tactic in online influence operations, looking into how flooding tactics are applied differently by three state actors. Next, it examines the growing intersections between this disinformation flood and cybersecurity, as well as implications for the threat landscape. The paper concludes by briefly considering the role of emerging technologies and trends in future attempts at online manipulation.

STATE AND STATE-SPONSORED DISINFORMATION

"In this struggle information has been rendered a target, disinformation a weapon, and the internet a battlefield."

— Piret Pernik, Hacking for Influence, 2018

In this increasingly fractured and complex information ecosystem, many actors engage in disinformation for a wide variety of reasons. Disinformation sometimes originates out of gratuitous fun – like the practice of 'journobaiting' – or as a way for individuals to gain attention and status. Online communities, such as those from 4Chan and Reddit, have been known to manipulate the media simply for the 'lulz' or to revel in the chaos they create. Others, such as ideology-driven groups as diverse as the anti-vaxxer movement and climate activists have, resorted to disinformation to further their cause.

Money is evidently a motivation too. A report called "The Price of Influence: Disinformation in the Private Sector" published in late 2019 by cybersecurity firm Recorded Future demonstrates how unscrupulous companies can resort to disinformation vendors and disinformation as a service to generate content to alternatively boost their image or tarnish that of their competitors.[8] In another example of for-profit disinformation, in May 2019, "Facebook removed hundreds of inauthentic pages and other assets operated by Israeli political marketing firm Archimedes Group" that were engaging in coordinated inauthentic behaviour in the context of elections taking place in Africa and Latin America.[9] Racking up to 2.8 million followers on their pages, Graham Brookie, director of the Digital Forensic Research Lab at the Atlantic Council, described it as "a real communications firm making money through the dissemination of fake news."[10] In addition, there have been reports related to the Hong Kong protests that influential Chinese-speaking Twitter users were offered money from sources with ties to the Chinese government in exchange for spreading favourable content.[11] Companies and individuals alike can be driven by financial revenue to engage in disinformation campaigns. Nonetheless, there are no hard and fast categories in this typology – a vast grey area exists where motivations and agents may overlap. However, the more competent and committed actors tend to be state or para-state entities. This will be the focus of this report.

In this configuration, disinformation campaigns are part of wider influence operations (IO) with the intent to effectively target a specific entity or foreign population (although disinformation can and is projected on national audiences) over a sustained period of time. The resources and tools mobilised by states far outstrip those of the actors aforementioned. Disinformation is churned out at scale, injected in a coordinated and systematic manner, often across multiple platforms, and using the full spectrum of overt (official government outlets, press) and covert (Advanced Persistent Threat groups, trolls, bots) means. Controlling a web of organisations, from entities masquerading as NGOs to co-opted proxy groups, the campaigns they run will be carried out over a sustained period of time and are "invariably designed to undermine governments, to split societies, to weaken national security and to strengthen the position of the aggressor state."[12]

EVOLVING TACTICS OF STATE-SPONSORED DISINFORMATION

"In an ever-changing, incomprehensible world the masses had reached the point where they would, at the same time, believe everything and nothing, think that everything was possible and that nothing was true."

— Hannah Arendt, The Origins of Totalitarianism, 1951

A number of experts have noted a certain shift in the disinformation tactics employed by states trying to sway public opinion and perception abroad. This change is best summarised by Sean Illing in an article for Vox published on 18 January 2020: "For most of recent history, the goal of propaganda was to reinforce a consistent narrative. But zone-flooding takes a different approach: It seeks to disorient audiences with an avalanche of competing stories."[13]

Sabrina Tavernise and Aidan Gardiner, writing in The New York Times, describe how citizens "numb and disoriented by information saturation - [struggle] to discern what is real in a sea of slant, fake, and fact."[14] Likewise, in her opening speech of the conference "Disinfo Horizon: Responding to Future Threats" on 30 January 2020, European Commission Vice-President Věra Jourová declared: "A bulk of disinformation efforts aim at exactly that: to blur the lines, to polarise, to make us indifferent."[15]

It would indeed appear that, at least for some state actors, rather than insidiously pushing a carefully crafted message – a master narrative – in a consistent manner, a flurry of separate, even possibly contradicting, statements may seek to muddy the waters. This is not to say that targeted attempts to persuade or convince the foreign audience have disappeared. Nor should it be construed as disputing the existence of overarching, familiar tropes or well-articulated strategies for the states pursuing disinformation operations. This is merely to suggest that, in our contemporary 'post-truth' and perspectivist world, flooding the information space to create maximum dissonance may sometimes prove just as effective in achieving certain foreign policy goals.

Adversaries aim to saturate the targeted audience with a multitude of conflicting stories to ultimately cause confusion to the point of indifference. The public is left with the notion that factual objectivity and impartiality are an illusion, amplifying uncertainty and undermining trust. Recent disinformation cases related to the COVID-19 pandemic have shown that many contradictory narratives cause a stronger manipulative impact than single coherent messages.[16]

This carefully constructed method of deception and manipulation can be described as collective 'gaslighting'. The term is traditionally used in psychology to describe a form of psychological violence through manipulating the victim's sense of judgement, trust, and perception by constantly sowing seeds of doubt and insecurity. Ultimately aimed at delegitimising the target's personal beliefs and values up to the inability to differentiate

between truth and lies, gaslighting is caused by the systematic creation and spread of disinformation.[17]

Russia, and Kremlin-affiliated groups, have been precursors in this strategy, something RAND researchers Christopher Paul and Miriam Matthews labelled a 'firehose of falsehood' as early as 2016. According to them, the defining characteristics of such a strategy are a high number of channels and messages and a shameless willingness to disseminate partial truths or outright fictions.[18]

There are telling examples of this strategy being put to use in order to obfuscate the likely implication of the Russian government in international incidents, such as the communication surrounding the downing of Malaysian Airlines Flight 17 (MH17) or the Skripal poisoning case.[19] In their comprehensive 2018 report analysing Moscow's reaction, Dr. Gordon Ramsay and Dr. Sam Robertshaw demonstrate that, rather than counter or promote a particular perspective, state-run Russian entities flooded "people's information channels with distraction, misdirection and falsehoods"[20] with the intention to confuse and impede a coherent Western response.

By no means is Russia the sole source of online disinformation, merely one of its most prolific and competent purveyors, pioneering this particular tactic. Despite the many differences in style and objectives between their disinformation operations, as has been convincingly demonstrated by experts such as Peter Mattis, "increasingly, China has interfered in foreign states in a manner similar to Russia."[21]

Arguably, China has started to embrace a variant of the flooding practice too.[22] Significantly, a 2020 Special Report by Freedom House notes that, since 2017, "Russian-style social media disinformation campaigns and efforts to manipulate search results on global online platforms have been attributed to China-based perpetrators."[23]

Previously, Chinese disinformation followed distinctively different methods and goals, typically focusing on promoting one single narrative presenting the Communist Party in a positive light. First and foremost, Chinese state media functioned as a mouthpiece to reach the domestic audience. On the international stage, the emphasis lied on improving the Chinese image rather than destabilising the foreign information environment.[24] However, Chinese officials and diplomats recently increased their activities on social media platforms that are banned in China – often in a confrontational and aggressive manner – aiming to target the foreign audience.[25] With regards to COVID-19 disinformation, Chinese official sources and diplomats not only over-exaggerated aid and assistance to Europe to blur public perception, but they also endorsed false claims and conflicting conspiracies about the virus and its transmission.[26] This set of coordinated tactics marks the departure from Beijing's traditional approach, indicating a strategic shift in Chinese disinformation that replicates the Russian playbook.

A report looking into PRC-linked information operations on the Hong Kong protests of the Australian Strategic Policy Institute describes "one-way floods of messaging, primarily at Hong Kongers", constituting even cruder attempts than those of the Kremlin.[27] According to data

released by Twitter in the summer of 2019 when it suspended 1,000 accounts believed to be linked to Chinese state actors, "almost all of the suspended accounts were disguised as personal or corporate accounts of marketing firms, international relations experts or bitcoin enthusiasts. Others posed as Hong Kong media outlets (…). Some suspended accounts even appeared to masquerade as the work of Chinese dissidents."[28]

Likewise, many of the stories in the barrage of fake news that hit Taiwan during its 2018 and 2020 election process could allegedly be traced back to the Chinese government or affiliated organs. To a large extent, these efforts did not seek to explicitly champion mainland China views, but instead to "devalue the credibility of all sources of information"[29] and distort the debate. Some of the stories traced back to Beijing were exaggerations; others outright fabrications, including accusations that Taiwanese President Tsai Ing-wen's PhD was fake or that "pro-democracy protesters in Hong Kong had been offered money to kill police officers in suicide attacks."[30]

There are even signs that Iran, "makes systematic use of inauthentic websites and social media personas", disseminates content that is "a mirror of its state propaganda but (…) seldom wholly fabricated"[31], and may be adapting its playbook. A 2018 investigation by cybersecurity firm FireEye identified a large-scale suspected Iranian IO leveraging inauthentic websites and fake social media accounts to target audiences across several continents. Based on these findings, Facebook took down over 600 accounts and pages associated with the operation and Twitter also suspended a number of accounts. In a 2019 follow-up report into a potentially connected operation, FireEye uncovered "accounts in the network [that] also posted a small amount of messaging seemingly contradictory to their otherwise pro-Iran stances" as well as social media accounts "utilizing fake American personas that espoused both progressive and conservative political stances"[32] to build a broader audience-base and mask their influence activities online.

Similarly, in 2019, the Citizen Lab at the University of Toronto, analysed an "Iran-aligned network of inauthentic personas and social media accounts" dubbed 'Endless Mayfly' uncovering a "multi-platform, multi-narrative disinformation campaign"[33] which lead to the publication of over 135 fabricated articles. This operation used inauthentic websites and online personas to spread false and divisive information primarily in Saudi Arabia, the United States, and Israel.

The prevalence of true state-sponsored flooding disinformation campaigns or the exact exposure of foreign audiences to them remains inherently hard to ascertain. Nor should we conflate exposure to disinformation with uncritical acceptance of it.[34] Empirical evidence as to the effectiveness of these operations remains scant. What is apparent nonetheless is that these efforts do exist in sufficient amount and potency to sometimes alter the information space by successfully subverting the very idea of truth and pushing fringe ideas into the limelight of the digital public sphere.

Moreover, despite discrepancies in their 'flooding' tactics – from Russia's firehose to China and Iran's less direct and disruptive approaches to swamp the conversation – the examples

outlined above all exhibit, to varying degrees, a coordinated, high-volume and cross-platform approach with little to no commitment to truth. They also share the characteristic, apart from Iran, of being the preferred tactic in reaction to a precipitating crisis, an event which these regimes had likely not anticipated.

It is a well-known reality that malign actors learn from one another, recycling methods so that the challenge transcends Russia or any single state. Worse still, it could be argued that all such campaigns have mutually reinforcing effects. Similar or repurposed influence tactics can be deployed irrespective of the particular political or ideological goals being pursued. As Peter Pomerantsev and Michael Weiss remark: "We stand before a deluge of disinformation—the Kremlin's use of disinformation is, and will be increasingly, used by other states."[35]

At the very least, disinformation, including the flooding tactic, has a corroding impact, contributing to the perception that the integrity of all media is equally questionable, that the entire information space is polluted. It would appear that the socio-technological conditions of our digital age are present to allow that pollution to fester. In 2019, Reuters Institute Digital News Report found that "across all countries, the average level of trust in the news in general is down 2 percentage points to 42% and less than half (49%) agree that they trust the news media they themselves use."[36] Simultaneously, over 80% of respondents to a Eurobarometer survey on online disinformation perceive fake news a problem in their own country and to democracy in general.[37]

The feeling that "there are no facts, only interpretations" is reinforced by prominent figures regularly lambasting legitimate outlets, making it "easier for saboteurs and hostile foreign powers to inject forged news into the confusing swirl of disinformation that destroys democracies from within."[38]

DISINFORMATION AND CYBERSECURITY: A GROWING OVERLAP

"In other words, we continue to focus on the walls of the castle, while our enemies are devising methods to poison the air."

— Philipp Lohaus, From cybersecurity to information warfare, AEI, 2017

The intersections of disinformation with cybersecurity are manifold. Two aspects however are of particular interest here.

State-sponsored disinformation campaigns can be executed fully or partially in and through the digital tools afforded by cyberspace.

It is abundantly clear that cybersecurity has a role to play in curbing the tide of disinformation on a technical level. Numerous studies have focused on ways in which cybersecurity principles, best practices or tools can prove useful in tackling disinformation, in fields as different as health and election processes. From building a defensive perimeter around an infrastructure to prevent website hijacking to monitoring spoofed company domain names or assisting in the takedowns of malevolent social bots, cybersecurity can help mitigate part of the threat.

Yet, the impact of disinformation on cybersecurity can manifest itself in less obvious ways. In this context, social media accounts have become high-value targets for states, intent on seeing their propaganda suffuse. Using social engineering, honeypots and phishing campaigns, malign actors increasingly seek to compromise the real accounts of users.

Real accounts present an undeniable advantage. They are legitimate, as opposed to impersonated, and "the more legitimate an account appears, the more likely that the message will get amplified."[39] This is precisely what happened when Russia was caught "testing new disinformation tactics in an enormous Facebook campaign"[40] in African countries including Mozambique, Cameroon, Sudan and Libya late last year. Nathaniel Gleicher, Facebook's head of cybersecurity policy, said "some of the Russian-run pages and groups also used compromised Facebook accounts that once belonged to real people but had been stolen and repurposed by hackers."[41]

What is even better than a real account? A (Twitter) verified one. Due to their influential nature and the trust-based system underlying social networks, they can be efficiently leveraged to snowball content in an organic way. Attacks aiming for verified Twitter accounts were reported in 2017 by the digital rights group Access Now, who dubbed the technique the 'DoubleSwitch'. The group indicated that hackers targeted accounts of journalists and human rights activists in Venezuela and Myanmar, some of which were verified and "had a large following", in order to "spread false information."[42]

Another case of investigative journalism by the independent newsroom ProPublica, unveiled that hacked Twitter accounts were instrumentalised to spread disinformation about the COVID-19 outbreak and the Hong Kong protests. Among the 10,000 suspected fake Twitter accounts that were traced between August 2019 and January 2020, the investigation found hacked profiles from users around the world that were "involved in a coordinated influence campaign with ties to the Chinese government".[43]

Such examples illuminate the symbiotic correlation between cyberattacks and disinformation campaigns, with the purpose to destabilise adversaries with non-traditional, hostile means.

By helping to create a saturated information ecosystem, malicious states are looking to target the non-technical vulnerabilities of an organisation.

The steady decay of trust and the distorted perception of reality discussed previously could have serious implications from a cybersecurity perspective. Although the hard evidence as to this is only progressively emerging, given the speed at which threat actors reinvent themselves, it is necessary to hypothesise likely evolutions in order to pre-empt all possible attack vectors. Two such scenarios are reviewed below.

Cognitive hacking

Cyberspace is commonly accepted to be composed of three distinct but closely interrelated layers: a physical, a logical and a cognitive or social layer. The physical layer comprises geographic features and physical network components like hardware, wires and cables. The logical layer is where data resides in digital format and it is responsible for routing data packages to their final destinations. Finally, the social layer consists of the information and human beings that interact with it.

This is a clear reminder that cyberspace encompasses more than hard assets, networks and computing devices. Despite the mantra of 'the human as our weakest link', we fail to properly buttress the 'cognitive security' of the organisations we seek to defend.

According to Ofcom, in 2019, 66% of adults in the United Kingdom reported using online sources for their news, placing it second behind television. Of this proportion, 49% named social media as their preferred platform.[44] With citizens increasingly turning to the internet to share and seek information, it appears we could be inherently psychologically ill-equipped to deal with the complexities and pitfalls of the digital news ecosystem.

Indeed, malicious actors can readily exploit a person's cognitive vulnerabilities and dependencies, gaming our inclinations to react to eye-catching pictures or sensationalist headlines, and leverage the social media algorithms that favour viral (regardless of newsworthiness or truth) stories. As a result, content that is attention-grabbing, popular or elicits high-arousal emotions is more susceptible to inspire consumer engagement. A situation

of online information consumption that Stanford University psychologist Sam Wineburg sums up as follows: "We are all driving cars, but none of us have licenses."[45]

These innate predispositions become particularly salient in the case of world events. "Unconstrained by reality or facts", disinformation can be "crafted with considerable latitude to appeal to hopes, fears, wishes and curiosity" and tap into the media hype around crises to lure readers in.[46] Speaking to US political news-site The Hill, Steve Grobman, the senior vice president and chief technology officer at computer security company McAfee, confirmed that "there is clearly a trend of using topics that have high levels of emotional sensitivity in disinformation campaigns."[47]

The recent outbreak of the COVID-19 pandemic has coincided with an unprecedented spike in disinformation campaigns, capitalising on negative repercussions coupled with a high uncertainty as well as the fear and urgency around its spread. Clickbait-style content was carefully curated to prey on our gullibility and deep human instincts.

The physical transmission of the virus globally seemed to mirror the evolution of Covid-19 related cases of disinformation. The first disinformation cases focusing on conspiracies connected to the outbreak in Wuhan shifted their narratives in relation to the geographic spread, however, once European countries were affected and the first quarantine measures were introduced, the contents transferred to match with national and local discourses.[48]

As infections have surged, scammers have started flooding platforms from TikTok to YouTube with weaponised information, sending alarmist messages or emails alluding to health recommendations laced with malicious software to targets across the world.[49] Cybersecurity company CheckPoint reported that since January 2020, "over 4,000 coronavirus-related domains registered globally" and that these were "50% more likely to be malicious than other domains."[50]

As of the writing of this report, it would appear that a significant amount of COVID-19-related disinformation has been the work of cyber criminals and fraudsters. But state actors have piggybacked on this. A senior US State Department notably accused Russia of spreading falsehoods about the coronavirus during an appearance before a Senate subcommittee on 5 March 2020. Lea Gabrielle, the head of the State Department's Global Engagement Centre, commented that Russian disinformation efforts included "state proxy websites, official state media, as well as swarms of online, false personas pushing out false narratives".[51] Further, also Chinese officials have engaged in the spread of COVID-19 disinformation in the form of conflicting statements related to the origins of the virus and defaming European healthcare personnel.[52]

Several reports also point to a number of Advanced Persistent Threat (APT) groups, using COVID-19 themed lures to deploy their malware. As a blog post from the Director of Threat Research at Bitdefender explains, a "lure based around fake news has a significant chance of undermining targets' mental defenses and cyber-hygiene training (...). Victims interact with news lures for several reasons, which include a drive to be 'up-to-date' or current; a sense of urgency; socio-political polarization; curiosity; or fear."[53]

Impaired situational awareness

Agents of disinformation, by polluting the information, make it difficult and time-consuming to identify who to trust or distrust online. This production of confusion is something that state-backed APT groups, the most tenacious adversaries from a cybersecurity standpoint, have greatly contributed to. There are the infamous election interference cases mentioned in the introduction, notably the 2016 breach of the Democratic National Committee (DNC). There, a likely Russian-affiliated actor, alternatively dubbed Fancy Bear or APT28, resorted to phishing operations to acquire a trove of confidential documents.[54] These were subsequently altered and released in order to seed disinformation and in the hopes of manipulating voter sentiment – a technique known as 'tainted leaks'.

At the organisational level, a corrupt information ecosystem with massive data-dumps containing erroneous material, such as the one described above, can frustrate the efficiency of cyber threat intelligence (CTI). In an interview, Citizen Lab's Senior Researcher John Scott-Railton highlighted that "Part of the power of tainted leaks is that they make some think 'Hey, this might be true.'"[55] With analysts also looking to open-source information and social media as a source of data to provide situational awareness and decision-support, not being able to count on the validity and trustworthiness of the information underlying the decision-making process can be conducive to policy-paralysis.

More specifically, a skilled threat actor of the kind this paper concerns itself with could resort to disinformation to produce 'deception-in-depth' when conducting cyber-attacks. These steps to further confuse or mislead defenders, to obfuscate the true motives or entities behind the attack, could conceivably impair situational awareness, leading to slower response times or incorrect assessments of the threat.

Here, disinformation gambits can serve to cast doubt on the findings of investigation, just sufficiently to provide plausible deniability to the attacker. This goes to the heart of the attribution problem. Technical attribution, understood as the capability to pinpoint the perpetrator behind an attack, is inherently difficult due to the many existing ruses to evade detection, such as spoofing Internet Protocol (IP) addresses. It is nevertheless a key element in CTI activities, one that serves to establish deterrence and accountability; it is equally crucial to adopt the best technical and tactical tools to prevent, detect and respond to a particular threat actor.

To successfully hide their traces, state actors deploy various proxies, covering up their tracks so as to avoid direct associations to disinformation and cyberattacks. These proxy groups vary in their level of professionalisation and operational capabilities, ranging from intelligence agencies to institutionalised networks of state-sponsored experts and single individuals. In particular, the Russian state has a long history of deploying so-called web brigades and troll farms, who design and spread false and misleading information online. The most infamous group operating from Russia is the Internet Research Agency (IRA) based in Saint Petersburg.

Since 2013, the group operates with a monthly budget of approximately one million Euros, employing 80 to 100 professionals to activate false identities on social media and disseminate disinformation in a coordinated manner.[56,57]

Add disinformation tactics to the already challenging activity of attribution and it can be used as an additional false-flag to "muddy specific attribution claims, leaving an electorate exposed to the coexistence of 'multiple truths' and a fractured narrative of the past."[58] The Sony Pictures Entertainment (SPE) breach in November 2014 is a notable case in point. Following a crippling attack on SPE, an attribution debate played out over several weeks, pitting the Obama administration against vocal sceptics in the cybersecurity community.

On the one hand, the US government rapidly mounted a case pointing to North Korea as the source of the attack, allegedly acting in retaliation to the release of a movie parodying the country's political establishment. On the other hand, prominent researchers alternatively fingered hacktivists, Sony insiders and Russia. A significant part of the confusion was linked to the public persona that initially claimed responsibility for the breach, a group calling itself 'Guardians of Peace' (GOP), that demanded monetary compensation.

An authoritative report into the incident by a coalition of security companies, including Novetta, Symantec, AlienVault and Kaspersky, suggested the Lazarus Group, an allegedly North Korean threat actor, was behind the hack and the bogus GOP group. "I think they are quite willing to dispose of identities and use a certain amount of disinformation in their campaigns, which is one the reasons I think that the security research community has a hard time, until now, with clustering all of this activity and understanding that it is all inter-related" noted Juan Andres Guerrero-Saade, senior security researcher with Kaspersky Lab's Global Research and Analysis Team.[59] The incident demonstrates "how the digital nature of a cyberattack permits the attacker to leave a smoke screen of disinformation and false clues, making attribution problematic."[60]

LOOKING TO THE FUTURE

"For people who think disinformation and cybersecurity aren't related, things are going to change very quickly for you. The issue of disinformation is increasingly becoming something that security teams are expected to address."

— Melanie Ensign, Head of Security and Privacy Communications, Uber, 2020

A number of looming challenges, some of them technological, others societal, will further blur the line between disinformation and cybersecurity.

Advances in artificial intelligence will increasingly be used "to create sophisticated digital deception."[61] Readily available software has already democratised the possibility of producing synthetic media, sometimes referred to as 'deepfakes', highly convincing but false image and audio recordings. Manipulated videos of former US President Barack Obama and Mark Zuckerberg, uttering sentences they had never truly spoken, went viral.

Meanwhile voice emulators have proved sufficiently convincing to dupe people as to the identity of their interlocutors. In March 2019, cyber criminals successfully used "artificial intelligence-based software to impersonate a chief executive's voice and demand a fraudulent transfer of €220,000 (…)."[62] It is a short step from there to forging the speech of a prominent public figure.

In a similar vein, 'fake' writing, with developments in the field of Natural Language Processing (NLP), will lead to con artists or trolls being able to mimic the style of reputable news outlets or influential public figures. The New Yorker carried out an experiment on this very topic, by having OpenAI's, an artificial-intelligence company, GPT-2 system co-write part of an article by generating coherent paragraphs matching the theme and tone of the human author.[63] The firm has declined to release the trained model "due to our concerns about malicious applications of the technology."[64]

Indeed, tools such as these add "scale, language fluency, and the ability to mirror the jargon and writing style of any profession or, with enough samples, any individual"[65], making these developments very appealing to con artists and trolls alike. As Renee DiResta points out, "one of the most challenging shifts […] will occur when automated accounts –bots – begin to do a passable job of presenting themselves as human when chatting with people. We know that this is coming (…)."[66]

As the old proverb goes, seeing (and possibly also hearing in the cases presented above) is believing. Increasingly realistic and difficult-to-detect, doctored footage or textual content will considerably complicate our ability to separate fact from fiction and play right into the hands of disinformation actors – state or otherwise.

At the societal level, our propensity to share many of our habits and hobbies online will add to the trove of personal details available on the web. Basic cyber hygiene practices, like enabling two-factor authentication or a security-conscious approach to one's digital footprint, are not necessarily engrained among the general population. It represents a boon for attackers and disinformation agents who can harvest the information readily available on social media to micro target an audience.

In 2019, a research team from NATO Strategic Communications Centre of Excellence conducted an experiment in support of a military exercise in an Allied country, whereby using open-source intelligence techniques, setting up honey-pot pages[i] and social engineering, they tried to gather information on and influence military personnel. Having managed to engage a number of soldiers on social media, the team was able to "identify all members of certain units, pinpoint the exact locations of several battalions, gain knowledge of troop movements" and even "to instil undesirable behaviour during the exercise."[67] These findings highlight two things. First, it demonstrates how a malicious actor can harvest personal data online to create actionable intelligence, spread corrupt information and influence attitudes. Second, it shows that social media companies are still having a hard time cracking down on networks of fake accounts and lagging behind in implementing necessary (some of them in terms of cybersecurity, others privacy-related) countermeasures.

i Pages designed to lure or attract a certain group of people.

RECOMMENDATIONS

In this perpetual cat-and-mouse game between defenders and attackers, governments, academic institutions, tech firms, and non-profits need to work together. Current efforts are too disjointed between various stakeholders and often reactive, a "whack-a-troll" approach that does not sufficiently address digital literacy or imposing costs on purveyors of disinformation.[68] Several recommendations can help remedy this.

It needs to start with the community recognising disinformation as "a new breed of cybersecurity threats"[69] – and cybersecurity as a mitigating measure. **Cybersecurity has the power to keep tabs on actors engaging in disinformation**, to get rid of bogus websites or accounts that have connections to criminals or malevolent foreign-backed enterprises, as well as to ensure people are equipped with the appropriate skills or tools to discern between click-bait-style disinformation and legitimate content. Naturally, a tech-centric solution is not sufficient.

It requires resolute action at the political level. Here, the authors welcome the extensive efforts to mitigate potential hybrid threats that European stakeholders have undertaken in the last years, such as the EU Action Plan against Disinformation[70] entailing inter alia the stronger collaboration with social media providers as outlined in the Code of Practice on Disinformation,[71] the European Cybersecurity Act and the mandate extension of the European Union Agency for Cybersecurity (ENISA),[72] as well as the EU Cyber Diplomacy Toolbox[73] and its enhanced framework enabling to impose sanctions and defensive cyberattacks.[74] It is crucial to **implement integrated political approaches that treat the nexus of disinformation and cybersecurity holistically**. All future efforts to mitigate hybrid threats should therefore complement each other, following a comprehensive coordinated European strategy.

In tackling the COVID-19 outbreak and the "infodemic"[75] that ensued, the EU has proven it can act rapidly and collectively. Initiatives taken in this space, including the voluntary Code of Practice, have achieved their first tangible results. Online platform providers made sudden adjustments to tackle disinformation, demonstrating that the private sector can rise to the occasion with the right amount of political pressure.

However, the business model of several social media platforms still favours the creation and distribution of manipulative and sensationalistic content. It will require an ever-closer collaboration with these providers who, through their contemporary role in the public sphere, **should open themselves up to more accountability**. This should go hand in hand with an EU-wide commitment "to resolutely counter disinformation with transparent, timely and fact-based communication on their actions"[76] and call-out disinformation aimed at the European Union.

The Commission has named China for the first time as a source of harmful and manipulative disinformation in a high-level report,[77] despite harsh warnings from Beijing against releasing

these accusations. In the future, **the European Union should not refrain from openly and publically revealing the tactics, aggressors, and potential security implications of hybrid threats – akin in general and disinformation in particular**. In order to further build up strategic autonomy and to transparently inform the general public, the EU must not give in to such kinds of threats or warnings.

Lastly, the global pandemic has further proven that the fluid digital space requires a constant redevelopment of strategies, actions, and policies. **The ever-changing online environment thus calls for a thorough analysis, evaluation, and adaptation of political measures in close collaboration with the civil society and the private sector**, in order to safeguard political resilience and the rights and freedoms of the European community.

About the Institute

For over 20 years (est. 1996), the Austrian Institute for European and Security Policy (AIES) is researching on various issues regarding the European Union and the European Integration Process. In doing so, the focus lies on European Security and Defence Policy with regards to the European Neighbourhood Policy, transatlantic relations as well as the Austrian Foreign-, European and Security Policy.

The AIES is not only doing research on these subject areas, but also provides demand-oriented analysis and recommendations for contract partners It is also organizing internal workshops, training courses and public discussions. In the course of this interdisciplinary work the AIES team of experts is using a network of research, political, economic, social and military institutions, facilities of the civil society, the media as well as international organisations.

https://www.aies.at

AUSTRIAN INSTITUTE FOR EUROPEAN AND SECURITY POLICY

BIBLIOGRAPHY

[1] High Level Group of Experts on Fake News and Online Disinformation. (2018). *A multi-dimensional approach to disinformation*. Retrieved from https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

[2] Martens, B., Aguiar, L., Gomez-Herrera, E., & Mueller-Langer, F. (2018). *The digital transformation of news media and the rise of disinformation and fake news*. Retrieved from https://ec.europa.eu/jrc/communities/sites/jrccties/files/dewp_201802_digital_transformation_of_news_media_and_the_rise_of_fake_news_final_180418.pdf

[3] Lin, H. (2019). *The existential threat from cyber-enabled information warfare*. *Bulletin of the Atomic Scientists, 75*(4), 187–196. https://doi.org/10.1080/00963402.2019.1629574

[4] Knight Commission on Trust, Media and Democracy. (2019). *Crisis in Democracy: Renewing Trust in America*. Retrieved from the Aspen Institute website: https://csreports.aspeninstitute.org/documents/Knight2019.pdf

[5] Wall Street Journal. (2020). *Facebook Executives Shut Down Efforts to Make the Site Less Divisive* https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499

[6] The Kofi Annan Commission on Elections and Democracy in the Digital Age. (2020). *Protecting Electoral Integrity in the Digital Age.* Retrieved from the Kofi Annan Foundation website: https://www.kofiannanfoundation.org/app/uploads/2020/01/f035dd8e-kaf_kacedda_report_2019_web.pdf

[7] Accenture Security. (2019). *Cyber ThreatScape Report*. Retrieved from https://www.accenture.com/_acnmedia/pdf-107/accenture-security-cyber.pdf

[8] Recorded Future. (2019). *The Price of Influence: Disinformation in the Private Sector*. Retrieved from https://go.recordedfuture.com/hubfs/reports/cta-2019-0930.pdf

[9] The Atlantic Council's Digital Forensic Research Lab. (2019, May 25). Inauthentic Israeli Facebook Assets Target the World. *Medium*, Retrieved from https://medium.com

[10] Debre, I., & Satter, R. (2019, May 16). Facebook busts Israel-based campaign to disrupt elections. *The Associated Press*, Retrieved March 23, 2020, from https://apnews.com/7d334cb8793f49889be1bbf89f47ae5c

[11] ProPublica. (2020). *How China Built a Twitter Propaganda Machine Then Let It Loose on Coronavirus*, https://www.propublica.org/article/how-china-built-a-twitter-propaganda-machine-then-let-it-loose-on-coronavirus

[12] Bugajski, J. (2019). *The Geopolitics of Disinformation*. Retrieved from The Center for European Policy Analysis StratCom Program website: http://infowar.cepa.org/The-geopolitics-of-disinformation

[13] Illing, S. (2020, February 6). *The impeachment trial didn't change any minds. Here's why*. Vox, Retrieved February 24, 2020, from https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation

[14] Gardiner, A., & Tavernise, S. (2019, November 18). *"No One Believes Anything": Voters Worn Out by a Fog of Political News*. The New York Times, Retrieved from https://www.nytimes.com

[15] Jourová, V. (2020). *Opening speech of Vice-President Věra Jourová at the conference "Disinfo Horizon: Responding to Future Threats"* [Transcript]. Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_20_160

[16] EUvsDisinfo. (2020). *The Kremlin and Disinformation about Coronavirus*, https://euvsdisinfo.eu/the-kremlin-and-disinformation-about-coronavirus/?highlight=conflicting

[17] Zinkanell, M. (2020). *Disinformation during Covid-19 from a European Perspective*, Austrian Institute for European and Security Policy (AIES), FOKUS 3/2020, 3 May 2020, https://www.aies.at/publikationen/2020/fokus-20-03.php

[18] Paul, C., & Matthews, M. (2016). *The Russian "firehose of falsehood" propaganda model*. Rand Corporation, 2-7. Retrieved from: https://www.rand.org/pubs/perspectives/PE198.html

[19] Ramsay, G., & Robertshaw, S. (2019). Weaponising news: RT, Sputnik and targeted disinformation. King's College London. The Policy Institute Centre for the Study of Media, Communication & Power. January 2019. https://www.kcl.ac.uk/policy-institute/assets/weaponising-news.pdf

[20] Anderson, J., & Rainie, L. (2017). *The future of truth and misinformation online*. Pew Research Center, 19. Retrieved from: https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/

[21] Nemr, C., & Gangware, W. (2019). *Weapons of Mass Distraction: Foreign State-Sponsored Disinformation in the Digital Age*. Retrieved from the United States Department of State website: https://www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distraction-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf

[22] Jeangène Vilmer, J.-B., & Charon, P. (2020, January 21). *Russia as a Hurricane, China as Climate Change: Different Ways of Information Warfare*. War on the Rocks, Retrieved February 26, 2020, from https://warontherocks.com/2020/01/russia-as-a-hurricane-china-as-climate-change-different-ways-of-information-warfare/

[23] Cook, S. (2020). *Beijing's Global Megaphone*. Retrieved from The Freedom House website: https://freedomhouse.org/report/special-report/2020/beijings-global-megaphone

[24] AXIOS. (2020). *China takes a page from Russia's disinformation playbook*, https://www.axios.com/coronavirus-china-russia-disinformation-playbook-c49b6f3b-2a9a-47c1-9065-240121c9ceb2.html

[25] Ohlberg, M. (2019). *Propaganda beyond the Great Firewall,* Merics, Mercator Institute for China Studies, https://www.merics.org/de/china-mapping/propaganda-beyond-the-great-firewall

[26] EUvsDisinfo. (2020). *EEAS Special Report Update: Short Assessment of Narratives and Disinformation around the Covid-19 Pandemic*, 01.04.2020, https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic/

[27] Uren, T., Thomas, E., & Wallis, J. (2019). *Tweeting Through the Great Firewall: Preliminary Analysis of PRC-linked Information Operations on the Hong Kong Protests*. Retrieved from the Australian Strategic Policy Institute: https://www.aspi.org.au/report/tweeting-through-great-firewall

[28] Feng, E. (2019, August 20). *How China Uses Twitter And Facebook To Share Disinformation About Hong Kong.* National Public Radio, Retrieved February 26, 2020, from https://www.npr.org/2019/08/20/752668835/how-china-uses-twitter-and-facebook-to-share-disinformation-about-hong-kong

[29] Watts, J. (2018). *Whose truth? Sovereignty, disinformation, and winning the battle of trust.* Retrieved from The Atlantic Council website: https://www.atlanticcouncil.org/in-depth-research-reports/report/whose-truth-sovereignty-disinformation-and-winning-the-battle-of-trust/

[30] Su, A. (2019, December 16). *Can fact-checkers save Taiwan from a flood of Chinese fake news?* The Lost Angeles Times, Retrieved March 2, 2020, from https://www.latimes.com/world-nation/story/2019-12-16/taiwan-the-new-frontier-of-disinformation-battles-chinese-fake-news-as-elections-approach

[31] Brooking, E., & Kianpour, S. (2020). *Iranian digital influence efforts: Guerrilla broadcasting for the twenty-first century.* Retrieved from The Atlantic Council website: https://www.atlanticcouncil.org/in-depth-research-reports/report/iranian-digital-influence-efforts-guerrilla-broadcasting-for-the-twenty-first-century/

[32] Revelli, A., & Foster, L. (2020, February 12). *"Distinguished Impersonator" Information Operation That Previously Impersonated U.S. Politicians and Journalists on Social Media Leverages Fabricated U.S. Liberal Personas to Promote Iranian Interests*. Retrieved March 5, 2020, from https://www.fireeye.com/blog/threat-research/2020/02/information-operations-fabricated-personas-to-promote-iranian-interests.html

[33] Lim, G., Maynier, E., Scott-Railton, J., Fittarelli, A., Moran, N., & Deibert, R. (2019, May 14). *Burned After Reading: Endless Mayfly's Ephemeral Disinformation Campaign.* Retrieved March 26, 2020, from https://citizenlab.ca/2019/05/burned-after-reading-endless-mayflys-ephemeral-disinformation-campaign/

[34] Guess, A. M., Lockett, D., Lyons, B., Montgomery, J. M., Nyhan, B., & Reifler, J. (2020). "Fake news" may have limited effects beyond increasing beliefs in false claims. Harvard Kennedy School Misinformation Review, *1*(1). Retrieved from: https://misinforeview.hks.harvard.edu/article/fake-news-limited-effects-on-political-participation/

[35] Ehrett, J. S. (2017). *Confronting Disinformation Warfare*. Retrieved from the Yale Journal of Law & Technology website: https://yjolt.org/blog/confronting-disinformation-warfare

[36] Newman, N., Fletcher, R., Kalogeropoulos, A., & Kleis Nielsen, R. (2019). *Reuters Institute Digital News Report 2019*. Retrieved from the Reuters Institutie website: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_0.pdf

[37] European Commission. (2018). *Fake News and Disinformation Online,* Flash Eurobarometer 464, April 2018, https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/instruments/flash/yearFrom/2018/yearTo/2020/surveyKy/2183

[38] Bugajski, J. (2019). *The Geopolitics of Disinformation*. Retrieved from The Center for European Policy Analysis StratCom Program website: http://infowar.cepa.org/The-geopolitics-of-disinformation

[39] Rashid, F. Y. (2019, August 29). *Disinformation Attacks Aren't Just Against Elections*. Retrieved March 6, 2020, from https://duo.com/decipher/disinformation-attacks-aren-t-just-against-elections

[40] Alba, D., & Frenkel, S. (2019, October 30). *Russia Tests New Disinformation Tactics in Africa to Expand Influence*. The New York Times, Retrieved March 4, 2020, from https://www.nytimes.com/2019/10/30/technology/russia-facebook-disinformation-africa.html

[41] Alba, D., & Frenkel, S. (2019, October 30). *Russia Tests New Disinformation Tactics in Africa to Expand Influence*. The New York Times, Retrieved March 4, 2020, from https://www.nytimes.com/2019/10/30/technology/russia-facebook-disinformation-africa.html

[42] Access Now - Digital Security Helpline. (2017). *The "Doubleswitch" social media attack: a threat to advocates in Venezuela and worldwide*. Retrieved from https://www.accessnow.org/doubleswitch-attack/

[43] ProPublica. (2020: *How China Built a Twitter Propaganda Machine Then Let It Loose on Coronavirus*, 26.03.2020, https://www.propublica.org/article/how-china-built-a-twitter-propaganda-machine-then-let-it-loose-on-coronavirus

[44] Jigsaw Research. (2019). *News Consumption in the UK: 2019*. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0027/157914/uk-news-consumption-2019-report.pdf

[45] Steinmetz, K. (2018, August 9). *How Your Brain Tricks You Into Believing Fake News*. Time, Retrieved March 5, 2020, from https://time.com/5362183/the-real-fake-news-crisis/

[46] Choy, M., & Chong, M. (2018). Seeing *Through Misinformation: A Framework for Identifying Fake Online News*. Retrieved from: https://arxiv.org/ftp/arxiv/papers/1804/1804.03508.pdf

[47] Miller, M. (2019, November 14). *Veterans face growing threat from online disinformation*. The Hill, Retrieved March 5, 2020, from https://thehill.com/policy/technology/470390-veterans-face-growing-threat-from-online-disinformation

[48] EU Disinfo Lab. (2020). *Covid-19 Disinformation: Narratives, Trends, and Strategies in Europe*, 02.04.2020, https://www.disinfo.eu/publications/covid-19-disinformation-narratives-trends-and-strategies-in-europe/

[49] Robinson, T. (2020, March 6). *Coronavirus sparks phishing, disinformation, tabletop exercises and handwashing*. SC Media, Retrieved March 6, 2020, from https://www.scmagazine.com/home/security-news/government-and-defense/coronavirus-sparks-phishing-disinformation-tabletop-exercises-and-handwashing/

[50] Check Point. (2020, March 5). *Update: Coronavirus-themed domains 50% more likely to be malicious than other domains*. Retrieved March 5, 2020, from https://blog.checkpoint.com/2020/03/05/update-coronavirus-themed-domains-50-more-likely-to-be-malicious-than-other-domains/

[51] Blake, A. (2020, March 6). *Russia behind coronavirus disinformation spreading online: State Department official*. The Washington Times, Retrieved March 6, 2020, from https://www.washingtontimes.com/news/2020/mar/6/russia-behind-coronavirus-disinformation-spreading/

[52] Reporters without Borders. (2020). *Beware of China's coronavirus disinformation, RSF says*, April 18 2020, https://rsf.org/en/news/beware-chinas-coronavirus-disinformation-rsf-says

[53] Botezatu, B. (2017, February 2). *Beware of Fake News - From a Cybersecurity Standpoint*. Retrieved 8 May 2020, https://businessinsights.bitdefender.com/fake-news-cybersecurity

[54] Symantec Security Response Team. (2018, September 18). *Subverting Democracy: How Cyber Attackers Try to Hack the Vote*. Retrieved March 7, 2020, from https://symantec-blogs.broadcom.com/blogs/election-security/election-hacking-faq

[55] Reed, J. (2017, May 26). *Q & A With Citizen Lab on "Tainted Leaks" and Russia's Disinformation Campaign*. Retrieved March 3, 2020, from https://www.justsecurity.org/41404/citizen-lab-tainted-leaks-disinformation/

[56] Carnegie Europe. (2020). *Russia's Long-Term Campaign of Disinformation in Europe*, 19.03.2020, https://carnegieeurope.eu/strategiceurope/81322

[57] National Review. (2019). *Countering Foreign Disinformation on Social Media*, 16.05.2019, https://www.nationalreview.com/2019/05/countering-foreign-disinformation-social-media- russian-campaign/

[58] Egloff, F. J. (2019). *Contested public attributions of cyber incidents and the role of academia*. Contemporary Security Policy, *41*(1), 55–81. https://doi.org/10.1080/13523260.2019.1677324

[59] Zetter, K. (2016, February 24). *The Sony Hackers Were Causing Mayhem Years Before They Hit the Company*. Wired, Retrieved February 27, 2020, from https://www.wired.com/2016/02/sony-hackers-causing-mayhem-years-hit-company/

[60] Zetter, K. (2016, February 24). *The Sony Hackers Were Causing Mayhem Years Before They Hit the Company*. Wired, Retrieved February 27, 2020, from https://www.wired.com/2016/02/sony-hackers-causing-mayhem-years-hit-company/

[61] Fraga-Lamas, P., & Fernandez-Carames, T. M. (2019). *Fake News, Disinformation, and Deepfakes: Leveraging Distributed Ledger Technologies and Blockchain to Combat Digital Deception and Counterfeit Reality.* Manuscript submitted for publication.

[62] Stupp, C. (2019, August 30). *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case.* The Wall Street Journal, Retrieved March 7, 2020, from https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

[63] Seabrook, J. (2019, October 31). *Can a Machine Learn to Write for The New Yorker?* The New Yorker, Retrieved March 26, 2020, from https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker

[64] Radford, A. (2019, December 13). *Better Language Models and Their Implications.* Retrieved March 26, 2020, from https://openai.com/blog/better-language-models/

[65] Fong, J. (2020, March 4). *Text generation algorithms could fill the internet with fake writing*. Vox, Retrieved February 27, 2020, from https://www.vox.com/recode/2020/3/4/21163743/ai-language-generation-fake-text-gpt2

[66] Dias, N. (2018). *The Big Question: How Will "Deepfakes" and Emerging Technology Transform Disinformation?,* Retrieved from The National Endowment for Democracy website: https://www.ned.org/wp-content/uploads/2018/10/How-Will-Deepfakes-and-Emerging-Technology-Transform-Disinformation.pdf

[67] Willemo, J. (2019). *Trends and Developments in the Malicious Use of Social Media*. Retrieved from NATO Strategic Communications Centre of Excellence website: https://www.stratcomcoe.org/trends-and-developments-malicious-use-social-media

[68] Patriarca, S., 2020 (2019, October 12). *Disinformation, Cybersecurity, And Online Influence*. Retrieved May 8, 2020, https://iiciis.org/international/2019/10/12/disinformation-cybersecurity-and-online-influence/

[69] New Knowledge. (2018, September 20). *The State of Disinformation 2018*. Retrieved March 6, 2020, from https://cdn2.hubspot.net/hubfs/4326998/NK_StateofDisinformation.pdf

[70] European Commission. (2018). *Action Plan against Disinformation*. JOIN(2018) 36 final, 5 December, 2018, https://ec.europa.eu/commission/sites/beta-political/files/eu-communication-disinformation-euco-05122018_en.pdf

[71] European Commission. (2019). *Code of Practice on Disinformation one year on: online platforms submit self-assessment reports*. 29 October 2019, https://ec.europa.eu/commission/presscorner/detail/en/statement_19_6166

[72] European Parliament. (2019). *Regulation (EU) 2019/881 of the European Parliament and of the Council on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act)*, 17 April 2019, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0881&from=EN

[73] European Council. (2017). *Draft Council Conclusions on a Framework for a Joint EU Diplomatic Response to Malicious Cyber Activities ("Cyber Diplomacy Toolbox")*. 7 June 2017, http://data.consilium.europa.eu/doc/document/ST-9916-2017-INIT/en/pdf

[74] European Council. (2019). *Cyber-attacks: Council is now able to impose sanctions*, 17 May 2019, https://www.consilium.europa.eu/en/press/press-releases/2019/05/17/cyber-attacks-council-is-now-able-to-impose-sanctions/

[75] World Health Organization. (2020). *Novel Coronavirus(2019-nCoV) Situation Report – 13*, 02.02.2020, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[76] European Council. (2020). *Joint statement of the Members of the European Council* 26 March 2020, Brussels, https://www.consilium.europa.eu/media/43076/26-vc-euco-statement-en.pdf

[77] POLITICO. (2020). *European Commission accuses China of peddling disinformation*, 10 June 2020, https://www.politico.eu/article/european-commission-disinformation-china-coronavirus/