Social Media Manipulation of Elections

Group 5

Even Hyldmo	evenhy@stud.ntnu.no
Ivar Olav Moen	ivaromoe@stud.ntnu.no
Ketil Østnor	ketilost@stud.ntnu.no
Kristian Havstein	kristhav@stud.ntnu.no
Kristian Tørseth	krto@stud.ntnu.no
Linett Simonsen	linetts@stud.ntnu.no
Michael Tareke Gebremariam	michaetg@stud.ntnu.no
Ole André Hauge	oleahau@stud.ntnu.no
Sigrid Anne Hafsahl Karset	sakarset@stud.ntnu.no

Date:	23.11.2020
Version:	2.0
Supervisor:	Bernhard M. Hämmerli
Page count:	133
Appendices:	0
Keywords:	Social media, manipulation, election, coordinated inauthentic behavior, democracy, disclosure, mitigation

NTNU | Norwegian University of Science and Technology

Department of Information Security and Communication Technology
Gjøvik, Norway

Executive Summary

By Ole André Hauge

This book portrays the situations, concepts, technologies, and methods used in social media manipulation of elections, and gives the reader the situational awareness needed to avoid, accept, mitigate, transfer, and continue research on the subject effectively. There are nine chapters, including a general introduction setting the context for the other papers:

- *"Manipulation and How Elections Were Influenced Before Social Media"*, by Ivar Olav Moen
- *"Election Manipulation After the Emergence of Social Media"*, by Linett Simonsen
- *"Coordinated Inauthentic Behavior*", by Sigrid Anne Hafsahl Karset
- *"Undermining Democracy - The Effect of Manipulation Through CIB in Social Media"*, by Even Hyldmo
- *"Social Media Awareness and Operational Education"*, by Kristian Tørseth
- *"Disclosure and Mitigation of SMME – Technical Solutions, Limitations, and Challenges"*, by Kristian Havstein
- *"The Russian Internet Research Agency's Interference in the 2016 US Election"*, by Ketil Østnor
- *"The Severity of Coordinated Inauthentic Behavior"*, by Micheal T. Gebremariam

The research was conducted using systematic literature reviews in the fields of information technology and security, computer-, political-, and social science, and information from the intelligence domain. Some of the sub-topics utilize empirical research, meaning that case studies and experience within the topic are used in addition to theory.

The second chapter explores the concept of manipulation and discusses relevant tools and techniques used for manipulating political processes. Ivar found that there are several forms of manipulation, such as propaganda and social engineering, which he exemplifies by conducting case studies on election manipulation. He discovered that the hostile actors' requirement to stay covert affected the manipulation methods they applied.

Democratic processes during 2016 were illegitimately influenced through the use of social media. Linett provides an analysis of how design-weaknesses in social media platforms are exploited to influence public opinion. Her findings show examples of tactics and techniques used by state actors to influence public opinion through social media. She further discovered that the data-driven advertising model used by social media platforms is one reason why political influence campaigns may succeed and that it will remain hard to regulate election manipulation through social media if its effects are not measured.

Coordinated inauthentic behavior (CIB) has seen a great upsurge in recent years. Sigrid describes CIB, which has been the basis for the removal of several pages, groups, and accounts. However, various platforms may have different interpretations of what constitutes CIB, which affects the mitigation efforts and techniques from those platforms. She further discovered that CIB and fake news are two separate topics.

Chapter five is solely focused on manipulation and CIB's effect on democracy. Even found that the emergence of social media has meant that governments, can conceal their manipulating behavior with the use of CIB. He further discusses how this is said to challenge democracy, as people might not be able to make informed decisions. The discussion of who has the responsibility to fight this type of manipulation is important, yet no clear answer has surfaced.

Kristian T. portrays how social media is increasing the political polarization in the US, and how social media platforms neglect to protect their users from false or incorrect information. He found that social media amplifies polarization and that social media is the perfect tool to undermine the integrity of a democratic election. He goes on to discover non-technical methods to mitigate the problem like awareness, critical thinking, and validation of content.

Chapter seven gives an overview of current problems and solutions for detection and mitigation on social media platforms, particularly by machine learning. Kristian H. found that the most used techniques and tools for social media manipulation are a mixture of human-, machine-, and hybrid-elements. He further discovered that most of the mitigation efforts focus on artificial intelligence and machine learning and that there are limitations to the use of this technology. He also explores other mitigation avenues outside of the artificial intelligence domain.

As discussed by Sigrid, foreign coordinated inauthentic behavior (FCIB) is probably the most used type of manipulation online. This statement is further strengthened by the finding of Even concerning the Russian government's use of manipulation.

In his case study on the Russian Internet Research Agency's (IRA), Ketil found that the IRA used FCIB in their manipulation efforts. He discovered that the IRA created several accounts and pages across platforms and used memetics to reach younger audiences. They used CIB to amplify conspiratorial narratives. Ketil estimates that the content created by the IRA on Facebook and Instagram combined may have reached as many as 264 million people.

The severity of CIB is further discussed in chapter nine by Michael where he found that CIB has had a great impact on several elections and crises on a global scale. He goes on to discuss how the social media companies are failing to handle the problem.

In conclusion, the act of manipulation is something everyone is familiar with. However, since manipulation could be used for hostile attempts of influencing political processes, it is something every member of a democratic society should be aware of. It may constitute a potent enabler to undermine states transitioning to democracy, or enables a state to undermine its population or a foreign state's democratic election. As the number of social media users escalates, the potential impact of CIB increases greatly. The use of automation for CIB and manipulation by government cyber troops and private firms is increasing in scope and technical abilities. If the social media platforms and the users of the internet do not employ and constantly improve their defenses and remediation against this type of behavior, the consequences will be severe. Nations need to be aware of the possibility of foreign nations trying to manipulate democratic elections in the future. Foreign nations like Russia might have an interest in who is elected as head of state. Their covert operations show how far they are willing to go to succeed in manipulating a democratic election. As Even puts it: "It remains a paradox that the platform to share information, strengthening one of the principles of democracy[,] also may account for undermining the principle itself".

Due to the likelihood and severeness of CIB and the use of social media manipulation, several recommendations are proposed by the authors to take precautionary actions and to provide directions for future research. The recommendations are summarized as follows:

- An increased awareness of the potential problem among the population.
- An increased knowledge and usage of critical source and information evaluation.
- A stronger cooperation between the public and private sector to respond to antidemocratic influential campaigns in social media.
- Public contribution to the removal of inauthentic users, by reporting suspected users to the platforms.
- Investment in human moderation by social media companies to backstop current trends.
- Government regulation of social media companies and mandates that ensure transparency.

The future research suggested by the authors are summarized as follows:

- How can people become aware of social manipulation in their everyday life?
- How were elections manipulated in the 19th century and before?
- The reason for the effectiveness of manipulation as a tool used by nation-states to influence the public.
- The implications of foreign electoral intervention attempts on citizen behavior.
- Accurate automated detection and mitigation of CIB and synthetic media.
- Collect further evidence to investigate the topic further. E.g., using Analysis of Competing Hypothesis.

Two major limitations are common across the papers. The first is the lack of previous research studies on the topic. Since the topic is new the low number of studies have at times made it difficult to find validated information to support ideas and statements. The second is cultural bias. The effect of the *WEIRD-phenomenon* (western, educated, industrialized, rich, and democratic) may also have affected the results in the papers. The group has been aware of this, however the lack of knowledge in foreign non-western languages like Chinese and Russian, have made it hard to use other sources. The first limitation will fix itself given time, as more research is conducted on the topic. Future studies should focus on the use of non-western research and preferably collaborative research to improve diversity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

By Ole André Hauge

> "It's the gradual, slight, imperceptible change in your own behavior and perception that is the product" (Rhodes, 2020).

Since the emergence of social media, society has welcomed the technology with open arms. During the last decade, there has been a shift in the purpose of social media platforms. It has evolved from being a product for the user to have the users be the product. Social media has become a purely capitalistic platform. User data is now mined and regularly fed to algorithms, which use this to continually self-improve in predicting user behavior. The unparalleled accuracy of the predictions and sometimes the user data itself is sold to other parties, organizations, and companies. Today the public disputes whether user data has become the new oil (B, 2018), (J, 2014). However, adequate research on this topic has not been conducted.

Platforms evolved to be more efficient and accurate by implementing psychological mechanisms, designed to create a user addiction which cleverly facilitates the collection of ever more user data (Rhodes, 2020). An example can be found on almost every information feed on social media platforms, where one can pull down to refresh the feed. This design behavior is based on "the law of effect" (Thorndike, 1927), a law in learning theory, contending that behavior which is rewarded is repeated. It is the foundation of the "variable ratio schedule" used in slot machines and is resistant to extinction: The users know that a reward is due by updating their feed, but not how many times they have to repeat the action to get the reward. Tricking them to keep updating, hoping for a new reward. This is popularly referred to as "the gambler's fallacy" (Ayton, 2004).

The public became aware of the social medias power to manipulate elections after the 2014 and 2016 elections. This awareness was further strengthened by the Cambridge Analytica (CA) scandal, where it became evident that president Donald Trump got help from CA with big data analysis during his election campaign. Since then the focus on the phenomenon has increased, as well as the use of it.

Thus, with (1) a growing social media addiction, (2) a growing interest in the platform as a tool for manipulating the public, and (3) both utilizing an exponentially growing technology, the society is facing a new problem; albeit an evolution of propaganda and manipulation. The effectiveness of social media to reach people and share information fast makes it a perfect platform for the manipulation of the public, and might even be seen as a threat to western democracies.. There is no clear method or technology in place to fight or mitigate this phenomenon effectively. The area evolves fast and is relatively new, thus the researchers are struggling to keep up with the development. Companies, such as Facebook, Instagram, and Twitter are starting to implement methods, algorithms, and policies to mitigate the problem (Gleicher and Stamos, 2018), (Cohen, 2020), (Statt, 2020).

Research shows that the organized use of social media for manipulation increased by a massively 150% from 2017 to 2019 (Bradshaw and Howard, 2019). Research has also proved that social media manipulation (SMM) techniques have been used in as many as 48 countries the last years (Bradshaw and Howard, 2018). Further research has been conducted on the different techniques and effectiveness of SMME. However, there is a lack of broader research that can help societies, researchers, and governments achieve an understanding in the many facets that this problem encompasses

This book aims to portray the situation, concepts, technologies, and methods used in SMME today, and give the reader a greater situational awareness needed to mitigate and continue research on the subject effectively. There are eight chapters: (2) Manipulation and How Elections Were Affected Before Social Media, (3) Election Manipulation After the Emergence of Social Media, (4) Coordinated Inauthentic Behavior (CIB), (5) Undermining Democracy - The effect of manipulation through CIB in social media, (6) Social Media Awareness and Operational Education, (7) Disclosure and Mitigation of SMME – Technical Solutions, Limitations and Challenges, (8) Foreign Coordinated Inauthentic Behaviour - A Case Study on the Russian Internet Research Agency's Interference in the 2016 US Election, and (9) The Severity of Coordinated Inauthentic Behavior. Each chapter targets a sub-topic and summarizes some future areas of research that can aid in developing solutions to this growing societal problem.

Bibliography

Ayton, P., F. I. (2004), 'The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?'.
URL: *https://doi.org/10.3758/BF03206327*

B, M. (2018), 'Re-floating the titanic: Dealing with social engineering attacks'.
URL: *https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#2775b1a3aa96*

Bradshaw, S. and Howard, P. N. (2018), 'Challenging truth and trust: A global inventory of

organized social media manipulation'.
URL: *http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf*

Bradshaw, S. and Howard, P. N. (2019), 'The global disinformation order 2019 global inventory of organised social media manipulation'.
URL: *https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf*

Cohen, D. (2020), 'Twitter details several networks of accounts removed for coordinated inauthentic behavior', *Adweek* . Accessed 27.09.2020.
URL: *https://www.adweek.com/digital/twitter-details-several-networks-of-accounts-removed-for-coordinated-inauthentic-behavior/*

Gleicher, N. and Stamos, A. (2018), 'Removing bad actors on facebook', *Facebook* . Accessed 06.09.2020.
URL: *https://about.fb.com/news/2018/07/removing-bad-actors-on-facebook/*

J, T. (2014), 'Re-floating the titanic: Dealing with social engineering attacks'.
URL: *https://www.wired.com/insights/2014/07/data-new-oil-digital-economy*

Rhodes, L., O. J. (2020), *The Social Dilemma*, Netflix.

Statt, N. (2020), 'Tiktok is banning deepfakes to better protect against misinformation', *The Verge* . Accessed 27.09.2020.
URL: *https://www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference*

Thorndike, E. L. (1927), 'The law of effect', *The American Journal of Psychology* **39**, 212–222.
URL: *http://www.jstor.org/stable/1415413*

Chapter 2

Manipulation and How Elections Were Influenced Before Social Media

By Ivar Olav Moen

Executive Summary

This paper presents the concept of manipulation and discusses relevant tools and techniques used for manipulating political processes. This is further exemplified by exploring real-life cases of election manipulation from the 20th century.

The paper draws attention to the fact that there are many different views on manipulation and how to define it even though it is a concept that most people is fundamentally familiar with to some extent. Some authors claim that manipulation needs to include deception when communicating with a victim, while others state that it only needs some kind of persuasion in order to change the views and opinions of the victim. Considering all the definitions explored in the paper, it is however quite certain that the act of manipulation is a delicate and sophisticated process that exploits fundamental human behaviour in order to change the outcome of an event.

Manipulation is a crucial part of the influencing operations done by states within their sphere of influence and has been in use for a long time, even before social media developed into what we have today. Several tools can be used, and one of the most used tools is propaganda. Although this concept also has many different definitions it can be summarized to include most kinds of communication aimed at changing the views that certain groups have on a specific matter. One state can for instance use propaganda to form how the population of a country should view specific political groups.

The process of manipulation includes social engineering in which specific actions are aimed at changing views, opinions and acts of the victim. Some people are more prone to experi-

encing social engineering, but this can change according to changes in their everyday lives. In order to be effective, the process of social engineering needs to be tailor-made to target a specific group or person. When done correctly, the effects of social engineering can be a crucial part of influencing operations.

During the Italian elections of 1948, the US government was involved on many different levels of the political processes. The US made sure that the Christian Democratic Party won the election over the Communist Party in order to reduce the influence of communist forces. This was done by using techniques such as propaganda aimed at Italian citizens, economic aid aimed at specific political groups, manipulation of important diplomats through communication and deployment of covert agents in the Italian society. This joint effort helped changing the views and actions of the Italian population and serves as an example of election manipulation.

US influencing operations in Chile during the 1960s and 1970s show similar techniques and results as in the Italian election and further exemplifies how manipulation can be used to affect democratic processes. By using previously mentioned tools and techniques, the US made sure that communist groups in Chile were minimized. The actions in Chile were however less covert and utilized economic aid in an aggressive manner in order to support the targeted political groups. It therefore shows a slightly different approach to manipulation.

The paper is limited by its scope and its intention of primarily being an introduction to the concepts of manipulation and its uses in influencing operations targeted at election processes. It does however commit to the task and serves naturally as a springboard to the rest of the chapters in this book where several related concepts are discussed more thoroughly.

2.1 Introduction

While manipulation of elections is a common phenomenon seen in elections all over the world and often mentioned by the modern media today, it is not a new invention and can be seen in the history of elections long before social media appeared.

Manipulation is used by most people in their everyday life and is not limited to politicians and their apparatus. However, not everyone is aware of their manipulative actions or how they are being affected by other people's manipulation, but it is nevertheless present. As an example, even small children will try to manipulate their parents into for instance getting more candy than usual because they have learned that by behaving in a certain way, they are awarded with what they want (Nordhelle, 2009). Although a banal example, it shows that manipulation is used by virtually everyone. Because it is used by so many people in various situations, one could likely make a separation between everyday manipulation and hostile manipulation, like the manipulation that could be present when someone wants to influence an election process.

Manipulation of elections can be seen as a broad concept, and within its landscape we also find similar terminology such as lobbying, vote gathering and election campaigns. While some parts of these terms might be perceived as something negative, others are vital for a democratic process. One could easily ask the question of where one ends and the other starts.

Since everyone is using manipulation, it is of interest to explore its definitions. What is considered to be manipulation, and what is not? In this paper the concept of manipulation will be explored and discussed, presenting various definitions and uses. In order to exemplify the manipulation techniques that are presented, real-life examples from the last century will be discussed to show some ways manipulation was used to influence elections before social media developed into what we have today.

2.2 Methodology

This paper has been undertaken with a qualitative approach to the subject matter. In order to create a basis for discussion of the research topic, a literature review was conducted. This included approaching relevant research papers, books, reports, and other textual publications for the sake of considering their relevance to the matter.

Relevant academic sources have been found by using acknowledged online academic search engines and literary databases. A selection of the encountered publications has been used as a basis for this paper. The selection was done in order to stay within the scope of this paper as being both an introductory chapter to the book it is a part of, but also functioning as an independent paper introducing several different concepts. Because the aim of this paper is to look at manipulation and how it was done before social media developed into what we know today, an effort to identify publications produced both before and after the onset of social media was undertaken. This was done in order to better discuss the topic in a context relevant to the time period in question.

The chosen publications have been presented and compared throughout the paper, creating a continuous discussion relating to the research topic.

2.3 Manipulation in general

Since manipulation is such as broad concept, it is firstly important to get a clearer understanding of what the concept includes and how it can be defined.

Manipulation is defined by Cambridge Dictionary as “controlling someone or something to your own advantage, often unfairly or dishonestly” (Cambridge Dictionary, n.d.*b*). Already by a semantical understanding of the term, one could see manipulation as something negative, but this is a blunt description, and the concept should be further nuanced. According to Todd, manipulation involves deception in which an individual tries to influence another person covertly. Within the act of manipulation are hidden factors in the communication that are used to influence a person’s actions, perspectives, and thoughts (Todd, 2013, p. 1).

This is different from coercion in which an individual is overtly influenced in order to change his or her actions, perspectives and thoughts. The act of coercion is often seen as something negative, but also much more explicit than manipulation needs to be and can be defined as "to achieve by force of threat" or "to restrain or dominate by force" (Merriam-Webster, n.d.). This creates a situation in which the victim of the coercive action cannot resist doing what they are being forced to. It is also obvious for the victim that the applicant of coercion wants to influence the other person, this is not necessarily the case with manipulation. In order to successfully manipulate someone, it has to be done with precision and intellect, not blunt force. For instance, Rudinow states that "[...]manipulation seems delicate, sophisticated, even artful in comparison with the hammer-and-tongs crudity of coercion and seems also to be similarly restricted in its application to beings with some minimum level of sophistication, complexity, and intelligence" (Rudinow, 1978, p. 339).

It already becomes clearer that manipulative actions need to be done in a certain, delicate manner in order to succeed. Perhaps manipulation could be compared to "tricking" someone into doing something, although that could also be a too simple comparison. However, Todd argues that by applying deception in communication with a victim, and thereby making the victim change his goals or create new goals, one is manipulating the victim (Todd, 2013, p. 2). This indicates that manipulation is a demanding task in which it is important not only to know who your counterparts are, but also know how to best influence what they think and how they act.

Based on these notions of manipulation, a picture of it being something negative is quickly created. However, we have related concepts that are often seen as much more positive. In a democratic society, politicians are expected to run election campaigns and commit to vote gathering in order to win their respective elections. By doing such actions, the politicians are in fact trying to influence the population to do certain actions although it is not necessarily seen as something negative.

According to the Cambridge Dictionary, an election process is defined as "the period of time immediately before an election when politicians try to persuade people to vote for the" (Cambridge Dictionary, n.d.*a*). Although the term "persuade" has been used, it is most likely not seen as a negative action such as coercion mentioned earlier. By holding rallies, press conferences, debates, create advertisements and participate in many different forms for communication on a vast number of platforms, the politicians try to persuade the public into voting for them. Some might claim that politicians in general are liars, but assuming this is not the case, an election campaign should be seen as a process in which politicians announce what kind of work they are willing to carry out should they be elected. This is of course the basis of an ideal election process without actions such as coercion or negative forms of manipulation. The core difference between election campaigns and election manipulation is that the first includes overt actions where it should be obvious what the politician wants to achieve, while manipulation includes covert actions where someone is tricked into acting in a certain way.

2.4 Manipulative Tools

2.4.1 Propaganda

Propaganda is a phenomenon many people have experienced, and it is often used in relation to debates, news broadcasts, articles and even during election processes. According to the Cambridge Dictionary propaganda is defined to be "information, ideas, opinions, or images, often only giving one part of an argument, that are broadcast, published, or in some other way spread with the intention of influencing people's opinions" (Cambridge Dictionary, n.d.*c*).

While that is a broad and fitting definition of what propaganda can be, the concept is also in need of being refined as it has many nuances, and several authors have different definitions. Kellen states that the purpose of propaganda is not only to change opinions, but it also aims "to intensify existing trends, to sharpen and focus them, and, above all, to lead men to action" (Ellul, 1965, p. VI). This indicates that propaganda can easily be used to strengthen views and perspectives already present within a society. A simple comparison could possibly be to how we use stereotypes to easily classify the behaviour of certain people in our everyday lives. If we for instance see people from a certain nation as hard-working, it would be easier for us to further extend this view every time we get confirmation of this trait. Thereby strengthening our perception of the people as being hard-working.

Ellul further highlights the fact that there are many different definitions, and that these often are comparable and move about the same landscape. One of these definitions include: "Propaganda is the expression of opinions or actions carried out deliberately by individuals or groups with a view to influencing the opinions or actions of other individuals or groups for predetermined ends and through psychological manipulations" (Ellul, 1965, p. XI-XII). Although many definitions have similarities, he also highlights the fact that several authors have disagreeing views. Depending on what sources are consulted, one could likely find different approaches to the definition of propaganda from various factions of the propaganda landscape.

Since this is not a study in how propaganda can be defined, the perspectives already brought out can be used as a basis. Based on these, an important part of propaganda is that there needs to be a certain degree of psychological manipulation in the spreading of opinions and information in order for it to be considered propaganda. An example could include the difference between stating that a person was robbed on the streets of London in the evening news. This is merely a presentation of facts. However, if the news story involves a particular biased description of the perpetrator and his action such as: A young asylum seeker from a certain nation, having an especially negative behaviour and appearance, violently attacked a poor, old man walking home from the local church after the evening service. In this banal example, the first part exemplifies a situation in which the news only serves to inform factually about a criminal offence that took place, while the purpose of the second is to form the opinion of the reader about the behaviour of a person with a certain ethnicity. Similar examples could be seen in news media all over the world today and has most definitely been present earlier in history as well.

However, propaganda could be so much more. In his discussions about propaganda, Ellul states that propaganda to some extent could include all of the following areas:

> Psychological action: The propagandist seeks to modify opinions by purely psychological means [...].
>
> Psychological warfare: Here the propagandist is dealing with a foreign adversary whose morale he seeks to destroy by psychological means so that the opponent begins to doubt the validity of his beliefs and actions.
>
> Re-education and brainwashing: Complex methods of transforming an adversary into an ally which can be used only on prisoners.
>
> Public and human relations: These must necessarily be included in propaganda [...]. They serve to make him conform, which is the aim of all propaganda (Ellul, 1965, p. XIII).

He deliberately explains propaganda as something extensive to further call attention to the fact that propaganda can be extremely different according to the situation and context in which it is applied. One state or group could use it for one purpose in a certain context, while another state uses it for something completely different. It does not necessarily need to be a state actor at all, propaganda could also be used by private organizations. The use of propaganda constitutes a highly sophisticated influencing technique where the propagandist "builds his techniques on the basis of his knowledge of man, his tendencies, his desires, his needs, his psychic mechanisms, his conditioning - and as much on social psychology as on depth psychology" (Ellul, 1965, p. 4). This means that the propaganda needs to be tailor-made for the target community in order for it to function properly, and that is possibly why it tends to have such a tremendous effect on a population's opinion and understanding of whatever is being influenced.

Detailed examples on the use of propaganda in manipulation of elections will be explored briefly further on in this paper, but also to a greater extent in other chapters of this book.

2.4.2 Social engineering

Although social engineering for many people is associated with modern phishing attempts and various forms of cyber-attacks, it is important to remember that it uses basic human behaviour in order to influence someone and is well-known in the social sciences as well. Social engineering could in general terms be seen as "psychological manipulation, skilled or otherwise, of an individual or set of individuals to produce a desired effect on their behaviour" (Harley, 1998, p. 9).

This definition implies that the psychological manipulation indeed does not need to be of negative or destructive nature. This also coincides with the previously explored definitions of manipulation. A non-hostile example of social engineering could for instance be when someone tries to manipulate their friends into choosing one activity over the other by lying. On the other hand, a situation including hostile social engineering could be when criminals approach you on the internet claiming to be a Nigerian prince needing help with getting rid of an absurd amount of gold, while they in reality are looking for ways to make you lose your money.

Although Harley (1998) in his paper mostly explores social engineering in the context of information security, many of the principles and examples brought up could be applied onto a more analogue context, for instance in situations where social engineering is used outside of social media.

In order to better understand social engineering, some of the techniques used should be mentioned. Harley mentions the term direct psychological manipulation, which could cover many different actions such as "seduction and bribery, intimidation, [...], extortion and blackmail" (Harley, 1998, p. 12). Although these tend to be quite hostile, they help show some manipulative actions that could be undertaken by someone willing to manipulate other people. Other examples could include the act of lying and biased presentation of information, two actions commonly claimed to be used by politicians and their apparatus. Whether it is used or not probably differs from country to country and what context the communication takes place in. The effect also varies according to the context in which it is applied.

People tend to behave differently and some people are more easily influenced than others. This does imply that some people are better victims than others, and depending on the context, situation and the person, most people will be influenced or manipulated by others at some point in their life. Harley highlights several factors that can make people vulnerable to social engineering and mentions phenomena such as gullibility, curiosity, courtesy, greed, diffidence, thoughtlessness and apathy (Harley, 1998, p. 13-14). Although these characteristics of people could be seen as something negative, most people have some degree of these traits at some points in life, making ordinary people vulnerable to social engineering. The characteristics do not need to be present at all times, but could appear differently according to different life situations. A young student has different needs than a 40-year-old or a pensioner. They will all be susceptible to social engineering differently throughout their life.

To give another type of example, politicians might target specific communities where they know that their reception can be influenced and then carry out social engineering actions. Perhaps one politician easily establishes good relations with the poor parts of society, another one with a certain ethnicity and a third with the rich and well-educated communities. By using the previously mentioned characteristics, the communication from the politician can be tailored to the needs of the community and thereby creating a positive relation that ends in a higher number of votes during an election.

Since there is a great difference in how social engineering is used, and not all of them are prominently negative, it is of interest to explore negative forms of social engineering in greater depth. According to the RAND Corporation the term "hostile social manipulation" is used to define situations in which states "seek to gain competitive advantage by manipulating political, social, and economic conditions in target countries by various informational means"

(RAND Corporation, n.d., p. X). Although the report focuses on modern social engineering situations including modern media outlets such as social media, it highlights issues regarding the use of fabricated information and created narratives to be used in election campaigns. These campaigns are not only seen through traditional news providers, but also for instance through advertisement, books and the work done by PR businesses.

Although not necessarily focused on specific election processes, the Soviet hostile social manipulation done during the union's existence shows an interesting example of how nation-states can and have influenced both its own population and the populations of other countries in order to achieve what they want. The Soviet Union was dependent on increasing their area of political influence around the globe, and attempted to do so by social manipulation amongst many other tactics. Even though the social manipulation techniques within the Soviet society began long before the union arose, its uses become clear when examples from the 1920s and further on are explored. For instance, during Vladimir Lenin's reign in the 1920s, he "designed and executed propaganda campaigns to discredit domestic and foreign adversaries and to foster support for the Communist ideology", and outside their borders, "Soviet leadership sponsored the dissemination of disinformation in Western Europe [...] to malign émigré groups" (RAND Corporation, n.d., p. 33-34). Already in the youth of the union, techniques and tactics seen today were implemented in both home affairs and foreign policy. Not only did this mean that they influenced their own population, but also the population of the enemy or countries in which they had political interests in.

As time progressed, the Soviet approach to social manipulation evolved and many different state departments were created with responsibilities according to the domain and methods to be used, such as the creation of the infamous KGB. During the 1950s the state used "devices of covert action as forgeries, planted press articles, planted rumors, and controlled information media" in order to shape the opinion of the general population (RAND Corporation, n.d., p. 36). Although the apparent use of propaganda as a tool, one can easily assume that social engineering was used during these influencing operations, and that this most likely shaped the development of Soviet politics.

This helps to show examples of how a state can operate without going in detail on the exact methods used in their operations. However, in order to better understand exactly how these manipulative actions can be used, the next section will exemplify this by looking at a couple of historical events in which these techniques were successfully used.

2.5 Case studies of foreign intervention in elections before social media

2.5.1 Case 1 – US involvement in the Italian elections of 1948

After World War II Italy suffered difficulties such as many other countries in warn-thorn Europe. One of the greater changes was that the Italian state became a republic in 1946, which meant that the elections of 1948 were the first general elections that were held in the new republic. In the period before the elections, many parties had their saying in who to elect into Italy's first parliament. This was clearly an event that could have great consequences both within and outside Italy itself, and several states such as the US, Great Britain and the Soviet Union had interests in its outcome. Although several states intervened in the elections, this section focuses on the American influencing operations in order to give examples of how manipulation of elections were done in the past. Miller goes into depth on this matter in his article about the US involvement in the Italian elections of 1948 (Miller, 1983).

The US perceived Italy as a strategic nation within its sphere of influence and saw it as a loss if the Communist Party in Italy would have been victorious after the election. At the same time, the Soviet Union wanted the Communist party to win as it would cause them to involve Italy in their sphere of influence. This was crucial for both parties due to the Cold War that had just begun. While the US was afraid of a Communist victory in the Italian election, the Soviet Union wanted it to happen. The US on the other hand wanted the Christian Democratic party to be victorious.

In early 1947, the US began supporting the Christian Democratic party, and their leader Alcide De Gasperi, with funds in order to strengthen the party, a party the US saw as crucial to their interests in the region (Miller, 1983, p. 36). As a response to the US support, De Gasperi of the Christian Democratic party, forced the leftists out of the government and initiated an economic reform. The leftists began violent campaigns and the country started heading into a civil war. As a response to the threat of an armed revolution, the US continued their support for De Gasperi and his allies, making it clear that De Gasperi was not alone.

In the beginning of 1948, it seemed like the Communists would in fact win after public opinion polls were checked. The US intensified their operations to defeat them and strengthen the Christian Democratic Party, an operation which in the end succeeded. On the election day, De Gasperi and his Christian Democratic party clearly won over their opponents in the Communist party.

This was a superficial walk-through of a complex political matter which of course included many more events that could be brought up in another context. However, in general terms this is how the history went, but what measures did the US in fact initiate in order to achieve their wanted outcome?

For instance, economic aid was extremely important for the US success in their election manipulation process:

> Economic aid was critical to the survival of the De Gasperi government. Without this additional aid, the Italian economy would collapse before Marshall Plan funds arrived. Elections would have to be held soon, and if the country's economic position deteriorated further, the Left would sweep De Gasperi from power and Italy out of the U.S. sphere of influence (Miller, 1983, p. 42).

Although using even more US money on the European post-war soil met some political challenges in the US, this was decided in the end. By increasing the power of the political party of their choice, the US managed to indirectly let the population's interest in the Christian

Democratic party increase. Even though the direct communication between the people and the party was done by the Italians themselves, the economic means to involve in such communication were given by the US government. Therefore, economic support can be seen as one way to manipulate elections.

In addition to economic aid, the US were involved in Italy with military troops after the end of World War II. These were however planned to be withdrawn in the end of 1947. As the deadline for their withdrawal came closer, the leftists increased "[...]their political muscle-flexing with strikes, mass rallies, assaults on police stations, and occupation of factories." (Miller, 1983, p. 43). The US responded by presenting their military power and threatened of US military involvement in Italy should the Communist Party arrange a coup. This resulted in the leftist receding with their violent campaigns. Although the chances of the US wanting to be involved in a violent and militarized conflict in the Italy were low, the act of intimidation manipulated parts of the Italian society which then resulted in less participation and involvement for the Communist Party because of fear of the US. This could of course be seen as coercion, but helps to show that the scopes or boundaries of the concepts are hard to identify.

The US realized that they did not want to back up a threat of military force, and then began planning other ways of influencing the Italian population. For instance using covert operations with local agents. At one point an army chief wanted it arranged so "that a list of potential Italian agents be prepared and passed on to the Central Intelligence Agency (CIA) for possible use in covert operations in Italy" (Miller, 1983, p. 43). The US also used diplomatic staff in order to influence important members of The Vatican, which were an important part of the Italian society:

> The Vatican shared American apprehensions about a Communist coup, and with the encouragement of U.S. diplomats, the Church edged toward full participation in the anti-Communist coalition. In spite of Christian Democratic ties to the Church, Pope Pius XI1 and many of his advisers were ambivalent about Italy's nascent democracy. They recalled past clashes with democratic movements in Italy and shared U.S. doubts about the capability of DC leadership. Nevertheless, by the fall of 1947 Catholic leaders concluded that the Communist threat was so great that the Church must cast its lot with De Gasperi, the DC, and democracy. [. . .] The Vatican also informed U.S. officials that it would welcome American intervention to defeat a Communist attempt to seize power (Miller, 1983, p. 44).

As shown through the covert actions of the US, we see another example of manipulative actions. By communicating with key individuals of a society, it is possible to influence what they communicate out to the rest of the society and thereby their followers. When the masses change their opinions, the manipulation manifests itself.

American propaganda was also heavily used in Italy, especially in the times when the Communist Party gained increasing support from the Italian society. For instance in 1948, "the U.S. propaganda underlined the importance of Marshall Plan aid to Italian economic recovery and the fact that no Communist nation was participating in the European Recovery Program" in addition to exaggerating the amount of money Italy was to receive and that this would be decreased in case the Community Party won the general elections (Miller, 1983, p. 49). The US government also "flooded Italy with newsreels demonstrating the benefits of U.S. aid. Leading American film distributors pooled their resources and sent documentaries and government films free of charge" (Miller, 1983, p. 49). The general population was directly influenced by the American views camouflaged as legitimate and non-suspicious media outlets, not unlike what we see today in the use of social media in election campaigns.

Without digging deeper into the history of the Italian elections of 1948, we can conclude that the US used recognizable techniques for manipulating the Italian election. These includes actions such as the use of propaganda aimed at Italian citizens, economic aid to foster support from both government and the population itself, diplomatic relations to affect people in important positions as well as the deployment of covert agents. This helps indicate that the old ways of manipulating elections are not necessarily too different from what we see today, they were just carried out using the technology, politics, world economy and the world view existing in the late 1940s.

2.5.2 Case 2 – US involvement in the Chilean politics throughout the 1960s and 1970s

Throughout the 20th century, Chile experienced a lot of political difficulties. The wealthy parts of the population had for a long time been given preferential treatment and polarization between the different parts of the society developed. Several politicians attempted to fix this, but with varying degree of success causing tension to rise steadily. This culminated in a military coup in 1973 staged by the armed forces, which also led to the death of the former president Salvador Allende and a shift in political power from the left-wing to the right-wing (Britannica, n.d.). This in turn led to a period of military dictatorship lasting until the 1990s which contributed to social unrest in the Chilean society. Although many of the reasons as to why this coup happened can be explained by internal factors related to the Chilean society and politics, it is fairly certain that US involvement in Chile contributed to the outcome and the coup as well.

The US government has admitted to being involved covertly in Chile between 1963 and 1973 (U.S. Senate Select Committee, 1975, p. 1). This involvement included many different efforts, not only did they attempt to influence the outcome of the presidential elections of 1964, but they continued to influence the society until the political environment eventually led to the military coup in 1973. Most importantly, the US has admitted that "[t]he goal of covert action is political impact" (U.S. Senate Select Committee, 1975, p. 1), meaning that the actions done in Chile were clearly parts of an influencing operation aimed at manipulating the Chilean politics. Some of the actions undertaken "[. . .]includes covert action, clandestine intelligence collection, liaison with local police and intelligence services, and counterintelligence" (U.S. Senate Select Committee, 1975, p. 1). However, one of the most important measures undertaken by the US was economic support. As previously explained, an effective influencing method is the use of economic support targeting various groups and organizations in the targeted society, which was also used by the US in Chile:

> What did covert CIA money buy in Chile? It financed activities covering a broad spectrum, from simple propaganda manipulation of the press to large-scale support for Chilean political parties, from public opinion polls to direct attempts to foment a military coup. The scope of "normal" activities of the CIA Station in Santiago included placement of Station-dictated material in the Chilean media through propaganda assets, direct support of publications, and efforts to oppose communist and left-wing influence in student, peasant and labor organizations (U.S. Senate Select Committee, 1975, p. 1).

In the section on propaganda in this paper, definitions of propaganda were explored, and one such definition stated that the aim of propaganda is "[...]influencing the opinions or actions of other individuals or groups for predetermined ends and through psychological manipulations" (Ellul, 1965, p. XI-XII). Relating to the situation in Chile, we see an example of the US attempting to influence the opinions of the Chilean population during the presidential election of 1964 in which they staged an immense anti-communist propaganda campaign: "Extensive use was made of the press, radio, films, pamphlets, posters, leaflets, direct mailings, paper streamers, and wall painting. It was a "scare campaign", which relied heavily on images of Soviet tanks and Cuban firing squads and was directed especially to women[...]" (U.S. Senate Select Committee, 1975, p. 15). This is just a brief description of some of the tools used by the US, but shows that the list is extensive. In this case, by targeting a specific social group of the Chilean society, the messages conveyed could be customized to give the wanted output. Much in the same ways as described in this book's chapters on social media manipulation of elections.

But why did the US go through such extensive measures to influence the politics of Chile? For instance, when attempting to prevent Allende from becoming the elected president during the elections of 1970, the US did not succeed. It is noted that by letting Allende come to power it "[...]would threaten hemispheric cohesion and would represent a psychological setback to the U.S. as well as a definite advance for the Marxist idea" (U.S. Senate Select Committee, 1975, p. 48). Although Allende was elected during that specific election, and other US efforts failed during their operations in the 1960s and 1970s, this helps to show that the motivation and goal of US involvement nevertheless was to minimize the extent of influence the Communists and Marxists could have on the Chilean society.

The fact that the US extensively and deliberately targeted communist and left-wing forces in the Chilean society most likely strengthened the polarization in the society that eventually ended in the coup. Although they attempted at doing covert influencing operations, the U.S. Senate Select Committee report of 1975 highlights the fact that "there was simply too much unexplained money, too many leaflets, too many broadcasts" for the US involvement in Chile to remain covert (U.S. Senate Select Committee, 1975, p. 54). Because these influencing operations and their effects were so obvious in Chile, it was highly visible not only for the population of Chile, but also other countries in Latin America. Slowly an expectation of US involvement was created, making attempts at covert action challenging. Because of this, both the US government and the groups that received US support in Chile suffered negatively in view of the public because they were being known to be taking part in obstructing the democratic processes. This shows how challenging influencing operations can be. Although states all over the world attempt influencing operations in their spheres of influence, the public and especially those being influenced will likely be critical of such operations once they are revealed to the public.

It is evident that the involvement of the US government in Chile influence the local politics in a way that contributed to its development. By using social engineering, economic support, propaganda and other manipulative tools, it was possible to manipulate their political processes. This brief presentation of the involvement in Chilean politics is comparable to the US involvement in the Italian elections discussed in the previous section, and helps to show another example of how election processes can be manipulated and what this can result in.

2.6 Conclusion

The act of manipulation is something everyone is familiar with. Fortunately, not all have experienced its negative effects. However, since manipulation could be used for hostile social engineering or propaganda in attempts of influencing political processes it is something every member of a democratic society should be aware of.

Throughout this paper several examples of influencing operations aimed at manipulating election processes have been shown. It is evident that manipulation was highly possible even before the rise of social media. Although this paper mainly exemplified election manipulation as done by the US with some examples mentioning Soviet techniques, it could easily be exemplified in the perspectives of other countries such as Russia, China or many others. However, one will likely find similar tools, techniques and results as the ones presented here.

This paper is limited by its scope and the fact that it functions as a brief introduction to this vast topic. Further exploration of this topic could include looking at other case studies in which attempted influence did not succeed and situations in which other nations than the US attempted influencing operations. This can be important because it further shows that manipulative actions are done by everyone, even other states. Why is manipulation of the general population such an effective technique used by nation-states to influence politics? How can people become aware of social manipulation in their everyday life? How were elections manipulated in the 19th century and before? These are questions that can be addressed in a different context and have likely been explored by researchers already. The other chapters of this book will amongst other things look at topics related to what has already been mentioned in this paper. They will also explore manipulation using modern day technology, showing that the techniques used today have roots in the methods used long before today's advanced technology was developed.

Bibliography

Britannica (n.d.), 'Chile - the military dictatorship, from 1973'. Accessed 01.11.2020.
URL: *https://www.britannica.com/place/Chile/The-military-dictatorship-from-1973*

Cambridge Dictionary (n.d.*a*), 'election campaign'. Accessed 03.10.2020.
URL: *https://dictionary.cambridge.org/dictionary/english/election-campaign*

Cambridge Dictionary (n.d.*b*), 'manipulation'. Accessed 04.10.2020.
URL: *https://dictionary.cambridge.org/dictionary/english/manipulation*

Cambridge Dictionary (n.d.*c*), 'propaganda'. Accessed 07.10.2020.
URL: *https://dictionary.cambridge.org/dictionary/english/propaganda*

Ellul, J. (1965), 'Propaganda: The formation of men's attitudes'. Toronto: Vintage Books Edition, 1973.

Harley, D. (1998), 'Re-floating the titanic: Dealing with social engineering attacks'. Accessed 30.08.2020.
URL: *https://smallbluegreenblog.files.wordpress.com/2010/04/eicar98.pdf*

Merriam-Webster (n.d.), 'Coerce'. Accessed 13.09.2020.
URL: *https://www.merriam-webster.com/dictionary/coerce*

Miller, J. (1983), 'Taking off the gloves: The united states and the italian elections of 1948', *Diplomatic History, 7(1), p. 35-55* .
URL: *http://www.jstor.org/stable/24911419*

Nordhelle, G. (2009), 'Manipulasjon: Forståelse og håndtering (manipulation: Understanding and handling)'. Oslo: Gyldendal akademisk.

RAND Corporation (n.d.), 'Hostile social manipulation: Present realities and emerging trends'. Accessed 30.08.2020.
URL: *https://www.rand.org/content/dam/rand/pubs/research_reports/RR2700/RR2713/RAND_RR2713.pdf*

Rudinow, J. (1978), 'Manipulation', *Ethics, 88(4), 338-347* . Accessed 13.09.2020.
URL: *http://www.jstor.org/stable/2380239*

Todd, P. (2013), 'Manipulation', *International Encyclopedia of Ethics, pages 3139-3145* .
URL: *DOI: 10.1002/ 9781444367072.wbiee585*

U.S. Senate Select Committee (1975), 'Covert action in chile 1964-73'. Accessed 30.08.2020.
URL: *https://archive.org/details/Covert-Action-In-Chile-1963-1973/page/n3/mode/2up*

Chapter 3

Election Manipulation After the Emergence of Social Media

By Linett Simonsen

Executive summary

Democratic processes such as the 2016 United States presidential election and the 2016 United Kingdom European Union membership referendum were exposed to extensive influence operations (Datatilsynet, 2019, Norwegian National Security Authority, 2020*b*). In both of these operations, design "weaknesses" in social media platforms were being exploited. This chapter provides an analysis of how democratic elections may be illegitimate influenced through the use of social media. The main focus is on partisan electoral interventions and how foreign state actors are using social media platforms in general, and Facebook in particular, in an attempt to influence public opinion. Through a literature review, the following areas are examined:

- Traditionally we have PR, influence, tricks, cunnings, and deceptions. What is the difference between these traditional forms and the new ones in social media?
- What tactics and techniques may state actors use in order to influence public opinion through social media?
- How may public authorities, private companies and individuals tackle disinformation campaigns in social media?

Currently, the dominant business model of the internet is based on surveillance. Shoshana Zuboff, who coined the term 'surveillance capitalism' in 2014, warns that the shift from a mass to an individual-oriented structure of consumption may undermine personal autonomy and erode democracy (Laidler, 2019). Corporations are increasingly mining users' information to predict and shape their behavior. Despite possible benign intents, the products supplied by these corporations may be exploited by rouge or state actors. As the Cambridge Analytica scandal showed, the combination of advanced technology and huge amounts of personal data can apparently be used as a powerful tool for partisan electoral interventions (Datatilsynet, 2019). The findings in this chapter show that algorithms, bots, trolls, disinformation and machine learning are examples of tactics and techniques state actors may use in order to influence public opinion through social media. The data-driven advertising model used by social media platforms, e.g., Facebook, is one reason why political influence campaigns may succeed. It is, however, hard to find empirical data on the exact effects, including the real-world behavioral consequences, of various foreign influential campaigns. In their article on how to protect elections from social media manipulation, Aral and Eckles argue that it is hard to regulate election manipulation through social media if we do not measure its effects (Aral and Eckles, 2019). They argue that an organized research agenda that informs policy is needed to protect democracies from foreign antidemocratic election interference operations (Aral and Eckles, 2019). In this chapter it is argued that the prevalence of microtargeting and the lack of human editing are the main reasons why campaigns in social media differs from more traditional forms such as PR, influence, tricks, cunnings and deceptions. Based on the findings from the literature review, this chapter will highlight the following key recommendations regarding ethics and advisable policies:

- Awareness campaigns may help citizens gain knowledge about the inner workings of social media and help them develop the skill of source criticism. Even though partisan electoral interventions can be conducted in a subtle manner, awareness about the possibility and likelihood of such operations happening may improve societal resilience.
- A stronger cooperation between the public and private sector is needed in order to respond to antidemocratic influential campaigns in social media. Citizens may also be of great help in the process of uncovering foreign electoral intervention attempts.
- Political parties need to be aware of the risk of foreign influence in upcoming elections. The parties do also need to be aware of the fact that information about a person's political affiliation is a special category of personal data (c.f. GDPR). Given that the Norwegian Broadcasting Corporation (NRK) discovered that six out of ten parliamentary parties in Norway are sharing such information with Facebook (Gundersen, 2019), there may be a need for more comprehensive ethical guidelines on what tools and products political parties should or should not implement on their websites.
- There is a need for further research on what the implications of foreign electoral intervention attempts are on real-world citizen behavior.

This chapter does not cover all aspects of the presented scope. The primary focus is to give a brief overview of how social media may be used in election manipulation, as well as highlight key recommendations regarding ethics and advisable policies.

Abstract

In recent years, social media platforms such as Facebook and Twitter have increasingly been exploited by foreign state actors in order to influence and disrupt democratic elections in other states. This chapter analyzes how foreign powers are using social media to achieve this goal. To answer the research question, this chapter uses a literature review as a methodology for collecting data in a systematic manner. Findings show that the advertising model used by social media platforms is one reason why political influence campaigns may succeed. One of the weaknesses with the model is that it enables foreign state actors to target specific citizens in other states without significant risks. Several tactics and techniques may be used in foreign-controlled interference operations in social media, including algorithms, bots, trolls, disinformation and machine learning. Election interference operations do not only break regulations, they do also have the capability to undermine the integrity of political elections, and hence, democracies. Based on the findings from the literature review, this chapter will highlight key recommendations regarding ethics and advisable policies.

Keywords: Disinformation, social media, election manipulation, microtargeting

3.1 Introduction

Manipulation tactics and techniques evolves as new digital communication technologies emerges. Traditionally we have PR, influence, tricks, cunnings, and deceptions. What is the difference between these traditional forms and the new ones in social media? The aim of this chapter is to describe and analyse election manipulation after the emergence of social media. The chapter will mainly focus on partisan electoral interventions, i.e. situations where foreign powers try to determine election results in other states (Levin, 2016*a*). Different manipulation tactics and techniques will be presented, and pros and cons of the data-driven advertising model will be discussed. The chapter will conclude with some recommendations regarding ethics and advisable policies on how public authorities, private companies and individuals may tackle disinformation campaigns in social media, and how to protect democratic elections in the digital age where new digital communication technologies are used to confuse and mislead citizens (Ilves et al., 2020).

Foreign interference in democratic processes is a complicated security threat. According to a report from The Norwegian Intelligence Service, different influence campaigns may have different objectives (Etterretningstjenesten, 2020). Examples of objectives may be to weaken citizens trust in democratic election processes, governments, politicians or the media (Etterretningstjenesten, 2020). Other objectives may be to steer the public debate in a certain direction, cast doubt on facts, or discredit certain opinion leaders, in order to to undermine trust and democratic electoral integrity (Etterretningstjenesten, 2020). Foreign state actors may also try to influence governmental decisions in order to archive their strategic objectives (Norwegian National Security Authority, 2020*b*). Such partisan electoral interventions may take place both openly and in secret (Norwegian National Security Authority, 2020*b*).

3.2 Methodology

The presented research questions will be investigated through the use of qualitative methods. A literature review will in this chapter be conducted in order to gather insights. In the literature review, publications that relate to the research question will be collected, evaluated and analysed. The number of sources reviewed in the methodology part is of limited extent, and the collected literature are mainly gathered from peer-reviewed journals. In the literature review, some newspaper articles have also been analyzed. In addition to these sources of information, sources of information does also include public reports, e.g., reports published by various authorities and Non-Governmental Organizations (NGOs). Throughout this chapter, Russia's interference in the 2016 US Presidential Election will be used as the main example on how state actors may interfere in other states' democratic elections.

3.3 Social media manipulation of elections

Illegitimate attempts at influencing democratic elections are nothing new (Levin, 2016*b*, Langemyr and Bakke, 2020). Great powers have for decades deployed partisan electoral interventions as a major foreign policy tool (Levin, 2016*b*). With the emergence of Web 2.0 and social media, democratic states may, however, be more vulnerable for such influence operations than ever before (Langemyr and Bakke, 2020). In a report by the Kofi Annan Commission on Elections and Democracy in the Digital Age, it is argued that new information and communication technologies pose difficult challenges for electoral integrity (Ilves et al., 2020). The commission highlights how social media have been used by foreign governments to interfere in elections around the globe, and that "disinformation has been weaponized to discredit democratic institutions, sow societal distrust, and attack political candidates" (Ilves et al., 2020).

According to an annual risk report, published by the Norwegian National Security Authority (NSM) in 2020, liberal democracies are vulnerable to data misuse, disinformation campaigns and strategic influence (Norwegian National Security Authority, 2020*a*). An Official Norwegian Report (NOU 2020: 6) conveys that as election outcomes may have major commercial and political consequences, it is reasonable to assume the existence of actors who have a great interest in influencing the outcome of democratic elections in their favor (NOU 2020: 6, 2020). According to the report, the risk is significant if the likelihood of success, motive and capacity are present (NOU 2020: 6, 2020). Norway, as an open society, is vulnerable for partisan electoral interventions (Norwegian National Security Authority, 2020*b*). According to a report from The Norwegian Intelligence Service, both China and Russia are states that have an interest in intervening in democratic electoral processes in Norway (Etterretningstjenesten, 2020).

Information and content on social media platforms can easily be altered or replaced by half-truths, fake content and similar (Norwegian National Security Authority, 2020*b*). When this happen, it can be hard for the citizens to ensure adequate source criticism (Norwegian National Security Authority, 2020*b*). A research study from the United States indicate that it is harder for older people than for young people to ensure an adequate level of source criticism (Jakobsen, 2019). In Norway, however, a research study indicate the opposite, i.e., that younger people do struggle more then older people in identifying and evaluating the sources upon which a given text relies (Jakobsen, 2019). According to NSM, the risk of misuse of open information for illegitimate purposes may be hard to sufficiently reduce without at the same time also weakening democratic values (Norwegian National Security Authority, 2020*b*).

3.3.1 Social media: A game changer?

Social media platforms have increasingly become an important part of many people's daily lives. The platforms provide easy access to news and relevant content, and facilitate sharing of content such as ideas and thoughts. Social media can be used to spread information about important issues such as public health and political issues to a large number of people. For many years, political institutions have been using social media to communicate with and engage voters. Through social media, political institutions can reach people they may not reach through other and more traditional media forms. Even though political institutions currently have the ability to use Facebook in order to target certain groups of individuals, their launched advertisements are publicly available. Such transparency can be viewed as crucial for building trust.

According to an article by Bennett and Lyon on data-driven elections, it is currently a pervasive assumption that elections can be won and lost on the basis of which party has the better data (Bennett and Lyon, 2019). The consequences of such an assumption may harm democratic values (Bennett and Lyon, 2019). As Bennett and Lyon argue, political microtargeting can be considered as a form of surveillance, and surveillance does, according to Bennett and Lyon, harm democratic values (Bennett and Lyon, 2019).

According to the European Parliamentary Research Service (EPRS), the tools and techniques used by state actors in order to disrupt or influence democratic processes in other states are constantly evolving (Berntzen, 2018). As new technologies are developed, rouge or state actors try to find ways to exploit or misuse these technologies in order to achieve their goals. According to the EPRS, algorithms, automation and machine learning are some of the techniques foreign state actors are currently exploring and using in order to boost the efficiency of influential campaigns (Berntzen, 2018).

3.3.2 Manipulation tactics and techniques

A variety of different tactics and techniques may be used in order to manipulate elections through social media. Some of these are the use of algorithms, bots, trolls, disinformation and machine learning. In the following text, a brief description of these different tactics and techniques will be given.

Algorithms

Social media platforms such as Facebook and Twitter use algorithms in order to "predict what users are interested in seeing, spark engagement and maximise revenues" (Berntzen, 2018). In short, an algorithm is simply a list of rules that may be used in order to find an answer to a particular problem. Facebook may for instance use algorithms in order to filter and prioritise content that a particular user receives based on the user's likes and shares. According to EPRS, social media platforms prioritize content that "sparks an emotional reaction and/or confirms already existing biases" as such content does tend to engage users the most (Berntzen, 2018). According to Benkler et al. (2018, p. 280), the "economic rewards of producing media that use anger, outrage, ridicule, and tribal bonding are immediate and significant". Using algorithms for this purpose may create echo chambers, isolating users within "social spaces that reinforce beliefs among like-minded users, contributing to political polarisation" (Berntzen, 2018). When Facebook data on 87 million users illegitimate were shared with Cambridge Analytica, these data were for instance used to target and mobilise voters prior to the 2016 US presidential election (Berntzen, 2018). The effects are still unknown.

Bots

One definition of a bot, or a robot, is as an automated account programmed to behave and interact in a human way, in particular on social media (Berntzen, 2018). In 2017 it was estimated that 15 % of all Twitter users were bots (Prier, 2017, as cited in Schia and Gjesvik, 2020). Bots operating on social media platforms can for instance be used to spread disinformation and so-called 'fake news'. In 2018 Twitter suspended up to 70 million accounts on their platform (Berntzen, 2018). The main reason for this was the growing concern about the impact of disinformation bots (Berntzen, 2018). According to the EPRS, some of the same bots that were used to spread disinformation prior to the 2017 French presidential election had earlier been used to spread disinformation prior to the 2016 US presidential election (Berntzen, 2018). This may indicate, according to the EPRS, that there exists a "black market for reusable disinformation bot networks" (Berntzen, 2018). As the accuracy of deep neural networks continue to improve, it is expected that the next generation of bots will use natural language processing (NLP) (Berntzen, 2018). With the use of NLP, it is likely that it will be much harder for citizens to separate a bot from a human than what is the case today.

Trolls

According to the EPRS, trolls may be defined as "human online agents, sometimes sponsored by state actors to harass other users or post divisive content to spark controversies" (Berntzen, 2018). The Kremlin-backed Russian Internet Research Agency (IRA), which interfered in the 2016 US presidential election, is one prominent example of coordinated, state-sponsored trolling (Berntzen, 2018). To shed light on the trolling activities initiated by Russia, Twitter did in 2018 disclose data on millions of tweets, images and videos that were linked to the 'troll farms' in St. Petersburg (Berntzen, 2018). However, it is important to highlight that not all trolls are sponsored by state actors. Trolls may very well be just ordinary citizens engaging in trolling activities (Berntzen, 2018).

Disinformation

According to a paper by Schia and Gjesvik at the Norwegian Institute of International Affairs (NUPI), disinformation has been described as "the deliberate creation and sharing of false and/or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain" (Digital, Culture, Media and Sport Committee, 2019, as cited in Schia and Gjesvik, 2020, p. 2). Disinformation has in international politics increasingly being used as a "tool for altering perceptions on a particular issue or subverting the political discourse by inserting false information" (Digital Forensic Research Lab, 2018, as cited in Schia and Gjesvik, 2020, p. 2).

Misinformation and propaganda have existed since ancient times. Even though misinformation has been designated as a major risk by the World Economic Forum, it is still not clear exactly how important misinformation and propaganda is in the age of social media (Bovet and Makse, 2019). According to a paper by Bovet and Makse, research studies do indicate that false news in social media is characterized by a much faster and broader diffusion than real news (Bovet and Makse, 2019). The attraction of the novelty of false news is one possible explanation for this tendency (Bovet and Makse, 2019). In 2018, Facebook did remove 583 million fake user accounts in order to combat the sharing of false news on their platform (Berntzen, 2018).

Social media platforms are vulnerable for false news as they were built for advertising and users' engagement. Social media does also provide users with a lot of information, and hence, making it hard for users to maintain an adequate amount of source criticism. In social media "the distinction between sketchy and reputable news sites" is also diminished (Yochai Benkler and Roberts, 2018). The ability for false and misleading information to rapidly spread on social media platforms have been described as societal vulnerability, and also, a threat to democracy (Schia and Gjesvik, 2020). Even though it is still uncertain to which extent the effectiveness of these forms are, the consequences can be serious (Schia and Gjesvik, 2020).

Machine learning

As deep-learning algorithms evolve, it may become easier for rouge or state actors to target specific individuals in more engaging ways. Using artificial intelligence (AI) and machine learning (ML), rich content such as sound, video and images can be tailored to users' personalities, interests and backgrounds (Berntzen, 2018). In the future it is not unlikely that bots using NLP will be used to further engage users in online discussions (Berntzen, 2018), or perhaps even try to steer the discussion in a certain direction. For the citizens it can be harder to distinguish between real and fake content, as the fake content may be highly realistic (Berntzen, 2018). As deep learning algorithms evolve, it may be easier to alter and manipulate rich content. It is for instance likely that the quality of manipulated videos will increase in the near future, and such powerful technologies can also be used by rouge or state actors. So-called 'deep fakes' can for instance be used to make it look like a person, for instance a politician, did say or do something that he or she did not say or do (Berntzen, 2018). It is likely that such content may be used by foreign states to influence democratic processes in other states.

3.4 Discussion

3.4.1 Social media and the 2016 US presidential election

According to a study from the University of Oxford, the US and the USSR/Russia intervened in one of every nine competitive national level executive elections between year 1946 and 2000 (Levin, 2016*b*). Hence, partisan electoral interventions are nothing new. The emergence of web 2.0 in general, and the emergence of social media platforms in particular, have, however, changed how such operations are being conducted. With the emergence of the Internet, geographical distance did become irrelevant (Schia and Gjesvik, 2020). Social media have later enabled actors around the world to reach millions of people both easily and cheap. Social media platforms do also allow for more direct forms of communication (Schia and Gjesvik, 2020). Such communication forms can, however, "bypass the media's traditional 'gatekeeping' role and thus be exploited to push false news stories (Schia and Gjesvik, 2020, p. 3). A further key feature is the anonymity the Internet offers to "those wishing to conceal their true identity" (Schia and Gjesvik, 2020, p. 3). The presence of online trolls have, according to Schia and Gjesvik, become a "defining element of online communication" (Schia and Gjesvik, 2020, p. 3).

A variety of techniques, including trolls, were in the 2016 US presidential election used to sway public discourse. Quantitative data on the effects of these influential campaigns are hard to find (Yochai Benkler and Roberts, 2018). However, one quantitative research study on American citizens' exposure to 'fake news' on Facebook were published by Hunt Allcott and Matthew Gentzkow in 2017. In their study, Allcott and Gentzkow found that prior to the US election, an average American adult would have seen about 1.14 fake stories (Allcott and Gentzkow, 2017). According to Benkler et al., Allcott and Gentzkow's fundamental point was that even if the fake news had a large audience in absolute terms, "this still translates into a tiny fraction of the campaign news-related websites that people visited, and an even tinier fraction of the stories to which they were exposed" (Yochai Benkler and Roberts, 2018, p. 286). However, there is a need for more thorough research on the effects of influential campaigns on social media.

In the aftermath of the 2016 US presidential election and the subsequent investigation into suspected Russian interference, disinformation campaigns have become a matter of critical importance on policy agendas around the world (Schia and Gjesvik, 2020). Following the US election, there have been public concerns on what role disinformation and trolls played in the election of Donald Trump (Allcott and Gentzkow, 2017). Prior to the presidential election, false content circulated in large numbers on social media platforms like Twitter and Facebook. It is, however, Facebook who have been described as sitting at "the center of the epistemic crisis after 2016" (Yochai Benkler and Roberts, 2018, p. 269). According to Benkler et al., "Facebook's willingness to sell advertising to Russian operatives and its hosting of a number of prominent Russian sockpuppet groups" put the company on the public agenda in the middle 2017 (Yochai Benkler and Roberts, 2018, p. 269). By early 2018 "the long-simmering story of Cambridge Analytica", did, according to Benkler et al., boil over and was "spilled onto Facebook's lap" Yochai Benkler and Roberts, 2018, p. 269).

3.4.2 Social media, trust and democracy

Currently, the dominant business model of the Internet is based on surveillance. Information about our online behaviour is being collected at large scale in order to give us customized advertising and content. Shoshana Zuboff, who coined the term 'surveillance capitalism' in 2014, have warned that the shift from a mass to an individual-oriented structure of consumption may undermine personal autonomy and erode democracy (Laidler, 2019). Corporations are increasingly mining users' information to predict and shape their behavior. Despite possible benign intents, the products supplied by these corporations, for instance Facebook, may be exploited by rouge or state actors. As the Cambridge Analytica scandal showed, the combination of advanced technology and huge amounts of personal data can be used as a powerful tool for partisan electoral interventions (Datatilsynet, 2019). According to Bennett and Lyon, the scandal would "not have been such without Facebook, for which both 'prediction' and 'personalisation' are central" (Bennett and Lyon, 2019, p. 3).

According to Matthew Crain and Anthony Nadler, the data-driven digital advertising model in social media has played a key role in facilitating political manipulation campaigns (Crain and Nadler, 2019). In this regard, it is important to highlight that IRA's influence campaign during the US presidential election were not a direct 'misuse' of social media platforms (Arild Bergh, 2020). What the IRA did was actually to use the platforms as they were intentionally designed (Arild Bergh, 2020, Yochai Benkler and Roberts, 2018). They did, however, use the platforms for other purposes than what the platforms considered as 'normal use'. According to FFI, evaluating a propose can be hard on social media platforms as it requires human editing (Arild Bergh, 2020). As social media platforms were not designed to require such editing processes, Benkler et al. argue that "the basic business of Facebook, when applied to political communication, presents a long term threat to democracy" (Yochai Benkler and Roberts, 2018). This threat is existing whether microtargeted manipulation is used by the "incumbent government to manipulate its population or by committed outsiders bent on subverting democracy" (Yochai Benkler and Roberts, 2018).

As Roose writes in the the New York Times, it is important to be aware of the fact that social media platforms are owned by corporations seeking to maximize growth and profitability (Roose, 2018). According to Roose, the social media apparatus that Russia 'exploited' in the 2016 US presidential election has been very profitable for these companies (Roose, 2018). As microtargeted behavioral marketing is Facebook and similar companies' core business, influential campaigns on social media can be challenging to tackle without implementing new regulations and policies.

According to a paper by Crain and Nadler, the digital advertising systems "have been built with capacities that can easily be weaponized” (Crain and Nadler, 2019, p. 372). When political operatives or foreign states weaponize such powerful technology, they may use it to “identify weak points where groups and individuals are most vulnerable to strategic influence” (Crain and Nadler, 2019, p. 372). In such cases, citizens’ data is turned against them (Crain and Nadler, 2019). Crain and Nadler argue that "the very capacities of digital ad systems that facilitate such weaponized communication need to be recalibrated to better serve democratic ideals" (Crain and Nadler, 2019, p. 372). They further argue that this approach is better than trying "to remove ’bad actors’ from abusing digital advertising systems (Crain and Nadler, 2019, p. 372). “Increasing advertising transparency, expanding the data rights of individuals, and attenuating advertisers’ capability to carve audiences into smaller and smaller segments” are some of the proposals presented by Crain and Nadler in order to address these issues (Crain and Nadler, 2019, p. 372).

In 2019, the Norwegian Broadcasting Corporation (NRK) discovered that six out of ten parliamentary parties in Norway were sharing information about their web site visitors with Facebook (Gundersen, 2019). As the shared information could indicate users’ political views, it may be argued that such information should have been better protected. Gisle Hannemyr at the University of Oslo argues that people should not need to be experts in Information Technology (IT) in order to avoiding the sharing of such sensitive data with Facebook (Gundersen, 2019). It may be argues the ethical guidelines is needed in order to better ’protect’ citizens. In addition, it is important to highlight that both individual citizens and the society as a whole can benefit form users being more aware of what information they share on social media (Norwegian National Security Authority, 2020*b*). NSM argues that as democratic societies are vulnerable for disinformation campaigns, all Norwegian citizens need to also be aware of this risk in order to improve societal resilience (Norwegian National Security Authority, 2020*b*).

3.4.3 Protecting elections in the digital age

Data-driven-electioneering is a global phenomenon, and according to Bennett and Lyon, Cambridge Analytica was working in approximately thirty countries before it closed down (Bennett and Lyon, 2019). The emergence of social media may be described as representing a "shift from geographic based targeting to more individualised messaging based on predictive models and scoring" (Bennett and Lyon, 2019, p. 4). In their paper, Schia and Gjesvik warns that "recent technological developments risk undermining established equilibrium for dealing with false or misleading information" (Schia and Gjesvik, 2020). In addition, Schia and Gjesvik argue that the lack of transparency characteristic of the social media platforms has made measuring the societal impact of disinformation accurately nearly impossible (Schia and Gjesvik, 2020).

In response to the Cambridge Analytica scandal, tens of millions of dollars have been spent by social media platforms on “disinformation research tracking online campaign messaging dissemination and consumption on social media” (Briant, 2020). It is possible that such effects may be effective in order to reduce foreign influential campaigns. However, as Emma

L. Briant The Brookings Institution argues, a focus “solely on tracking content will tend toward solutions such as content removal” ignores the issue of what truly motivates platforms to act (Briant, 2020). The tracking approach is also, according to Briant, “unable to identify the creators of content, the corporations, governments, or other organizations funding it, or their use and misuse of data” (Briant, 2020). Briant further argues that in order "to fully understand the issues raised by digital influence campaigns, or develop new ways to respond to them, it is necessary to also focus on exposing and responding to the rapidly expanding digital influence industry" (Briant, 2020). An tracking approach may also miss other ideas, and also, other responsible parties (Briant, 2020). According to Briant, efforts at “the policy level to regulate, legislate data privacy, and pressure platforms to remove non-compliant content” still have a long way to go (Briant, 2020).

Liberal and open societies are vulnerable for disinformation and influential campaigns (Norwegian National Security Authority, 2020*b*). It is likely that important democratic processes in the future may be exposed to antidemocratic influence campaigns by foreign states. In such campaigns, social media platforms may be exploited by foreign state actors in order to influence the outcome of democratic elections in their favor (Norwegian National Security Authority, 2020*b*). As the tactics and techniques evolve over time, partisan electoral interventions may be hard to detect (Norwegian National Security Authority, 2020*a*). As the social media platforms have implemented measures to prevent the sharing of fake news, actors have started using other kinds of influential activities, e.g., the sharing of polarized content published by established news platforms (Norwegian National Security Authority, 2020*a*). As influential campaigns in social media may be hard to detect for citizens, NSM argue that traditional media platforms have a responsibility to present their stories in a balanced way and to provide thorough information about which sources they used (Norwegian National Security Authority, 2020*a*). According to a report published by the Norwegian Defence Research Establishment, it is unrealistic for a democracy to stop influence attempts in a direct manner (Arild Bergh, 2020). In the report it is argued that influence attempts must be handled through increased knowledge about how such operations work and which groups it might affect (Arild Bergh, 2020). Resilience in the society may for instance be fostered through educational programs and awareness campaigns.

The findings from the literature review show that algorithms, bots, trolls, disinformation and machine learning are examples of tactics and techniques state actors may use in order to influence public opinion through social media. The data-driven advertising model used by social media platforms is one explanation for why political influence campaigns may succeed. There are, however, hard to find empirical data on what the exact effects, e.g., the real-world behavioral consequences, of such campaigns might be. In their article on how to protect elections from social media manipulation, Aral and Eckles argue that it is hard to regulate election manipulation through social media if we do not measure its effects (Aral and Eckles, 2019). They argue that an organized research agenda that informs policy is needed to protect democracies from foreign antidemocratic election interference operations (Aral and Eckles, 2019).

Based on the findings from the literature review, this chapter presents the following recommendations regarding ethics and advisable policies:

- Awareness campaigns may help citizens gain knowledge about the inner workings of social media and help them develop the skill of source criticism. Even though partisan electoral interventions can be conducted in a subtle manner, awareness about the possibility and likelihood of such operations happening may improve societal resilience.
- A stronger cooperation between the public and private sector is needed in order to respond to antidemocratic influential campaigns in social media. Citizens may also be of great help in the process of uncovering foreign electoral intervention attempts. Increased regulation on "personalisation algorithms" in social media may also be a powerful tool in mitigating the consequences of such influential campaigns.
- Political parties need to be aware of the risk of foreign influence in upcoming elections. The parties do also need to be aware of the fact that information about a person's political affiliation is a special category of personal data (c.f. GDPR). Given that the Norwegian Broadcasting Corporation (NRK) discovered that six out of ten parliamentary parties in Norway are sharing such information with Facebook (Gundersen, 2019), there may be a need for more comprehensive ethical guidelines on what tools and products political parties should or should not implement on their websites. The question concerning if political advertisement on social media should be allowed or not is not an easy question to answer (Bennett and Lyon, 2019). To implement a possibility for users to opt-in to political advertising on social media is one solution that may increase user control. In this regard, it is important to highlight that even though political influential campaigns in some cases do rely on targeted advertisements, the advertisements themselves may not necessarily refer to political content (Crain and Nadler, 2019).
- There is a need for further research on what the implications of foreign electoral intervention attempts are on real-world citizen behavior.

3.5 Concluding remarks

Social media have become an important part of many people's daily lives. The platforms provide easy access to news and relevant content, and can be used to spread information about important issues such as public health and important political issues to a large number of people. Social media may, however, also be used by foreign state actors in order to intervene in electoral processes in other states. In this chapter the main differences between traditional PR, influence, tricks, cunnings, and deceptions, and the new forms used in social media has been described. The chapter has also presented an overview of how social media may be used in election manipulation, and what tactics and techniques state actors may be used in order to influence public opinion through social media. The chapter has also highlighted some key recommendations regarding ethics and advisable policies, and how public authorities, private companies and individuals may tackle foreign electoral intervention attempts in social media.

Social media platforms have increasingly been exploited by foreign state actors in order to influence and disrupt democratic elections in other states. As design "weaknesses" in social media exist, there is a risk of these being exploited by rogue and foreign state actors for a variety of reasons. The main "weakness" presented in this chapter is the data-driven advertising model used by most social media. In the 2016 United States presidential election, Facebook's business model did for instance enable foreign state actors to target specific citizens in other states without significant risks. There are also other risks with these kinds of business models. As advertisements are Facebook's main revenue stream, the company wants people to actively use the company's platform, as this gives the company the possibility to show the users more advertisements. Hence, it is important for Facebook to engage their users, motivating them to participate in online discussions on the social platform, to share information about their lives, and to provide links to external content and similar. As argued by researchers at the Norwegian Institute of International Affairs, it is possible to describe the ability of false and misleading information to spread rapidly online as a societal vulnerability and a threat to democracy (Schia and Gjesvik, 2020). As social media platforms work in this way, it may be challenging to mitigate foreign-controlled interference operations in social media.

Based on the findings from a literature review, this chapter has highlighted the importance of awareness campaigns. Such campaigns may help citizens gain knowledge about the inner workings of social media and help them develop the skill of source criticism. Even though partisan electoral interventions can be conducted in a subtle manner, awareness about the possibility and likelihood of such operations happening may improve societal resilience. This chapter does also highlight the importance of more cooperation between the public and private sector. Such a cooperation is needed in order to respond to antidemocratic influential campaigns in social media. Citizens may also be of great help in the process of uncovering foreign electoral intervention attempts. Findings from the literature review also suggest that there is a need for political parties to be more aware of the risk of foreign influence in upcoming elections. As findings from the literature review indicate that six out of ten parliamentary parties in Norway share special categories of personal data with Facebook (Gundersen, 2019), comprehensive ethical guidelines on what tools and products political parties should or should not implement on their websites may be needed. This chapter does also highlight the need for further research on what the implications of foreign electoral intervention attempts are on real-world citizen behavior.

3.5.1 Limitations

This chapter have not covered all aspects of the proposed scope. The primary focus was to give a brief overview of how social media may be used in election manipulation, as well as highlighting key recommendations regarding ethics and advisable policies. The findings in this chapter is also based on a limited number of sources. It is reasonable to believe that this has affected the quality of the presented results.

Bibliography

Allcott, H. and Gentzkow, M. (2017), 'Social media and fake news in the 2016 election', *Journal of Economic Perspectives* **31**(2).

Aral, S. and Eckles, D. (2019), 'Protecting elections from social media manipulation', *Science* **365**(6456), 858–861.
URL: *https://science.sciencemag.org/content/365/6456/858*

Arild Bergh (2020), 'Påvirkningsoperasjoner i sosiale medier - oversikt og utfordringer'.
URL: *https://www.ffi.no/publikasjoner/arkiv/pavirkningsoperasjoner-i-sosiale-medier-oversikt-og-utfordringer*

Bennett, C. J. and Lyon, D. (2019), 'Data-driven elections: implications and challenges for democratic societies', *Internet Policy Review* **8**(4).
URL: *https://policyreview.info/data-driven-elections*

Berntzen, N. (2018), 'Computational propaganda techniques'.
URL: *https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_ATA(2018)628284*

Bovet, A. and Makse, H. A. (2019), 'Influence of fake news in twitter during the 2016 us presidential election', *Nature Communications* **10**(7).

Briant, E. L. (2020), 'We need tougher action against disinformation and propaganda'.
URL: *https://www.brookings.edu/blog/techtank/2020/07/15/we-need-tougher-action-against-disinformation-and-propaganda/*

Crain, M. and Nadler, A. (2019), 'Political manipulation and internet advertising infrastructure', *Journal of Information Policy* **9**, 370–410.
URL: *https://www.jstor.org/stable/10.5325/jinfopoli.9.2019.0370*

Datatilsynet (2019), 'Digital targeting of political messages in norway'.
URL: *https://www.datatilsynet.no/globalassets/global/dokumenter-pdfer-skjema-ol/rettigheter-og-plikter/rapporter/pa-parti-med-teknologien---engelsk.pdf*

Etterretningstjenesten (2020), 'Fokus 2020'.
URL: *https://www.forsvaret.no/aktuelt-og-presse/publikasjoner/fokus*

Gundersen, M. (2019), 'Nettsidene til 6 av 9 stortingspartier sender dine besøksdata til facebook'.
URL: *https://nrkbeta.no/2019/07/12/nettsidene-til-6-av-9-stortingspartier-sender-dine-besoksdata-til-facebook/*

Ilves, T. H., Persily, N., Stamos, A., Stedman, S., Chinchilla, L., Leterme, Y., Heyzer, N., Okolloh, O., Sweeney, W., Smith, M. and Zedillo, E. (2020), 'Protecting electoral integrity in the digital age: The report of the kofi annan commission on elections and democracy in the digital age'.

URL: *https://www.kofiannanfoundation.org/app/uploads/2020/01/f035dd8e-kaf_kacedda_report_2019_web.pdf*

Jakobsen, S. E. (2019), 'Hvem deler falske nyheter på nett?'.
URL: *https://forskning.no/media/hvem-deler-falske-nyheter-pa-nett/1278513*

Laidler, J. (2019), 'High tech is watching you'.
URL: *https://news.harvard.edu/gazette/story/2019/03/harvard-professor-says-surveillance-capitalism-is-undermining-democracy/*

Langemyr, H. B. and Bakke, S. (2020), 'Innlegg: Dataangrep mot demokratiet – et varsel om illegitim valgpåvirkning'.
URL: *https://www.dn.no/innlegg/datasikkerhet/hacking/cyberkriminalitet/innlegg-dataangrep-mot-demokratiet-et-varsel-om-illegitim-valgpavirkning/2-1-867612*

Levin, D. H. (2016*a*), 'Partisan Electoral Interventions by the Great Powers: Introducing the PEIG Dataset', *Conflict Management and Peace Science* **36**.

Levin, D. H. (2016*b*), 'When the Great Power Gets a Vote: The Effects of Great Power Electoral Interventions on Election Results', *International Studies Quarterly* **60**(2), 189–202.

Norwegian National Security Authority (2020*a*), 'Helhetlig digitalt risikobilde 2020'.
URL: *https://nsm.no/regelverk-og-hjelp/rapporter/helhetlig-digitalt-risikobilde-2020*

Norwegian National Security Authority (2020*b*), 'Risiko 2020'.
URL: *https://nsm.no/aktuelt/risiko-2020*

NOU 2020: 6 (2020), 'Frie og hemmelige valg — Ny valglov'.
URL: *https://www.regjeringen.no/no/dokumenter/nou-2020-6/id2703131/?ch=2#kap4*

Roose, K. (2018), 'Social media's forever war'.
URL: *https://www.nytimes.com/2018/12/17/technology/social-media-russia-interference.html*

Schia, N. N. and Gjesvik, L. (2020), 'Hacking democracy: managing influence campaigns and disinformation in the digital age', *Journal of Cyber Policy* .

Yochai Benkler, R. F. and Roberts, H. (2018), *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford University Press.

Chapter 4

Coordinated Inauthentic Behavior

By Sigrid Anne Hafsahl Karset

Executive summary

Coordinated inauthentic behavior is a term that has seen a great upsurge in the recent years. The phenomenon is described as the usage or sharing of one or multiple accounts in order to misrepresent, mislead, or artificially boost content on the internet in a coordinated manner. Coordinated inauthentic behavior has been the basis for the removal of several pages, groups, and accounts through out the last couple of years.

Facebook was the first platform that started using and defining the term coordinated inauthentic behavior. And shortly after, other big platforms such as Twitter, TikTok and YouTube followed as well.

The paper shows that various platforms on and off the internet may have different interpretations of what constitutes as coordinated inauthentic behavior. These definitions varies in order to fit the needs and the content of the given platform. However, the underlying meaning behind the term stays the same.

An important fact to note is that coordinated inauthentic behavior and fake news are two separate topics. Content spread by people engaging in coordinated inauthentic behavior can be false news and narratives. However, the content may also be authentic. The main difference between the two is that fake news are news stories that wholly or partially contains untrue information, while coordinated inauthentic behavior is groups, accounts, or pages lying about who they are and/or their purpose.

The impact of coordinated inauthentic behavior, as explained in this paper, have the potential to be highly severe. As the usage of the internet grows, coordinated inauthentic behavior has the potential to affect the unfolding of countries' elections and citizens views on other countries and conflicts, amongst other impactful subjects.

This paper shows that it is important to employ proper defences against this type of behavior, as well as good communication amongst platforms and governments. Since punishment beyond the digital world might not always be applicable due to the anonymity of the internet as well as different laws in different countries, the defence against it should be focused on. These defences need also constantly be improved in order to keep up with the ever-changing environment that is the Internet.

4.1 Introduction

On July 31 2018 Facebook reported Gleicher and Stamos (2018) having removed 32 pages and accounts from their platforms due to them being involved in "coordinated inauthentic behavior". Although Facebook have been removing content and accounts for various reasons before, this was the first instance where the term coordinated inauthentic behavior was used as their reasoning for the removals. From the first instance in July of 2018 until December of that year, Facebook reported having removed close to 2000 groups, accounts and pages due to the coordinated inauthentic behavior phenomenon Greene and Gleicher (2018), Facebook (2018), Gleicher and Rodriguez (2018), Gleicher (2018). The cause of this sudden upsurge in reported instances of the phenomenon is related to Facebook's update of their community standards, where they in June of 2018 included the term under the Misrepresentation section Facebook (n.d.*b*).

But what really is "Coordinated Inauthentic Behavior"? Is it something that has been around for a while, or is it something new? Is it just a Facebook "thing", or is it present elsewhere? And most importantly; should we be afraid of it?

4.2 Coordinated inauthentic behavior

Facebook defines Facebook (n.d.*a*) inauthentic behavior as the use or sharing of one or multiple accounts in order to misrepresent oneself or others, artificially boosting the popularity of content, mislead people or Facebook themselves about the purpose of a page/group/account, or engaging in behaviors designed to enable other violations of their community standards. Coordinated inauthentic behavior is the use of multiple Facebook assets, working in concert to engage in such inauthentic behavior.

Facebook might have popularised the term "coordinated inauthentic behavior", however, such behavior is nothing new. If we take a look at the dictionary definitions of the term, inauthentic behavior is someone acting "fake" or appears to be something they're not *Inauthentic* (n.d.). Coordinated inauthentic behavior then becomes a collective effort to convince one or more people that something is, or they are something they're not.

Using this definition, we can find more than enough cases of coordinated inauthentic behavior throughout history. The behavior exhibited by the Nazis in their propaganda against the Jews, or the Soviet's display of the US during the Cold War. Even the siege of Troy could fall under this definition of coordinated inauthentic behavior (If we believe that story to be true). Most history is filled with people showcasing some form of coordinated inauthentic behavior. However, in the mid to late nineteen hundreds two new inventions, that would have huge impact on the way coordinated inauthentic behavior is performed, was created. That is the Internet and the World Wide Web Andrews (n.d.), *History of the Web* (n.d.).

4.2.1 Inauthentic behavior on the Internet

With the invention and popularisation of the internet and the world wide web, people could now create online personalities and engage in inauthentic behavior without having to show their face or even leaving their house. With this new platform coordinated inauthentic behavior didn't have to be performed by multiple people. On this platform one person could imitate a hundred.

A true showcasing of the impact anonymity one can achieve on the internet occurred in Greater Manchester, England, in 2003. Where one young boy attempted to murder his own friend due to manipulation over the internet Bachrach (2006).

A boy named John had created multiple accounts and personalities to interact with another boy named Mark over an online chatroom. John used fake accounts in order to connect and become friends with Mark, as well as creating love interests for him. John also fabricated false stories of kidnappings and killings of the fake people Mark had bonded with over the chatroom, in order to emotionally upset him. When John, through a fake personality supposedly working for the Secret Service, had built a trust bond with Mark, he asked Mark to kill him (John). This was so that Mark could prove to the "Secret Service agent" that he was worthy of joining the service.

The text "U want me 2 . . . kill him . . . ?? that's wot ur askin me?" was one of the last messages Mark sent in the chatroom before heading out and stabbing John.

John fortunatley survived the stabbing, and the investigation of the incident led to the discovery of Johns PC. Analysis of the data and activity on this computer showed that John had created 193 different e-mail addresses and 98 accounts for the online chatroom, all of which had communicated with Mark. John had emotionally manipulated Mark over this chatroom, and eventually fabricated his own murder.

This story truly shows what one person alone can achieve by fabricating and coordinating behavior in multiple fake personas over the internet.

In the Manchester case only one person used the internet to display inauthentic behavior, imagine what hundreds of people could achieve if they banded together to perform inauthentic behavior. This is when we get the coordinated inauthentic behavior that Facebook is currently advocating against.

4.2.2 Cases of coordinated inauthentic behavior

To get a better understanding of the phenomenon, lets look at the first couple of recorded cases of coordinated inauthentic behavior on Facebook's platform.

Note that we're only looking at cases that happened after Facebook included the term coordinated inauthentic behavior in their community standards. There have of course been instances of coordinated inauthentic behavior before the update of the standards, but we are here looking at incidents where removal of accounts and/or pages has been justified by engagement in coordinated inauthentic behavior.

Case 1

July 31, 2018 Gleicher and Stamos (2018). This is the first recorded case of the removal of accounts and pages for engaging in coordinated inauthentic behavior. There were in total 32 pages and accounts that were removed in this incident. Most of these pages could be linked to the Russian-based IRA (Internet Research Agency), an agency which engages in online influence operations on behalf of Russian business and political interests Chen (2015). Pages named "Aztlan Warriors", "Black Elevation", "Mindful Being" and "Resisters" were removed. The pages engaged in advertising and propaganda spreading to over 300 000 followers. One of the pages, "Resisters", organized a protest in DC where over 2 600 people had planned to take part. Facebook managed to identify the coordinated inauthentic behavior in these pages and accounts, and luckily put a stop to the planned protest.

Case 2

The next case of removal happened on the 21th of August, 2018 Greene and Gleicher (2018). This time the pages and accounts where linked to both Iran and Russia. A network called "Liberty Front Press" was identified to have been engaging in coordinated inauthentic behavior. They were concealing their true identity and location while targeting people in the Middle-East, Latin-America, The UK, and The US. Further investigation into these accounts also revealed that they had been involved in traditional cyber attacks such as compromising accounts and spreading malicious software.

In the same report, Facebook announced that they also had removed several pages and accounts linked to Russian military intelligence services. These pages and accounts where focused on politics in Syria and Ukraine and predominantly spread pro-Russian and pro-Assad content.

These are the first of many cases where accounts, groups, and pages were removed due to coordinated inauthentic behavior. We can see that in these instances, coordinated inauthentic behavior was used to target people in other countries. Even though there are of course cases where coordinated inauthentic behavior is used to target people in their own country, much of the occurrences of the phenomenon has ties to foreign states. The reasons why in-country occurrences seems less frequent than the out-of-country instances might be due to the overall scale, the ease of detectability, and dedicated resources. A foreign state, as seen in the previously mentioned cases, will often use coordinated inauthentic behavior to influence people on a nationwide scale. An occurrence of such scale is often more detectable and desirable to avoid than smaller cases within a given country. However, as mentioned, the small scale cases of coordinated inauthentic behavior do occur, and is in no way more exempt to the rules than larger scale occurrences.

Facebook was the first big platform to include coordinated inauthentic behavior in their community standards and using the term as a cause of removal. Not long after, other platforms followed as well. Platforms such as Twitter, YouTube, and TikTok are all removing accounts, pages and groups due to coordinated inauthentic behavior Cohen (2020), Neuman (2019), Statt (2020).

Twitter reported TwitterSafety (2018) in August 2018 that they had removed 770 accounts for engaging in coordinated manipulation. These accounts were found to be originating from Iran even though many of them claimed to be US based. Some of the accounts were linked to Liberty Front Press which is the same network behind many of the accounts and pages removed by Facebook on August 21th, 2018, as mentioned earlier. Alphabet inc also reported having suspended several YouTube accounts linked to the Liberty Front Press network Schwartz (2018). This goes to show just how many opportunities a network can have to reach out to the general public. This also shows how important it is for the different social media platforms to work in unison in order to identify and combat coordinated inauthentic behavior.

4.2.3 Different views on coordinated inauthentic behavior

Coordinated inauthentic behavior is a term that can be interpreted and used differently. As previously described, coordinated inauthentic behavior can be as simple as someone collectively trying to convince people that something is, or they are, something they're not. However, when the term is being applied to different places, what falls under the term might differ in order to fit the given platform. Because of this, what one platform categorizes as coordinated inauthentic behavior might not be accepted as so on another platform.

An example of this was when the people of TikTok targeted Trumps rally in Tulsa, Oklahoma Lyons (2020). One of the more talked about subjects in June of 2020 was the poor attendance of Trump's presidential rally in Tulsa. Pictures an videos from the rally shows thousand of seats empty even though the event planners had reported a full set stadium Lorenz et al. (2020).

This turned out to be the workings of many young and old Trump-disagreers, initiated on the popular video posting platform TikTok. A TikTok Laupp (2020) (video) posted by Mary Jo Laupp is believed to have been the video that started it all. In the clip, Mrs. Laupp encouraged people to "go reserve tickets now and leave him standing alone there on the stage". This clip is said to have started a movement of people reserving tickets for the rally to artificially inflate the attendee numbers. The TikTokers coordinated this movement among themselves by sharing tips and advice on how to sign up for the rally and how to acquire all the details needed to sign up. The TikTokers only left up their videos for 24-48 hours in order to avoid wide spread detection Statt (2020).

This stunt was fast labeled as coordinated inauthentic behavior by platforms and news sites. However, when Facebook's head of cybersecurity, Nathan Gleicher, commented on the case Gleicher (2020), he explained that this stunt would not fall under Facebook's definition of coordinated inauthentic behavior since it did not involve the use of fake accounts, and it was used to influence people off the platform, as opposed to people on the platform. The stunt doesn't necessary fall under TikTok's community guidelines either TikTok (2020):

> "Do not: Engage in coordinated inauthentic activities (such as the creation of accounts) to exert influence and sway public opinion while misleading individuals, our community or the larger public about the account's identity, location or purpose"

The TikTok users did not try to mislead people about their identity, location, or purpose. Although they did use their accounts to share information and engage in behavior that exerted influence on the general public in an inauthentic manner.

4.2.4 CIB and fake news

Coordinated inauthentic behavior might sound familiar to or related to the popular term *fake news*. When reading articles about the phenomenon, you might come across the term fake news as well. However, is coordinated inauthentic behavior the same as fake news?

Although coordinated inauthentic behavior is false behavior, it is not fake news. Fake news is defined as false news stories that is spread, often using the internet, with purposes such as influencing political views, generating revenue, or discrediting/promoting a public figure or company *fake news* (n.d.). As an example of fake news, we can look at some of 2019 most viewed fake news stories on the internet Gilbert (2019). The most viewed news story, with over 29 million views Press (2019), claimed that Trump's father was a pimp and that his grandfather was a member of the KKK, none of which has been proven to be true. Another wildly viewed story of 2019 was that Nancy Pelosi diverted Social Security money for the impeachment inquiry, and the story of Alexandria Ocasio-Cortez proposing a motorcycle ban. These again were quickly proven not to be true.

As with this definition and examples of fake news as well as the aforementioned description of coordinated inauthentic behavior, we can see that fake news are news stories that wholly or partially contains untrue information, while coordinated inauthentic behavior is groups, accounts, or pages lying about who they are and their purpose. The big distinction between fake news and coordinated inauthentic behavior is that fake news can be spread by people engaging in coordinated inauthentic behavior. However, not every network involved in coordinated inauthentic behavior shares fake news. A network can very well share true stories and still be removed due to coordinated inauthentic behavior. Its the lying about the network's purpose, identity, and/or location that classifies it as a network consistent with coordinated inauthentic behavior.

4.2.5 Penalty

Coordinated inauthentic behavior is proven to be unwanted both on social media platforms as well as outside the digital world. However, what is the penalty for engaging in such behavior? And are these penalties effective?

If we look at the popular social media platforms Facebook/Instagram, Twitter, and TikTok, they do not define a particular penalty for engaging in inauthentic behavior. The community guidelines of the platforms specifies only that penalties for the violations of the guidelines depends on the severity of the violation as well as any previous history of violations Facebook (n.d.*b*), TikTok (2020), Twitter (2020). A penalty could be a warning, content removal, temporarily suspension, indefinite ban, or involvement of law enforcement.

When looking at the previously described cases of coordinated inauthentic behavior, with Liberty Front Press and the IRA, we have seen that the engagement in such behavior led to the removal of both content and accounts. Both law enforcement and congress were contacted regarding networks and their behavior. However, punishment beyond the digital world is a much more complex task. First of all one have to identify the people behind the networks and accounts, that in it self is convoluted. If the bad actors are identified, there is a substantial chance that the actor is based in another country. Foreign states do not necessary have the same laws as the investigating country, and they might not acknowledge the activity as a violating act.

No country has any direct law against coordinated inauthentic behavior, considering it is a fairly new term. However, there are laws that applies well to the situation such as, if we take the US as an example, the *18 U.S. Code §1038 - False information and hoaxes*[1] and the *15 U.S. Code §1125 - False designations of origin, false descriptions, and dilution forbidden*[2]. Under these laws, punishment such as fines and prison up to five years is applicable. As mentioned, if the bad actor is identified as a foreign national, there might be problems convicting him/her, so punishment beyond what can be done digitally is not always admissible.

The question remains whether or not these penalties for engaging in coordinated inauthentic behavior are effective or not.

In simple terms, removing an account or page will deny that certain account/page from reaching out to its followers. Therefore effectively putting a stop to its engagement in coordinated inauthentic behavior. However, since this being the internet, banning the account will not prevent the source from engaging in such activity from elsewhere. Due to the vastness and complexity of the internet, one can create new accounts and associated information with ease, and unless the investigating party can directly track the person's internet activity, easily avoid being tied to the previous account as well.

Unless one can stop the problem at its source, that being before the account/page can reach out to the internet, the preventative efforts will not be a hundred percent effective. That does not however mean that they aren't mitigating the problem.

[1] False information and Hoaxes: https://www.law.cornell.edu/uscode/text/18/1038

[2] False designations of origin, false descriptions, and dilution forbidden: https://www.law.cornell.edu/uscode/text/15/1125

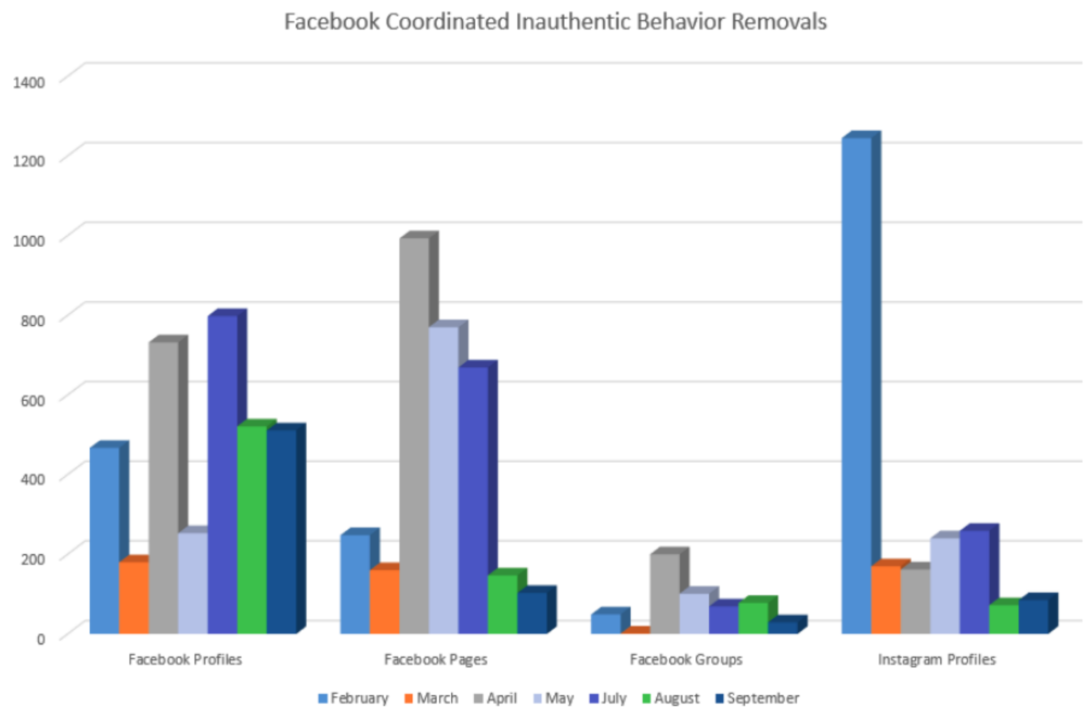


Figure 4.1: Facebook Coordinated Inauthentic Behavior Removals Hutchinson (2020).

The chart depicted above, Figure 4.1, showcases Facebook's removal of accounts, pages and groups that have been found to engage in coordinated inauthentic behavior this year. The chart shows a steady, if not declining amount of removals. Whether or not this decline is due to the effectiveness of the preventative efforts or not is up for discussion. Under a year worth of data might not be enough to safely declare anything. Other factors such as the bad actors employing new and better methods for masquerading their activity and evading detection might very well have an impact on these numbers as well.

4.2.6 Impact of coordinated inauthentic behavior

Social media has become one of the main sources for news the recent years. Close to 47 % of all US adults gets some of their news from social media platforms Newman (2020). These numbers have only inclined in the recent years Nielsen (2017), and is becoming more popular as the younger and more technological generations grows older.

Figure 4.2 shows that close to 52 % of all Facebook users in the US reads news from the platform.

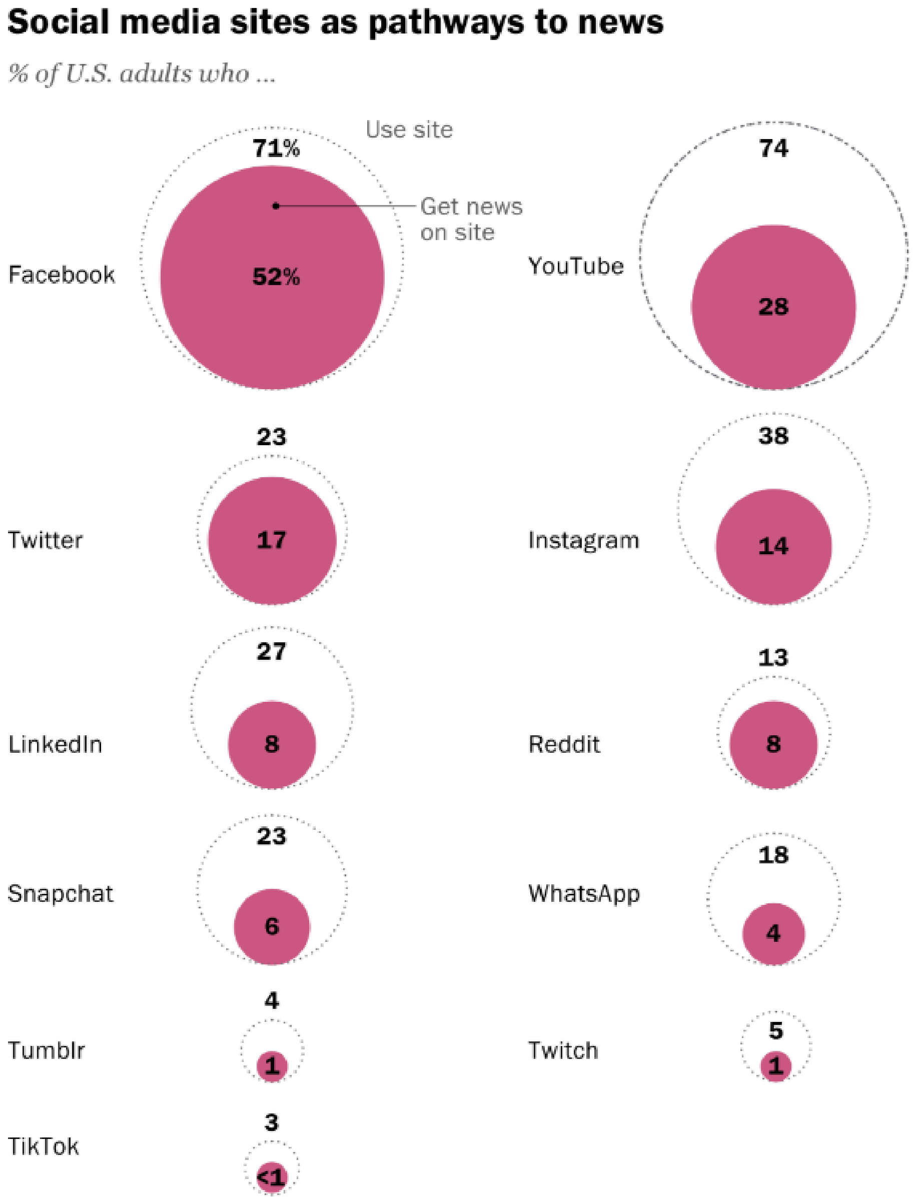


Figure 4.2: News from Social Media OWEN (2019).

These numbers and figures shows the news that the people are aware that they read and get of the platform. However, one can be influenced by other means than just news. Pictures, personal stories, memes, advertisements, and videos can all have a great impact on how a person interprets subjects and situations. People and groups engaging in coordinated inauthentic behavior will often try to influence people by showing them selective interpretations of situations and subjects through the aforementioned mediums. Such as seen in the cases of the IRA with the “Aztlan Warriors” and “Black Elevation”, amongst many of their other pages and groups.

A person’s interactions on social media and the internet as a whole works in the same way that advertisement does FELDWICK (n.d.). The more one is exposed to certain information, the more one is likely to be influenced by it. If a person is only exposed to the positive sides of a subject, that persons perception of that subject will likely be positive even though no research on the subject is conducted and there was no particular feeling towards it beforehand. And even if the person had opinions of the subject beforehand, there is a high probability of this opinion shifting if he or she is heavily exposed to one sided information. Pages, groups, and accounts engaging in coordinated inauthentic behavior utilizes this philosophy in order to influence users of the social media platforms.

People may lead their entire life based on the content they are exposed to on the internet. Who they are voting for, their political views, environmental views, perception of countries and religion, consumer behavior, and ethics are only a handful of subjects that can be influenced by what you see and read on social media. The impact of coordinated inauthentic behavior can therefore have substantial consequences on how people lead their lives up until the point of affecting how an entire country unfolds, such as with the outcome of presidential elections. The papers “Election Manipulation After the Emergence of Social Media”, “Undermining Democracy - The Effect of Manipulation Through CIB in Social Media”, and “Foreign Coordinated Inauthentic Behaviour - A Case Study on the Russian Internet Re-search Agency’s Interference in the 2016 US Election” dwells deeper into how great of an impact manipulation and coordinated inauthentic behavior can have on countries and states.

4.2.7 Future of CIB

The internet is an ever-changing environment, where new technology and practices are constantly expanding and being developed. This makes the future for how coordinated inauthentic behavior is being practiced and detected uncertain. Both the defenders and attackers will have to pay great attention to the environment, and adapt quickly in order to succeed.

The last decade has seen the rise of bots, artificial intelligence, and machine learning, amongst other technologies that have changed the way people operate on the internet. As will be discussed later in “SMME – Technical Solutions, Limitations, and Challenges”, the aforementioned technologies will have a substantial role to play in the evolution of coordinated inauthentic behavior, both on how it is performed as well as how it is fought.

4.3 Chapter conclusion

Although the term Coordinated Inauthentic Behavior only recently became a talked about subject, the basis of the phenomenon has been around since the dawn of time. Facebook was the platform that popularized the term. However, such behavior have always been unwanted on most platforms alike.

Coordinated inauthentic behavior on the internet isn't necessary about spreading false news and narratives. The news and posts shared by people engaging in this type of behavior may very well be authentic. It's falsifying the origin and purpose of the pages, groups, and accounts that defines what constitutes as coordinated inauthentic behavior, at least according to Facebook. Since the phenomenon does not in any way belong to Facebook, there are different interpretations of what can be categorized as coordinated inauthentic behavior. However, the impact of the phenomenon remains the same.

As the number of social media users escalates, the potential impact of coordinated inauthentic behavior increases greatly. If the social media platforms and the users of the internet do not employ and constantly improve their defences and remediation against this type of behavior, the consequences will be severe.

And because of this: Yes, coordinated inauthentic behavior is something we all should be afraid of.

Bibliography

Andrews, E. (n.d.), 'Who invented the internet?', *Histroy* . Accessed 13.09.2020.
URL: https://www.history.com/news/who-invented-the-internet#:~:text=The%20first%20workable%20prototype%20of,communicate%20on%20a%20single%20network

Bachrach, J. (2006), 'U want me 2 kill him?', *Vanity fair* . Accessed 13.09.2020.
URL: https://www.vanityfair.com/news/2005/02/bachrach200502

Chen, A. (2015), 'The agency', *The New York Times* . Accessed 19.09.2020.
URL: https://www.nytimes.com/2015/06/07/magazine/the-agency.html

Cohen, D. (2020), 'Twitter details several networks of accounts removed for coordinated inauthentic behavior', *Adweek* . Accessed 27.09.2020.
URL: https://www.adweek.com/digital/twitter-details-several-networks-of-accounts-removed-for-coordinated-inauthentic-behavior/

Facebook (2018), 'Removing myanmar military officials from facebook', *Facebook* . Accessed 06.09.2020.
URL: https://about.fb.com/news/2018/08/removing-myanmar-officials/

Facebook (n.d.*a*), '20. inauthentic behavior'. Accessed 12.09.2020.
URL: https://m.facebook.com/communitystandards/inauthentic_behavior/

Facebook (n.d.*b*), 'Community standards - all updates'. Accessed 06.09.2020.
URL: https://www.facebook.com/communitystandards/recentupdates/all_updates/

fake news (n.d.), *Dictionary.com* . Accessed 11.10.2020.
URL: https://www.dictionary.com/browse/fake-news

FELDWICK, P. (n.d.), 'How does advertising work?', *Advertising Association* . Accessed 07.11.2020.

URL: https://www.adassoc.org.uk/credos/how-does-advertising-work/#:~:text=They%20propose%20that%20advertising%20works,influence%20purchasing%20behaviour%2C%20often%20unconsciously.&text=Therefore%2C%20the%20way%20an%20ad,you%20have%20for%20the%20brand

Gilbert, B. (2019), 'The 10 most-viewed fake-news stories on facebook in 2019 were just revealed in a new report', *Business Insider* . Accessed 11.10.2020.
URL: https://www.businessinsider.com/most-viewed-fake-news-stories-shared-on-facebook-2019-2019-11?r=US&IR=T#3-aoc-proposed-a-motorcycle-ban-8

Gleicher, N. (2018), 'More information about last week's takedowns', *Facebook* . Accessed 06.09.2020.
URL: https://about.fb.com/news/2018/11/last-weeks-takedowns/

Gleicher, N. (2020), *Twitter* . Accessed 04.10.2020.
URL: https://twitter.com/ngleicher/status/1274823050647580673

Gleicher, N. and Rodriguez, O. (2018), 'Removing additional inauthentic activity from facebook', *Facebook* . Accessed 06.09.2020.
URL: https://about.fb.com/news/2018/10/removing-inauthentic-activity/

Gleicher, N. and Stamos, A. (2018), 'Removing bad actors on facebook', *Facebook* . Accessed 06.09.2020.
URL: https://about.fb.com/news/2018/07/removing-bad-actors-on-facebook/

Greene, C. and Gleicher, N. (2018), 'Taking down more coordinated inauthentic behavior', *Facebook* . Accessed 06.09.2020.
URL: https://about.fb.com/news/2018/08/more-coordinated-inauthentic-behavior/

History of the Web (n.d.), *World Wide Web Foundation* . Accessed 13.09.2020.
URL: https://webfoundation.org/about/vision/history-of-the-web/

Hutchinson, A. (2020), 'Facebook provides update on actions taken against coordinated inauthentic behavior', *SocailMediaToday* . Accessed 31.10.2020.
URL: https://www.socialmediatoday.com/news/facebook-provides-update-on-actions-taken-against-coordinated-inauthentic-b-1/586710/

Inauthentic (n.d.), *Oxford Dictionary* . Accessed 12.09.2020.
URL: https://www.oxfordlearnersdictionaries.com/definition/english/inauthentic?q=inauthentic

Laupp, M. J. (2020), 'Did you know you can make sure there are empty seats at trump's rally?', *TikTok* . Accessed 04.10.2020.
URL: https://www.tiktok.com/@maryjolaupp/video/6837311838640803078

Lorenz, T., Browning, K. and Frenkel, S. (2020), 'Tiktok teens and k-pop stans say they sank trump rally', *The New York Times* . Accessed 04.10.2020.
URL: https://www.nytimes.com/2020/06/21/style/tiktok-trump-rally-tulsa.html

Lyons, K. (2020), ‘K-pop fans and tiktok teens say they reserved tickets for trump’s tulsa rally to leave seats empty’, *The Verge* . Accessed 04.10.2020.
URL: https://www.theverge.com/2020/6/21/21298169/kpop-fans-tiktok-tickets-trump-tulsa-rally-empty-seats

Neuman, S. (2019), ‘Youtube channels suspended for ’coordinated’ influence campaign against hong kong’, *NPR* . Accessed 27.09.2020.
URL: https://www.npr.org/2019/08/23/753626357/youtube-channels-suspended-for-coordinated-influence-campaign-against-hong-kong?t=1601194902014

Newman, N. (2020), ‘Executive summary and key findings of the 2020 report’, *Digital News Report* . Accessed 01.11.2020.
URL: https://www.digitalnewsreport.org/survey/2020/overview-key-findings-2020/

Nielsen, R. K. (2017), ‘Where do people get their news?’, *Oxford University* . Accessed 01.11.2020.
URL: https://medium.com/oxford-university/where-do-people-get-their-news-8e850a0dea03

OWEN, L. H. (2019), ‘More americans than ever are getting news from social media, even as they say social media makes news ”worse”’, *NiemanLab* . Accessed 01.11.2020.
URL: https://www.niemanlab.org/2019/10/more-americans-than-ever-are-getting-news-from-social-media-even-as-they-say-social-media-makes-news-worse/

Press, A. (2019), ‘Fake stories on facebook have an estimated 159 million views this year’, *Market Watch* . Accessed 11.10.2020.
URL: https://www.marketwatch.com/story/facebooks-misinformation-problem-is-actually-getting-worse-group-says-2019-11-06

Schwartz, M. J. (2018), ‘Google suspends youtube accounts, content linked to iran’, *Bank info security* . Accessed 27.09.2020.
URL: https://www.bankinfosecurity.com/google-youtube-accounts-content-linked-to-iran-a-11415

Statt, N. (2020), ‘Tiktok is banning deepfakes to better protect against misinformation’, *The Verge* . Accessed 27.09.2020.
URL: https://www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference

TikTok (2020), ‘Community guidelines’, *TikTok* . Accessed 04.10.2020.
URL: https://www.tiktok.com/community-guidelines?lang=en

Twitter (2020), ‘Community guidelines’, *Twitter* . Accessed 18.10.2020.
URL: https://help.twitter.com/en/rules-and-policies/platform-manipulation

TwitterSafety (2018), *Twitter* . Accessed 27.09.2020.
URL: *https://twitter.com/TwitterSafety/status/1032055161978585088*

Chapter 5

Undermining Democracy – The effect of manipulation through CIB in social media

By Even Hyldmo

5.1 Executive summary

This paper is commissioned to examine the effects of manipulation, including disinformation and propaganda, through coordinated inauthentic behavior (CIB). More specifically the paper discussing whether it is creating an aftermath, where the societies effected are changing and democracies are undermined. The paper is answering the question through two parts, first by analyzing and evaluating the objective and consequence of manipulative content based on examples of accusations where manipulation have been used for political purposes. More specifically, whether the effects of manipulation through online content and CIB are undermining democratic states. The second part is addressing the problem in a theoretical manner by analyzing findings from the literature including an experiment. Through this two-parts approach the examples and empirical evidence may be anchored in theory, providing an answer to the effects of manipulation. Additionally, responsibility and recommendations are discussed to provide options for the management.

The findings draw attention to the fact that while limited geographically, manipulation has been around for centuries (Gleicher, 2018*a*). However, invention of the Internet and social media have paved way for information sharing, where findings indicate that producers of propaganda and manipulative content have found their way exploiting the opportunity to surge online (Stenslie et al., 2019), (Ofori-Attah and Martin, 2011). Both examples of online manipulation and the literature emphasize that democracies are under constant pressure, and under the threat of having their forms of government undermined (Kalsnes, 2019*a*,

p. 14), (Hwang, 2019), (Morgan, 2018), (Vilmer et al., 2018), (Nederhoed, 2019). Findings imply that through CIB, the above-mentioned effects of the almost omnipresent abundance of propaganda and manipulative content are amplified (Anderson and Rainie, 2020).

In conclusion, the effects of manipulative content are amplified by both social media and the rise of CIB, which may constitute a potent enabler to undermine both states transitioning to democracy, enables a government to perpetuate their rule or undermine a foreign state's democratic election. Furthermore, it remains a paradox that the platform initially designed to share information, strengthening one of the principles of democracy - informed voters (Hochhild, 2010), also may account for undermining the principle itself. Additionally, increased knowledge of the effects of manipulation and fake news could mitigate their effects, but may also lead to skepticism in factual news and debates (Kalsnes, 2019*a*). This skepticism may create a gap of mistrust between the voters and the government, which paradoxically could undermine democracy.

Due to the prospects of CIB and the severeness of its effect, several recommendations are proposed in order to take remedial actions, thus enabling the management to address the high impact, high probability risk. These recommendations are summarized as follows:

- Increase awareness of the potential problem among the population (Control/Transfer).
- Increase knowledge, and the usage of source and information evaluation (Control/-Transfer).

The two aforementioned recommendations may be implemented as mandatory throughout an education system. This measure is deemed cost-effective, as it may reduce the probability and maybe its effectiveness. The residual risk by learning about the problem, and information and source evaluation is assessed as medium probability and medium impact.

- Contribute to the removal of inauthentic users, by reporting suspected users to the website owners (Transfer).
- Collect further evidence to investigate the topic further. E.g. using Analysis of Competing Hypothesis.

The first measure may be non-cost-effective, as the increase of manipulative content may be higher than the amount removed. Manually identification of coordinated inauthentic users are assessed as ineffective, thus automatic removal may be necessary and could be the responsibility of platform owners. Although, automatic removal may restrict a public debate, thus challenging free speech and democracy. Additionally, implementing laws restricting creation of inauthentic users may be problematic due to local legislation and local the international access to the social media platform. The residual risk is assessed to stay as high probability and high impact, but may be reduced by the platform owners. The latter measure would provide more evidence to prove the existence of the threat. However, this measure is deemed non-cost-effective, as it would provide knowledge to a handful of people. The risk remains with a high impact and high probability.

During this report, several limitations have been identified. Some of the limitations are related to the literature used, both in terms of using primarily western literature, as well as the lack of state confidential findings with regards to manipulation attempts. Furthermore, the lion's share of the examples represents a weakness, as attributing and correlating the effects to the actual manipulative content or propaganda remains hypothetical. This also includes states correlating effects and actions when attributing a foreign state's attempt to interfere with their elections. This is since a state may have a covert agenda to paint a foreign state black when accusing a foreign state of propaganda or manipulation. This weakness requires further research and investigation to affirm.

Keywords— Democracy, CIB, Manipulation, propaganda

5.2 Introduction

"Comment is free, but facts are sacred"

C.P. Scott, 1921 (Singer and Ashman, 2009)

The well-known citation above, by C. P Scott (Singer and Ashman, 2009) may prove that he was ahead of his time, knowing what was yet to appear with the information age. The citation could be interpreted as the need to protect the facts, as they are scarce, while comments, views and opinions are flourishing in the society. Was Scott ahead of his time?

The act of misleading people has been around for centuries (Gleicher, 2018*b*), but has been geographically restricted (Johnsen, 2020). However, the invention of Internet and its progress during the 90s has paved way for information sharing, where the access to information are greater than ever (Stenslie et al., 2019), (Ofori-Attah and Martin, 2011). Simultaneously, producers of propaganda and manipulative content have found their way exploiting the availability and have seen an opportunity to surge and flourish online (Kelly et al., 2017). As clarified by Ivar Moen in chapter 2, manipulation as a tool for achieving a goal has a long history, and is seen before, during and after elections.

Globally, governments had increasingly used manipulation on social media platforms in 2017 (Kelly et al., 2017). The large-scale use of inauthentic users has been a part of this activity (Kelly et al., 2017). According to a publication from Freedomhouse in 2017, disinformation tactics has shown its importance in 17 other countries between 2016 and 2017 (Kelly et al., 2017), (Kalsnes, 2019*a*). According to a research project at Oxford University, 28 organized social media manipulation organizations existed in 2017 (Bradshaw and Howard, 2017). The disinformation tactics poses a serious threat to the democracy, as the society are unable to distinguish between factual information and manipulated content (Kalsnes, 2019*a*).

This raises several questions; What are these disinformation tactics; what objectives are achievable by using manipulative content during other states' elections; Is it possible to influence the citizens within a state to a degree where it is possible to control the outcome of an election?

This paper is dedicated to research whether social media is contributing to informed choices or limiting the factual debate to a degree where the voters are unable to make an informed choice, thus undermining a principle of democracy (Hochhild, 2010). In order to answer this question, both historical examples and research on the topic are consulted. Furthermore, the paper discusses whether there is a difference when propaganda and manipulative content are spread by coordinated inauthentic users. The key findings in this paper may provide insight in the topic. Additionally, it may contribute to a factual debate, discussing whether practitioners of information security management and the society as a whole are facing a paradigm shift related to how we perceive information channels.

5.3 Methodology

This paper has a IMRaD[1] -structure with minor customization and is based on a systematic literature review.

The theory and literature are the foundation of discussion throughout the paper, including relevant academic peer-reviewed articles, journals, books and other publications. News outlets have also been consulted, as these are deemed necessary in a holistic approach to the research. Note that using information from news outlets, some of the sources might have a covert or overt agenda by accusing another state of using manipulative content and bots.Parts of the literature have been found by using several search engines with Boolean operators and key words, including "Propaganda", "CIB", "Manipulation", "Social media", "bot", "election" and "democracy". By using other academic resources and literature, other conclusions may be reached.

Due to language barriers, the lion's share of the literature regarding this topic is by western authors, which may have affected the findings[2]. In order to provide a more nuanced picture, Russian, Ukrainian and other primary sources have also been applied, where linguist expertise have been consulted.

Providing sufficient attribution and evidence of an election result to manipulation and CIB has proven difficult and is therefore mainly theorized based on literature on the topic. Furthermore, there is a high possibility that dark figures of cases where propaganda, disinformation and manipulation have happened online, which has yet to be documented. Based on historical data, new examples of manipualtio and CIB are likely to occur after the information cut off for this paper as of 5th of November 2020.

As this paper is solely focusing on the effects of manipulation and CIB on democracy, it does not provide information about other societal changes.

In order or grasp the underlying aspects of this paper, please consult a more detailed overview of manipulation, CIB, social media and their effects in previous chapters 2, 3 and 4. This is since it is viewed as a prerequisite to have some prior knowledge of the topics discussed.

The term CIB is mostly used by Facebook (Gleicher, 2018*b*), while the more commonly used term is "bot" (Gorwa, 2017) when referring to coordinated inauthentic users. Please note that the term "bot" has different connotations in different countries (Gorwa, 2017), and therefore the author's connotation may have affected the research. In this paper, CIB is used to describe coordinated, automated non-human social media accounts used to spread used to increase propaganda, disinformation or manipulative content. Propaganda is here widely used about large scale dissemination of content, used to discredit or tribute a person, action or situation, and may be false or correct. Disinformation is here used to describe dissemination of untrue information, with an intent of controlling a narrative. Manipulation is used as an collective term of manipulative and manipulated content, propaganda and

[1] IMRaD is a structure of scientific texts and is an acronym for Introduction, Materials and methods, Results, and Discussion and inclusion. For more information, please see (Springer, 2020)

[2] Findings regarding human behaviour are often sampled in the West, thus making the study affected by an oversample of Western, Educated, Industrialized, Rich and Democratic (WEIRD) societies. For more information, please see (Azar, 2010)

disinformation, where a message is edited to fit a fictional reality or is intended to edit the reality of its recipients.

5.4 Manipulation for political purposes

> "In the future, you could through manipulation be able to achieve an objective, where you previously relied on weapons [...]" (Sandnes, 2020) (translated by Even Hyldmo)
>
> Chief Norwegian Cyber Defense Force Inge Kampenes, 2020

The citation above may predict what we have coming. However, according to Freedomhouse this is already happening (Kelly et al., 2017), as 30 countries including Turkey, Venezuela and the Philippines are accused of using "armies of opinion shapers" in social media to influence their respective population to counter government critics, spread the government views and control the online discussion (Kelly et al., 2017). The objective of these incumbent governments is to perpetuate their rule.

Instead of using censorship, which may be perceived as a direct intervention against the democracy and freedom of speech (Meserve and Pemstein, 2017), the use of CIB may be more effective, as it may be perceived as a more subtle approach from the government side (Kelly et al., 2017).

According to the Russian constitution (Kalsnes, 2019*a*), Russia is a democratic federation. Simultaneously, the Russian government are accused of using manipulation and disinformation in order to assert their power and preserve the current government (Kelly et al., 2017). In addition, the Russian national security strategy 2020 covers how Russia will tackle a "increased global information fight" by cultivating a Russian view (Kremlin, 2019). This could be viewed as a proclamation of involvement in spreading Russian propaganda. This view is supported by numerous examples of Russia's information warfare against Western governments (Galeotti, 2014), (Aro, 2016), (Kioski, n.d.). These examples include the Russian government's attempts of using bots and fake news to interfere and influence the outcome of elections in Western governments. A Russian controlled Facebook-page named "Heart of Texas" (Bertrand, 2017) became the "world's most famous manipulator of social media", by exploiting divergent emotions in American politics and culture (Shane and Mazzetti, 2018). Russia have understood that by mobilizing Americans, they could gain political power (Shane and Mazzetti, 2018). The Russian objective is to confuse the user of social media and to promote a favorable perception of Russia and its foreign affairs, through the use of manipulative content, disinformation or propaganda (Zakem et al., 2015).

As an example, Russian fabricated news flourished during the United States president election in 2016 (Kelly et al., 2017). The number of fake social media accounts used by Russia were almost as many as the American voters during the election (Shane and Mazzetti, 2018) The sign of successfully accomplishing the objective to influence the election may be the banner of the previous president of the United States Barack Obama with the imprinted text "Goodbye Murderer" (Shane and Mazzetti, 2018). The earliest promoter of the banner was a Russian operated Twitter account named @LeroyLovesUSA (Shane and Mazzetti, 2018). However, attributing the victory of President Donald Trump in 2016 to a coordinated manipulation campaign by Russia may be proved and disproved. As seen in isolation, the campaign may have worked or not worked, however the assessed manipulation was not done in a vacuum, but rather one operational line in a larger coordinated operation consisting of multiple lines against multiple fronts (Shane and Mazzetti, 2018).

While there are accusations against Russia, it is important to notice that Russia has been regarded as a geopolitical threat throughout the history by the West (Chapman, 2015). There might be a large share of dark figures with regards to Western manipulation and propaganda, but it may be undocumented.

Facebook suggested in 2020 that their automated systems of detecting CIB are not able to count more than 10% of the enforcement actions towards the automated accounts (Facebook, 2020*e*). According to professor Igor Panarin, the West have through information propagation ignited the Arab Spring and Euromaidan (Panarin et al., 2015).

Although strategies and accusations may find its way to the news and research, it is deemed necessary to look at the levels of CIB and sharing of propaganda and manipulative content online in order to anchor the theory in practice.

During a collection of automated content, there were registered a higher sharing of content on Polish Twitter, during a period where there were no political events, than during the elections in France, Germany, and United Kingdom (Gorwa, 2017), (Gallacher et al., 2017). However, there might be sleeping social media accounts, meaning that they are not in use until a specified moment in time.

According to a research project conducted by Bradshaw and Howard, 28 countries were identified having manipulated social media content, where 18 of these were having fake automated accounts (Bradshaw and Howard, 2017). The 28 countries are the following: "Argentina, Azerbaijan, Australia, Bahrain, Brazil, China, the Czech Republic, Ecuador, Germany, India, Iran, Israel, Mexico, North Korea, the Philippines, Poland, Russia, Saudi Arabia, Serbia, South Korea, Syria, Taiwan, Turkey, Ukraine, the United Kingdom, the United States, Venezuela and Vietnam" (Bradshaw and Howard, 2017).

Collaborating findings on the use of CIB, manipulative content and propaganda from the last decade have resulted in more than the 28 countries identified by Bradshaw and Howard (2017). as depicted in Table 5.1 below:

Table 5.1: list of discovered states and groups using CIB, manipulation, disinformation and propaganda the last decade 2010-2020, the targets and objectives are given by the sources or assessed based on the provided information

Originator	How	Assessed Objective/Targe
Afghanistan	CIB	Domestic: Farsi- and Dari speaking population (Facebook, 2015)
Argentina	CIB, Propaganda	Domestic: Maintain power (Rueda, 2012)
Australia	CIB	Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Facebook, 2020*d*)
Azerbaijan	CIB, Manipulation	Domestic: Remain in control (Pearce and Kendzior, 2012), (Geybulla, 2016), (Facebook, 2015)
Bahrain	Disinformation	Unknown (Kelly et al., 2017)
Brazil	CIB	Domestic: Political influence (Facebook, 2020*d*)
Canada	CIB	Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Facebook, 2020*b*); El Salvador, Argentina, Uruguay, Venezuela, Ecuador, and Chile (Facebook, 2020*d*)

Originator	How	Assessed Objective/Targe
China	Disinformation, Propaganda	Domestic: Maintain power (Kalsnes, 2019*a*), (G. King and Roberts, 2017) Foreign public debate or state structures: Taiwan (Benedictus, 2015)
Czech Republic	Distraction	Unknown (Bradshaw and Howard, 2017)
Egypt	CIB, Manipulation, Propaganda	Domestic: Protests (Revolution of 25 January 2011) (El-Khalili, 2013) Foreign public debate or state structures: Somalia, Yemen, Saudi Arabia, Sudan, Yemen, Tunisia, Iran, Turkey, Lebanon, Qatar, Libya, Comoros, Syria, Jordan, Morocco (Gleicher, 2019*f*), (Gleicher, 2019*e*); Arabic speaking public, Egypt and Gulf region, Middle East – North Africa (Stenslie et al., 2019)
Equador	CIB, Manipulation, Propaganda	Domestic: Opposition: Equador's president (Morla, 2015); Maintain power (Rueda, 2012) Foreign public debate or state structures: El Salvador, Argentina, Uruguay, Venezuela, Ecuador and Chile (Facebook, 2020*d*)
Finland	CIB	Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Facebook, 2020*d*)
France (Unknown personas)	CIB	Domestic: French population (Facebook, 2020*e*) Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Facebook, 2020*b*)
Georgia	CIB, Manipulation	Domestic: Maintain power (Gleicher, 2019*a*), (Facebook, 2020*a*), (Facebook, 2015); Political parties seeking political influence (Facebook, 2020*a*), (Facebook, 2015)
Germany (Unknown personas)	CIB, Disinformation, Propaganda	Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Bradshaw and Howard, 2017), (Facebook, 2020*b*)
Ghana	CIB, Manipulation	Foreign public debate or state structures: US (Facebook, 2020*e*), (Facebook, 2020*c*)
Great Britain	Disinformation	Domestic: Governmental political agenda (Kalsnes, 2019*a*, p. 102)

Originator	How	Assessed Objective/Targe
Honduras	CIB, Manipulation	Domestic: Maintain power (Gleicher, 2019*d*)
Hong Kong	CIB	Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Facebook, 2020*d*)
India	CIB, Manipulation, Propaganda	Domestic: Indian election 2014 (Bharatiya Janata party (BJP) and the Congress party) (Shearlaw, 2015) Foreign public debate or state structures: Gulf region, US, UK, Canada (Facebook, 2020*c*)
Indonesia	CIB, Propaganda	Domestic: Remain in control, opposition gaining political influence (Gleicher, 2019*f*); Saracen-linked group manipulates public debate through CIB (Gleicher, 2019*h*) Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Facebook, 2020*b*)
Iran	CIB, Manipulation, Propaganda	Foreign public debate or state structures: West (English-speaking audience), US (Trending, 2016), (Facebook, 2020*c*); Algeria, Afghanistan, Bangladesh, Bosnia, Egypt, Ghana, Israel, Libya, Mauritania, Morocco, Nigeria, Senegal, Sierra Leone, Somalia, Sudan, Tanzania, Tunisia, the US, UK and Zimbabwe (Gleicher, 2019*d*), (Gleicher, 2019*b*), (Facebook, 2015), (Facebook, 2015) Elections: Isareli election 2019 (Rubinstein, 2019)
Iraq	CIB, Disinformation	Domestic: Kurdistan seeking political influence (Facebook, 2020*f*)
Iraq	CIB	Domestic: Domestic audience (Facebook, 2020*d*)
Islamic State	CIB, Propaganda	Domestic: Maintain power (Alam, 2019), (Berger and Morgan, 2015)
Israel	Propaganda	Foreign public debate or state structures: Palestine; Perception of Israel (Benedictus, 2015)

Originator	How	Assessed Objective/Targe
Italy (Different groups)	Disinformation, Propaganda	Elections: European election in Italy 2019, Italian General election 2018 (Giglietto et al., 2019)
Maurtania	CIB, Manipulation	Domestic: Maintain power (Facebook, 2020*a*)
Mexico	CIB, Propaganda	Domestic (Election): Mexican president election (Vega, 2012) Foreign public debate or state structures: US (Facebook, 2015)
Muslim Brotherhood (Egypt, Turkey, Morocco)	CIB	Foreign public debate or state structures: Egypt, Libya, Tunisia, Yemen, Somalia, and Saudi Arabia (Facebook, 2015)
Myanmar	CIB, Manipulation	Domestic: Maintain power (Kelly et al., 2017), (Facebook, 2020*c*), (Facebook, 2020*a*), (Facebook, 2015)
New Zealand	CIB	Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Facebook, 2020*b*)
Nigeria	Propaganda	Domestic: Domestic audience by Islamic Movement (Facebook, 2015) Foreign public debate or state structures: US (Facebook, 2020*e*), (Facebook, 2020*c*) Elections: President election 2015 (Onafuwa, 2017) President election 2019 (Team, 2019)
North Korea	Manipulation, Propagand	Foreign public debate or state structures: South Korean (Seong-min, 2013)
Pakistan	CIB	Domestic: Domestic audience (Facebook, 2020*b*) Foreign public debate or state structures: India (Facebook, 2020*b*)
Poland	CIB, Manipulation	Domestic: Maintain power (Gorwa, 2017)
Romania	CIB	Foreign public debate or state structures: US (Facebook, 2020*d*)

Originator	How	Assessed Objective/Targe
Russia	CIB, Disinformation, Manipulation, Propaganda	Domestic: Russian opposition (Kalsnes, 2019*a*), (О. Гармажапова , 2014); Pro-Russian content (Connell and Vogler, 2017); Foreign public debate or state structures: Ukraine (Connell and Vogler, 2017) (Gleicher, 2019*d*), (Jaitner, 2015), (Kioski, n.d.), (Facebook, 2020*c*); North Africa, Madagascar, the Central African Republic, Mozambique, Democratic Republic of the Congo, Côte d'Ivoire, Cameroon, Libya (Galeotti, 2014), (Gleicher, 2019*g*); Undermine Finnish government (Kioski, n.d.) (Standish, 2017), (Aro, 2016); Georgia (Standish, 2017); Algeria, Egypt; (Facebook, 2020*b*); Western States, English speaking population, Germany, Spain, France, Hungary, Serbia, Georgia, Indonesia, Iran, NATO (Aro, 2016), (Kioski, n.d.), (Facebook, 2020*a*), (Gleicher, 2019*c*); Estonia, Latvia, Lithuania (Standish, 2017); Italy (Patalakh, 2020); Poland (Gorwa, 2017); Thailand (Gleicher, 2019*d*) Elections: US Election 2016 (Bogen, 2018) (Kalsnes, 2019*a*); French election 2017 (Ferrara, 2017*a*) Ukranian election 2019 (Connell and Vogler, 2017), (Gleicher, 2019*d*)
Saudi Arabia	CIB, Manipulation	
Serbia	CIB, Disinformation, Manipulation	Domestic: Maintain power (Rujevic, 2017)
South Korea	Disinformation	Domestic: Remain in control (Sang-Hun, 2013) (KH디지털2, 2013) Foreign public debate or state structures: North Korea (Sang-Hun, 2013)
Switzerland	CIB	Foreign public debate or state structures: Iraq (Facebook, 2020*d*)
Syria	CIB, Propaganda	Domestic: Maintain power (York, 2011), (Gorwa, 2017); Undermine regime (opposition) (York, 2011)

Originator	How	Assessed Objective/Targe
Taiwan	CIB	Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Facebook, 2020*d*)
Thailand	CIB, Manipulation	Domestic: Thailand (Gleicher, 2019*d*) Foreign public debate or state structures: US, China, Hongkong protests (Gleicher, 2019*d*)
The Philiphines	CIB, Propaganda	Domestic: Maintain power (Kelly et al., 2017) (Williams, 2017)
The Democratic Republic of Congo	CIB, Propaganda	Domestic: Political influence by Force des Patriotes (Facebook, 2020*d*)
Tunisia	CIB, Disinformation	Foreign public debate or state structures: Francophone countries in Sub-Saharan Africa (Facebook, 2020*f*)
Turkey	Disinformation	Domestic: Maintain power (Kelly et al., 2017), (Benedictus, 2015)
UAE	CIB, Manipulation, Propaganda	Foreign public debate or state structures: Qatar, Turkey, Iran, Sudan Libya, Comoros, Lebanon, Syria, Jordan, Morocco, Activity in Yemen, Muslim Brotherhood (Gleicher, 2019*f*), (Gleicher, 2019*e*)
Ukraine	CIB, Disinformation, Manipulation	Domestic: Political influence (Facebook, 2015), (Facebook, 2020*d*) Foreign public debate or state structures: Russia (Benedictus, 2015)
United Kingdom	CIB, Manipulation	Foreign public debate or state structures: English and Chinese speaking audiences, and Vietnam (Benedictus, 2015), (Facebook, 2020*d*)
Unknown	Manipulation	United Kingdom (Gorwa, 2017)
Unknown	Disinformation	Nigerian government (Assay, 2019, p. 226)
Unknown (Gleicher, 2019*f*)	CIB, Disinformation	Undermine Qatar (Cherkaoui, 2018),

Originator	How	Assessed Objective/Targe
US	CIB, Disinformation, Manipulation	Domestic: Domestic audience (Far right propaganda and anti-immigration) a head of US election 2020 (Facebook, 2020*a*), (Facebook, 2020*d*); Iraqi war (Kalsnes, 2019*a*, p. 101) Foreign public debate or state structures: Iraqi war (Kalsnes, 2019*a*, p. 101); Vietnam (Gleicher, 2019*a*); Kenya and Botswana (Facebook, 2015); Venezuela, Mexico, Bolivia (Facebook, 2020*b*)
Venezuela	CIB, Disinformation, Manipulation	Domestic: Maintain power (Kelly et al., 2017), (News, 2015), (Rueda, 2012) Foreign public debate or state structures: US (Facebook, 2015)
Vietnam	CIB, Disinformation, Manipulation, Propaganda	Domestic: Maintain power (Pham, 2013), (Staff, 2017) Foreign public debate or state structures: US., Vietnam, Spanish, and Chinese speaking audience (Gleicher, 2019*a*)
Yemen	CIB	Domestic: Unknown domestic target (Facebook, 2020*d*)

Based on the information in Table 5.1 above it is possible to say that the use of CIB, manipulation, propaganda and disinformation is widely spread on a global scale. However, Scandinavia and several countries in East-Asia, Africa and some in Latin-America and Europe are missing.

Furthermore, only a few have been related to an election of a foreign nation, while most of the cases are having a domestic audience in order to perpetuate power or to challenge the current government. The sources have either not been able to attribute the cases to an election, there is too many uncovered cases, or elections are not the main goal to conduct such behavior. A deduction from Table 1 is that the states affecting foreign states' population is often in proximity of the affected states' political and military sphere, in addition to US. Furthermore, most of the commonly viewed authoritarian regimes represent an overweight of states focusing on domestic matters when using social media. Additionally, the examples of elections interfered with are almost all Western, without Ukraine.

Simultaneously, Apuke holds that the threat towards democracy is real, as it can be used to create and fake political support and manipulate the public opinion (Apuke, 2018). However, attributing a campaign to a political victory or loss is difficult. Maybe the closest we can get is examples such as the 2012 presidential election in Mexico, where it might be concluded that the use of bots were a part of the campaign, as the many of the accounts ceased sharing once it was identified (Rueda, 2012).

As introduced, manipulation, disinformation and propaganda has been around for centuries. However, when introducing automated accounts, the ecosystem on social media has changed, where manipulation of the reality has become more effective and more widespread through the use of CIB (Ferrara, 2017*b*). The bots may amplify small voices, resulting in more attention around a relatively unknown phenomenon (Bradshaw and Howard, 2017). Additionally, Milan (2015) argues that the algorithms of social media platforms are customizing the content for the individual user, in such way that the more time spent and actions done within a specific content, the more similar content will show up in your social media account. According to Nizzoli et al. (2020), disinformation is effectively spread through coordinated behaviour online. As Horne, Nørregaard and Adal find in their study of content sharing in mainstream media:"news is homogeneous within communities and diverse in be-tween, creating different spirals of sameness" (Aro, 2016). What Horne, Nørregaard and Adal describe, may be perceived as an echo chamber. With the entrance of CIB, this echo chamber may be amplified and taken to new heights and polarize a population further. This is due to the social media algorithms (Muhammad, 2020), (Yuan et al., 2018), when in combination with CIB allows the sharing a larger amount of manipulative content and propaganda.

5.5 Manipulation of election experiment

> "[...] some of the most potent threats democratic threats in our time: fake news, disinformation and manipulation of elections"(Kalsnes, 2019*b*) (translated by Even Hyldmo)

As mentioned, identifying a causality between manipulation, disinformation, propaganda, CIB and election results may prove difficult. To identify whether there are a causality, the Norwegian state owned TV-channel "NRK' conducted an experiment in 2019 (Nederhoed, 2019). They researched how social media, propaganda, disinformation and manipulation may influence the outcome of a high school election (Nederhoed, 2019). The experiment was conducted over a period of six months, where the objective was to manipulate high school students at Lillestrøm to vote for "Senterpartiet", a Norwegian political party, at a local high school election (Nederhoed, 2019). The methods comprised propaganda, disinformation, manipulation through a meme account, fake "valgomat"[3], and an artificial account representing a human (Nederhoed, 2019). The results showed that "Senterpartiet" only increased by 1,1% in support and were still among the lowest supported political parties (Nederhoed, 2019). Even though the numbers were low, it was still an increase of 50% since the last school election (Nederhoed, 2019). Whether the increase was due to the experiment or not is hard to say. Kalsnes (2019*b*) criticized the experiment on several points; firstly, there were no control group and too many variables to evaluate the effect; secondly, manipulation through social media is well documented, thus the research did not contribute too anything new; thirdly, the timing were right in front of a real municipality election in Norway . The first argument by Kalsnes is important, as there were no control groups just as in a real election. Furthermore, there are also too many variables in an election to correlate a manipulation campaign to a victory. The third argument is also relevant, as the use of social media, manipulation, propaganda and disinformation and the "valgomat" may have been used by others outside the high school. In such, the experiment may have affected outsiders and thus, real political elections.

Looking at the results for the municipality election in Lillestrøm in 2019, "Senterpartiet" were the party that had the highest increased support with an increase of 7,3% (NRK, 2019). Whether the increase was a result of the previous six-months-experiment or not may stay a mystery for eternity. It might be as Ferrara (2017*a*) points out, that perceived anomalous accounts and usage patterns may be ignored by a society. It might be that a successful campaign must rely on cultural and behavioral knowledge. It was not mentioned any cultural or behavioral mapping of Lillestrøm high school stu-

[3] Valgomat is a software or application where a user is displayed a range of political topics, upon which the user is asked to select the topics that is important for the user. After several choices, a political party with corresponding views on political matters are shown to the user in order to make the user more confident, or reduce insecurity, when voting for a political party (Burns, 1999).

dents before conducting the experiment. In conclusion, the experiment did not show a real causality between manipulation and inauthentic behaviour and the election results.

5.6 Challenging democracy?

> "[...] I made a comment that I thought the idea misinformation on Facebook changed the outcome of the election was a crazy idea. Calling that crazy was dismissive and I regret it" (Zuckerberg, 2017)
>
> CEO of Facebook, Mark Zuckerberg, after the 2016 US president election

Even though it might be hard to identify a correlation between manipulation and election results, O'Carroll (2017) suggests that there is a possibility to hire bots and agents to conduct disinformation and manipulation campaigns to influence elections. Where there is a market - there are providers, that have been seen promoting and lobbying for governments, according to O'Carroll (2017). Hill (2017) holds that academics have warned how bots spreading political propaganda are threatening democracy.

Professor Hochhild (2010) points out that an essential part of the democracy is that the electors are making an informed choice. Therefore, one may argue that it is a prerequisite that the voters are informed in such way that their choice is based on the knowledge of the potential consequences of their choice. Given the increased information sharing through internet, it is natural to assume that the voters are more informed than ever. As Facebooks CEO Mark Zuckerberg (Alcantara et al., 2017) points out that internet and Facebook may ensure a good place for democracy, and that Facebook are giving people a voice, strengthening the democracy through participation (Zuckerberg, 2017). Hill (2017) argue that social media are not the ones that creates the content but is a tool to launch the propaganda.

In 2013 Kramer, Guillory and Hancock (Kramer et al., 2014) conducted a study, showing that without direct interaction between subjects, transfer of thoughts, ideas and points of view, through Facebook happened. With this in mind, it is possible for an actor to influence citizens and other nations to behave in a favorable manner for the actor. Although, a CIB creator stated that the use of these tactics cannot directly influence public opinion but the tactics include discrete persuation (Gorwa, 2017). An indirect approach may be harder to attribute, thus adding to the problem of all variables during an election. Both humans and inauthentic users are a part of social media platforms, which also functions as tools for social control (Bradshaw and Howard, 2017). As asserted by Bogen (Bogen, 2018), there are numerous of attempts, and to some degree successful attempts, by the Russian Federation to influence democratic elections in former Soviet states, including satellite states, and other nations. To do so, one would need barely a 100 people to steer and manipulate voters (Shane and Mazzetti, 2018). However, sabotaging a real election would require more resources and persons (Shane and Mazzetti, 2018).

(Freedom House, 2019) argues that governments have globally increased information manipulation on social media in 2017. The increased use of manipulation limits the voters' ability to make informed choices as they receive disinformation and not factual news through their primary source of information and news, social media (Kelly et al., 2017). This argument is seemingly in contrast with Zuckerberg's statement and an essential element of democracy as provided by Professor Hochschild (Hochhild, 2010).

But not only have CIB and manipulation been used to affect elections, it might be used to gather people to protest or even revolutionary purposes to overthrow governments (Khomko, 2015), (Brown et al., 2012), (York, 2011). It is not impossible that the use of manipulation tactics, unrelated to an election, may undermine a democratic state. This is since a revolution will throw out a potentially elected leader, or a democratic state turning autocratic may continue on its road without interference. The Arab Spring in 2011 represent an example of the former, as several groups in several states stood up against their governments (Knox, 2018). However, e.g. in Tunisia the president was deposed, and democratic election introduced (Knox, 2018).Furthermore, a state violating international laws or human rights could also avoid international pressure to stop their wrongdoings by using these tactics to drown out the topic (Brown et al., 2012), (York, 2011), as partially seen in Syria (York, 2011)

However, using manipulation domestically could also go wrong. According to Connell and Vogler (2017) the use of fraud may result in a counterblow, as the 2011 Russian Parliamentary election showed when evidence of fraud was identified, anti-government and anti-putin protests surged on social media. Kremlin allegedly perceived internet as "direct threat to government stability (Connell and Vogler, 2017).

The Norwegian Police Security Service (PST) have annually since 2016 included both propaganda and disinformation in social media as a threat in their threat assessment (Hugubakken, 2016-2020). If PST warns the population and present measures, does the population comply? According to Eurobarometer's survey, 85% of the participants stated that fake news partly or wholly is a threat to the democracy in Europe, while only 12% stated that it is not (Kalsnes, 2019*a*, p. 54). Newmann et al. (2020) conducted a research on whether the participants could trust the news most of the time, and whether they were concerned of what was real and fake on internet. Their findings showed that the country with most respondents that agreed that they could trust the news were Finland (56%), while the least agreeing participants was in South Korea (21%) (Newmann et al., 2020). 45% of the Norwegian respondents agreed to trusting the news most of the time (Newmann et al., 2020). At the same time, Brazilian respondents were the ones that was most concerned about fake and real news (84%), while the Netherlands was the least worried (32%), while Norwegian respondents totaled 42% (Newmann et al., 2020).

It remains unclear if the threat assessment by PST have affected to the Norwegian percentages. But, if 40% of a population does not trust the news, and is worried about disinformation, it could potentially lead do distrust between the government and its population. If manipulation online remains unaddressed, antidemocratic regimes may erode the trust in media and the government (Kelly et al., 2017), which in turn could undermine the current governmental form. The consequences could be that the citizens demand that the government is deposed, and thus indirectly undermines the democracy. Therefore, it could be argued that the abundance of manipulation, disinformation and inauthentic users 'sharing on social media may indirectly interfere with a democracy, by affecting the trust towards real political agendas. As Tufekci puts it: "When it's impossible to distinguish facts from fraud, actual facts lose their power" (Tufekci, 2019).

5.7 Responsibility

> "We'll keep working to ensure the integrity of free and fair elections around the world, and to ensure our community is a platform for all ideas and force for good in democracy" (Zuckerberg, 2017) CEO of Facebook, Mark Zuckerberg.

According to Jed Willard, director of the Franklin Delano Roosevelt Center for Global Engagement at Harvard, the population may prove important when responding to manipulation online (Standish, 2017). By having a positive narrive may be the best way to respond, while correcting false information may be less effective (Standish, 2017). When conducting a survey among the population in Norway, 89% responded that traditional media, such as news papers, radio and television have a large or really large responsibility when it comes to responsibility to hinder spread of fake news. (Kalsnes, 2019*a*, p. 53) On the other hand, 82% responded that technology companies like Facebook, Twitter and Google have a responsibility, while 78% responded that government and politicians have the responsibility, while 60% states that they themselves have a responsibility to hinder the distribution of fake news (Kalsnes, 2019*a*, p. 53).

According to Hill (2017), Woolley and Howard stated that social media must address the threat of computational propaganda and improve, if democracy is to survive social media. As the citation by Zuckerberg indicate, Facebook have already addressed the problem, by working to remove CIB and manipulation attempts (Gleicher, 2018*b*), to defend against states attempting to undermine and interfere with elections (Zuckerberg, 2017).

In somewhat contrast, the professors Rosenblatt and Dhar (2020), shared their opinions in Business insider (Rosenblatt and Dhar, 2020), where they stated that the lawmakers must take responsibility. A tech giant owning a social media platform may not directly have a societal responsibility. If the owner is ignorant, ignoring or satisfied with their platform being occupied by CIB, they may let it pass. On the other hand, if the public learns that a large part of the content on the platform is manipulated, it could potentially lead to loss of users. Lawmakers however, may put juridical restrictions (Rosenblatt and Dhar, 2020)

One could also direct the responsibility at a states defensive forces, as the word describes – it is defending a nation. Burns (1999) stated in 1999, that one of the tactics within information warfare is distorting the reality of an opponent or the public through propaganda, disinformation and demoralization, which can be done using technologies allowing such manipulation of information. He explained that the type of warfare are done by a state's military power, using civilian infrastructure in order to apply the influence in a timely manner (Burns, 1999). So it might seem natural that when the military is the actor, it should also be the defender. Burns (1999) contemplation indicate that a state's armed forces are using civilian infrastructure, such as the internet and social media to apply this form of warfare in the digital domain. Attributing CIB, manipulation and propaganda as part of information warfare, may seem like going down the rabbit hole. This is since the usage of civilian technologies such as the internet and social media platforms to spread disinformation and manipulatitive content are not directly linked to a state's armed forces. Thus, it may prove difficult to direct the responsibility to a States defense.

If the democracy is challenged by the use of manipulation of information and through CIB, how do we counter it? According to Freedomhouse (Kelly et al., 2017), time, resources and creativity are needed to counter this type of disinformation, as the countering should embrace the media and internet freedom. Furthermore, they argue that the responsibility should be divided between eductional institutions, governments of democratic countries, and technological companies (Kelly et al., 2017). This allows citizens to be educated on how to evaluate and investigate the sources and information presented, and report it (Kelly et al., 2017). Governments must regulate political propaganda, while tech companies must proactively remove of bots and fake accounts used to spread disinformation (Kelly et al., 2017).

It can be argued that the responsibility must be taken by every citizen. Accepting the risk of having democracy undermined is deemed non-expedient. Avoiding the risk by banning social media is deemed unfeasible, as the platforms would exist due to their economical gains, and could lead to dissatisfaction and protests (Bogost, 2019).

A combination of implementations of regulations, educational institutions teach students on how to assess presented information, and CIB are proactively removed by social media platforms. Thus, every citizen may bear the responsibility of using their knowledge to assess online information and participate in the political debate with scepticism.

5.8 Conclusion

This paper has drawn attention to whether manipulation through CIB are undermining democracy, by looking at examples of accused use of abovementioned tactics. There are, however, only a few examples of elections influenced by the use of CIB the last decade, manipulation, propaganda and disinformation (Giglietto et al., 2019), (Rubinstein, 2019), (Onafuwa, 2017), (Team, 2019), (Bogen, 2018), (Kalsnes, 2019*a*), (Ferrara, 2017*a*), (Connell and Vogler, 2017), (Gleicher, 2019*d*). These elections are mainly in Westernized countries, including Ukraine. However, there are potentially a large share that remains undocumented (Facebook, 2020*e*).

The findings indicate that CIB is contributing as an amplifier of the dissemination of manipulative content, and with or without it, the social media platforms are tools that can be used to undermine democracy (Kalsnes, 2019*a*, p. 14), (Hwang, 2019), (Morgan, 2018), (Vilmer et al., 2018), (Nederhoed, 2019), (Anderson and Rainie, 2020).

The literature suggests that CIB, manipulation, propaganda and disinformation not only affects elections of foreign states, but also lets a government to perpetuate their rule (Gleicher, 2019*a*), (Facebook, 2020*a*), (Facebook, 2015), or control the political discussion instead of using censorship (Kelly et al., 2017).

Additionally, the social media platforms are giving people the opportunity to share and participate in the democracy (Standish, 2017). However, due to the abundance of CIB, propaganda and disinformation on these platforms, the voters are not adequately informed to make an informed choice (Johnsen, 2020). Thus, the platform initially intended to strengthen democracy may simultaneously undermine it, as the voters may not conduct an informed choice based on factual information. It is possible to argue that the democracy also is threatened by increased distrust between the government and its population. If people are unresponsive of factual information, as it becomes infrequent among disinformation, people will not be able to conduct and informed choice.

When it comes to measures to accept, control avoid, or transfer the risk constitute by this phenomenon, a combination of the social media platform owners, governments, educational institutions and the citizens could be responsible (Kelly et al., 2017). Accepting the risk would be an ineffective solution. Avoiding the risk would be to shut down social media, an option deemed unfeasible. While transffering and controlling the risk are assessed more cost-effective. In order to control the effect of manipulation, disinformation, propaganda and CIB, the platform owners could proactively remove any attempts of abovementioned, while citizens could report suspicious activity (Kelly et al., 2017), (Gleicher, 2019*a*). Furthermore, governments exposed to CIB, manipulation, propaganda and disinformation should regulate the use of propaganda, and provide factual news. Simultaneously, education of source and information evaluation could be implemented in educational institutions in order to prepare future generations for this increased challenge.

This report has several limitations. Some are related to the literature used, while the main limitation is the absence of examples with a causality between an undermined democracy and manipulation. To be able to present such a causality, additional research with defined variables, control groups and in a controlled environment could be expedient. A states confidential information could potentially provide real evidence of a causality between outcome of elections and CIB and manipulation.

Bibliography

Alam, H. (2019), 'Twitter struggling to shut down bot and impersonation accounts created by isis facebook twitter flipboard email'. Accessed 26.09.2020.
URL: *https://www.npr.org/2019/11/01/775509366/twitter-struggling-to-shut-down-bot-and-impersonation-accounts-created-by-isis?t=1605806153781*

Alcantara, A., Malter, J. and Quart, J. (2017), 'Zuckerberg: Internet for the world is good for democracy'. Accessed 19.10.2020.
URL: *http://money.cnn.com/video/technology/business/2017/06/22/mark-zuckerberg-facebook-democracy-internet-connection.cnnmoney:CNNTech*

Anderson, J. and Rainie, L. (2020), 'Many tech experts say digital disruption will hurt democracy'. Accessed 19.09.2020.
URL: *https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2020/02/PI_2020.02.21_future-democracy_REPORT.pdf*

Apuke, O. D. (2018), 'The role of social media and computational propaganda in political campaign communication'. Accessed 15.10.2020.
URL: *https://www.researchgate.net/publication/328505241_the_role_of_social_media_and_computational_propaganda_in_political_campaign_communication*

Aro, J. (2016), 'The cyberspace war: propaganda and trolling as warfare tools'. Accessed 01.11.2020.
URL: *https://link.springer.com/content/pdf/10.1007/s12290-016-0395-5.pdf*

Assay, B. E. (2019), 'Social media and the challenges of curtailing the spread of fake news in nigeria', *Handbook of Research on Deception, Fake News, and Misinformation Online* **1**(1).

Azar, B. (2010), 'Are your findings 'weird'?'.
URL: *https://www.apa.org/monitor/2010/05/weird*

О. Гармажапова (2014), 'Где живут тролли в РФ: как работают интернет-провокаторы в Санкт-Петербурге и кто ими заправляет '.
URL: *https://news.finance.ua/ru/news/-/320589/gde-zhivut-trolli-v-rf-kak-rabotayut-internet-provokatory-v-sankt-peterburge-i-kto-imi-zapravlyaet*

Benedictus, L. (2015), 'Invasion of the troll armies: from russian trump supporters to turkish state stooges'. Accessed 01.11.2020.
URL: *https://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian*

Berger, J. and Morgan, J. (2015), 'The isis twitter census'. Accessed 20.09.2020.
URL: *https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf*

Bertrand, N. (2017), 'Texas secession movement: Russia-linked facebook group asked us to participate in anti-clinton rallies'. Accessed 18.10.2020.
URL: *https://www.businessinsider.com/russia-facebook-group-ads-texas-secession-secede-trump-clinton-2017-9?r=US&IR=T*

Bogen, Ø. (2018), *Russlands Hemmelige Krig mot Vesten*, Kagge Forlag AS, Nørhaven.

Bogost, I. (2019), 'When a country bans social media'.
URL: *https://www.theatlantic.com/technology/archive/2019/04/sri-lanka-social-media-ban-bigger-problem/587728/, note =Accessed 29.10.2020*

Bradshaw, S. and Howard, P. N. (2017), 'Troops, trolls and troublemakers: A global inventory of organized social media manipulation'. Accessed 03.10.2020.
URL: *https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/07/Troops-Trolls-and-Troublemakers.pdf*

Brown, H., Guskin, E. and Mitchell, A. (2012), 'The role of social media in the arab uprisings'. Accessed 30.10.2020.
URL: *https://www.journalism.org/2012/11/28/role-social-media-arab-uprisings/*

Burns, M. (1999), 'Information warfare: What and how?'.
URL: *https://www.cs.cmu.edu/~burnsm/InfoWarfare.html*

Chapman, J. (2015), 'Hollywood stereotypes: Why are russians the bad guys?'. Accessed 02.11.2020.
URL: *https://www.bbc.com/culture/article/20141106-why-are-russians-always-bad-guys*

Cherkaoui, T. (2018), 'A new kind of information warfare? cyber-conflict and the gulf crisis 2010–2017'. Accessed 29.10.2020.
URL: *https://www.polecom.org/index.php/polecom/article/view/90/294*

Connell, M. and Vogler, S. (2017), 'Russia's approach to cyber warfare'. Accessed 20.09.2020.
URL: *https://apps.dtic.mil/sti/pdfs/AD1032208.pdf*

El-Khalili, S. (2013), 'Social media as a government propaganda tool in post-revolutionary egypt'. Accessed 10.11.2020.
URL: *https://www.researchgate.net/publication/288694228_Social_media_as_a_government_propaganda_tool_in_post-revolutionary_Egypt*

Facebook (2015), 'October 2020 coordinated inauthentic behavior report'. Accessed 06.11.2020.
URL: *https://about.fb.com/news/2020/11/october-2020-cib-report/*

Facebook (2020*a*), 'April 2020 coordinated inauthentic behavior report'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2020/05/april-cib-report/*

Facebook (2020*b*), 'August 2020 coordinated inauthentic behavior report'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2020/09/august-2020-cib-report/*

Facebook (2020*c*), 'February 2020 coordinated inauthentic behavior report'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2020/03/february-cib-report/*

Facebook (2020*d*), 'July 2020 coordinated inauthentic behavior report'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2020/08/july-2020-cib-report/*

Facebook (2020*e*), 'March 2020 coordinated inauthentic behavior report'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2020/04/march-cib-report/*

Facebook (2020*f*), 'May 2020 coordinated inauthentic behavior report'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2020/06/may-cib-report/*

Ferrara, E. (2017*a*), 'Disinformation and social bot operations in the run up to the 2017 french presidential election'. Accessed 17.10.2020.
URL: *https://arxiv.org/pdf/1707.00086.pdf*

Ferrara, E. (2017*b*), 'Measuring social spam and the effect of bots on information diffusion in social media'. Accessed 15.10.2020.

URL: *https://www.researchgate.net/publication/319326974_Measuring_social_spam_and_the_effect_of_bots_on_information_diffusion_in_social_media*

Freedom House (2019), 'Freedom on the net 2019: The crisis of social media'. Accessed 01.11.2020.
URL: *https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf*

G. King, J. P. and Roberts, M. E. (2017), 'How the chinese government fabricates social media posts for strategic distraction, not engaged argument'. Accessed 27.10.2020.
URL: *https://gking.harvard.edu/files/gking/files/50c.pdf?m=1463683069*

Galeotti, M. (2014), 'The 'gerasimov doctrine' and russian non-linear war'. Accessed 18.10.2020.
URL: *https://inmoscowsshadows.wordpress.com/2014/07/06/the-gerasimov-doctrine-and-russian-non-linear-war/*

Gallacher, J. D., Kaminska, M., Yasseri, T., Kollanyi, B. and Howard, P. N. (2017), 'Social media and news sources during the 2017 uk general election'. Accessed 29.10.2020.
URL: *http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Social-Media-and-News-Sources-during-the-2017-UK-General-Election.pdf*

Geybulla, A. (2016), 'In the crosshairs of azerbaijan's patriotic trolls'. Accessed 20.10.2020.
URL: *https://www.opendemocracy.net/en/odr/azerbaijan-patriotic-trolls/*

Giglietto, F., Righetti, N. and Marino, G. (2019), 'Understanding coordinated and inauthentic link sharing behavior on facebook in the run-up to 2018 general election and 2019 european election in italy'. Accessed 13.10.2020.
URL: *https://osf.io/preprints/socarxiv/3jteh/*

Gleicher, N. (2018*a*), 'Coordinated inauthentic behavior explained'. (Accessed: 01.09.2020).
URL: *https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/*

Gleicher, N. (2018*b*), 'Coordinated inauthentic behavior explained'. Accessed 01.09.2020.
URL: *https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/*

Gleicher, N. (2019*a*), 'Removing coordinated inauthentic behavior from georgia, vietnam and the us'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2019/12/removing-coordinated-inauthentic-behavior-from-georgia-vietnam-and-the-us/*

Gleicher, N. (2019*b*), 'Removing coordinated inauthentic behavior from iran'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2019/01/removing-cib-iran/*

Gleicher, N. (2019*c*), 'Removing coordinated inauthentic behavior from russia'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2019/10/removing-more-coordinated-inauthentic-behavior-from-russia/*

Gleicher, N. (2019*d*), 'Removing coordinated inauthentic behavior in thailand, russia, ukraine and honduras'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2019/07/removing-cib-thailand-russia-ukraine-honduras/*

Gleicher, N. (2019*e*), 'Removing coordinated inauthentic behavior in uae, egypt and saudi arabia'.

Accessed 10.09.2020.
URL: *https://about.fb.com/news/2019/08/cib-uae-egypt-saudi-arabia/*

Gleicher, N. (2019*f*), 'Removing coordinated inauthentic behavior in uae, nigeria, indonesia and egypt'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2019/10/removing-coordinated-inauthentic-behavior-in-uae-nigeria-indonesia-and-egypt/*

Gleicher, N. (2019*g*), 'Removing more coordinated inauthentic behavior from russia'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2019/10/removing-more-coordinated-inauthentic-behavior-from-russia/*

Gleicher, N. (2019*h*), 'Taking down coordinated inauthentic behavior in indonesia'. Accessed 10.09.2020.
URL: *https://about.fb.com/news/2019/01/taking-down-coordinated-inauthentic-behavior-in-indonesia/*

Gorwa, R. (2017), 'Computational propaganda in poland: False amplifiers and the digital public sphere'. Accessed 30.10.2020.
URL: *http://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2017/06/Comprop-Poland.pdf*

Hill, R. (2017), 'Oxford profs tell twitter, facebook to take action against political bots'. Accessed 08.10.2020.
URL: *https://www.theregister.com/2017/06/20/oxford_profs_tell_social_media_to_fight_political_bots/?mt=1497973934083*

Hochhild, J. L. (2010), 'If democracies need informed voters, how can they thrive while expanding enfranchisement?', *Election Law Journal: Rules, Politics, and Policy* . Accessed 21.10.2018.
URL: *https://scholar.harvard.edu/jlhochschild/publications/if-democracies-need-informed-voters-how-can-they-thrive-while-expanding-en*

Hugubakken, T. (2016-2020), 'Pst'. Accessed 30.10.2020.
URL: *https://www.pst.no/alle-artikler/?FilterByValues=2*

Hwang, T. (2019), 'Maneuver and manipulation: On the military strategy of online information warfare'.

Jaitner, M. (2015), 'Russian information warfare: Lessons from ukraine - chapter 10 in kenneth geers (ed.), cyber war in perspective: Russian aggression against ukraine, nato ccd coe publications, tallinn 2015,'. Accessed 13.10.2020.
URL: *https://ccdcoe.org/uploads/2018/10/Ch10_CyberWarinPerspective_Jaitner.pdf*

Johnsen, R. (2020), 'Cyber defence tactics - part ii: Operations and tactics'. Accessed 15.04.2020.
URL: *https://learn-eu-central-1-prod-fleet01-xythos.s3-eu-central-1.amazonaws.com/5def77a38a2f7/3325617?response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%272020-01-16%2520Cyber%2520Tactics%2520Part%2520II.pdf&response-content-type=application%2Fp*

Kalsnes, B. (2019*a*), 'Falske nyheter - løgn, desinformasjon og propaganda i den digitale offentligheten'.

Kalsnes, B. (2019*b*), ‘Når elever blir manipulert’. Accessed 30.10.2020.
URL: *https://bentekalsnes.wordpress.com/2019/09/11/nar-elever-blir-manipulert/*

Kelly, S., Truong, M., Shahbaz, A., Earp, M. and White, J. (2017), ‘Manipulating social media to undermine democracy’.
URL: *https://freedomhouse.org/report/freedom-net/2017/manipulating-social-media-undermine-democracy, note =Accessed 15.09.2020*

KH디지털2 (2013), ‘11 cyber warfare agents face indictment - suspects accused of posting political comments online’. Accessed 29.10.2020.
URL: *http://www.koreaherald.com/view.php?ud=20131219000660*

Khomko, M. M. (2015), ‘Fueling the revolution: Social media’s effect on societal revolutions’. Accessed 19.10.2020.
URL: *https://repositorio.iscte-iul.pt/bitstream/10071/11198/1/Dissertation%20pdf.pdf*

Kioski, Y. (n.d.), ‘Yle kioski investigated: This is how pro-russia trolls manipulate finns online – check the list of forums favored by propagandists’. Accessed 01.11.2020.
URL: *http://kioski.yle.fi/omat/troll-piece-2-english*

Knox, R. (2018), ‘Transforming mexico: social movements, human rights and social media’. Accessed 18.10.2020.
URL: *http://etheses.whiterose.ac.uk/23180/1/Final%20PhD%20thesis%20-%20Transforming%20Mexico%20-%20Rupert%20Knox.pdf*

Kramer, A. D. I., Guillory, J. E. and Hancock, J. T. (2014), ‘Experimental evidence of massive-scale emotional contagion through social networks’.
URL: *https://www.pnas.org/content/pnas/111/24/8788.full.pdf*

Kremlin (2019), ‘Стратегия национальной безопасности Российской Федерации до 2020 года’.
URL: *http://kremlin.ru/supplement/424*

Meserve, S. A. and Pemstein, D. (2017), ‘Google politics: The political determinants of internet censorship in democracies’, *Political Science Research and Methods* **6**(2), 245–263.

Milan, S. (2015), ‘When algorithms shape collective action: Social media and the dynamics of cloud protesting’.
URL: *https://journals.sagepub.com/doi/pdf/10.1177/2056305115622481*

Morgan, S. (2018), ‘Fake news, disinformation, manipulation and online tactics to undermine democracy’, *Journal of Cyber Policy* **3**, 39–43.

Morla, R. (2015), ‘Ecuador’s correa recruits legion of social-media trolls’. Accessed 19.10.2020.
URL: *https://en.panampost.com/rebeca-morla/2015/01/26/ecuadors-correa-recruits-legion-of-social-media-trolls/*

Muhammad, Z. (2020), ‘Study reveals how social media algorithms create echo chambers’. Accessed 18.09.2020.
URL: *https://www.digitalinformationworld.com/2020/06/study-reveals-how-social-media-algorithms-create-echo-chambers.html#*

Nederhoed, L. (2019), ‘Folkeopplysningen 5 - make lillestrøm great again’.

Newmann, N., Fletcher, R., Schulz, A., Andı, S. and Nielsen, R. K. (2020), 'Reuters institute digital news report 2020'. Accessed 03.10.2020.
URL: *https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf*

News, V. (2015), 'Venezuela ruling party games twitter for political gain'. Accessed 18.10.2020.
URL: *https://www.voanews.com/americas/venezuela-ruling-party-games-twitter-political-gain*

Nizzoli, L., Tardelli, S., Avvenuti, M., Cresci, S. and Tesconi, M. (2020), 'Coordinated behavior on social media in 2019 uk general election'.
URL: *https://arxiv.org/abs/2008.08370*

NRK (2019), 'Valgresultat 2019'. Accessed 10.09.2020.
URL: *https://www.nrk.no/valg/2019/resultat/nb/sted/3030*

O'Carroll, T. (2017), 'Mexico's misinformation wars : How organized troll networks attack and harass journalists and activists in mexico'. Accessed 19.10.2020.
URL: *https://medium.com/amnesty-insights/mexico-s-misinformation-wars-cb748ecb32e9*

Ofori-Attah, K. and Martin, G. (2011), 'Internet technology enhanced classrooms'. Accessed 28.11.2020.
URL: *https://www.researchgate.net/profile/Kwabena_Ofori-Attah/publication/299534990_Internet_Technology_Enhanced_Classrooms/links/56fde0ce08ae1408e15b38ad/Internet-Technology-Enhanced-Classrooms.pdf*

Onafuwa, O. E. (2017), 'Propaganda or persuasion? a review of the nigeria 2015 presidential election campaign process via social media (part one)'. Accessed 14.09.2020.
URL: *https://www.researchgate.net/publication/321198325_Propaganda_or_Persuasion_A_Review_of_the_Nigeria_2015_Presidential_Election_Campaign_Process_via_Social_Media_Part_One*

Panarin, I., Jaitner, M. L. and (Ed.), K. G. (2015), 'Cyber war in perspective: Russian aggression against ukraine'. Accessed 17.10.2020.
URL: *https://ccdcoe.org/uploads/2018/10/CyberWarinPerspective_full_book.pdf*

Patalakh, A. (2020), 'Italy as the kremlin's 'trojan horse' in europe: Some overlooked factors'. Accessed 01.11.2020.
URL: *https://www.e-ir.info/2020/04/11/italy-as-the-kremlins-trojan-horse-in-europe-some-overlooked-factors/*

Pearce, K. E. and Kendzior, S. (2012), 'Networked authoritarianism and social media in azerbaijan', *Journal of Communication* **62**(2), 283–298.

Pham, N. (2013), 'Vietnam admits deploying bloggers to support government'. Accessed 20.10.2020.
URL: *https://www.bbc.com/news/world-asia-20982985*

Rosenblatt, H. and Dhar, V. (2020), 'Social-media platforms are undermining our democracy. lawmakers need to step up and protect it'.
URL: *https://www.businessinsider.com/social-media-platforms-facebook-google-twitter-undermining-democracy-2020-9?r=US&IR=T*

Rubinstein, R. (2019), 'Report: Iranian bot army trying to influence israeli elections'. Accessed

18.10.2020.
URL: *https://www.ynetnews.com/articles/0,7340,L-5455991,00.html*

Rueda, M. (2012), '2012's biggest social media blunders in latam politics'. Accessed 28.10.2020.
URL: *https://about.fb.com/news/2020/11/october-2020-cib-report/*

Rujevic, N. (2017), 'Serbian government trolls in the battle for the internet'. Accessed 12.10.2020.
URL: *https://www.dw.com/en/serbian-government-trolls-in-the-battle-for-the-internet/a-37026533*

Sandnes, O. T. (2020), 'Den digitale trusselen'. Accessed 20.02.2020.
URL: *https://www.tv2.no/spesialer/nyheter/den-digitale-trussele*

Sang-Hun, C. (2013), 'South korean intelligence agents accused of tarring opposition online before election'. Accessed 01.11.2020.
URL: *https://www.nytimes.com/2013/06/15/world/asia/south-korean-agents-accused-of-tarring-opposition-before-election.html?_r=0*

Seong-min, K. (2013), '北, 對南 사이버테러요원 3000명... 댓글 다는 전문 요원만 200여명'. Accessed 15.10.2020.
URL: *https://www.chosun.com/site/data/html_dir/2013/08/13/2013081300176.html*

Shane, S. and Mazzetti, M. (2018), 'The plot to subvert an election'. Accessed 21.10.2020.
URL: *https://www.nytimes.com/interactive/2018/09/20/us/politics/russia-interference-election-trump-clinton.html*

Shearlaw, M. (2015), 'From britain to beijing: how governments manipulate the internet'. Accessed 18.10.2020.
URL: *https://www.theguardian.com/world/2015/apr/02/russia-troll-factory-kremlin-cyber-army-comparisons*

Singer, J. B. and Ashman, I. (2009), '"comment is free, but facts are sacred": User-generated content and ethical constructs at the guardian'.
URL: *https://www.tandfonline.com/doi/full/10.1080/08900520802644345?scroll=top&needAccess=true* note = *Accessed 19.09.2020, year = 2020*

Springer (2020), 'Overview of imrad structure'.
URL: *https://www.springer.com/gp/authors-editors/journal-author/overview-of-imrad-structure/1408*

Staff, R. (2017), 'Vietnam unveils 10,000-strong cyber unit to combat 'wrong views''. Accessed 21.10.2020.
URL: *https://www.reuters.com/article/us-vietnam-security-cyber/vietnam-unveils-10000-strong-cyber-unit-to-combat-wrong-views-idUSKBN1EK0XN*

Standish, R. (2017), 'Why is finland able to fend off putin's information war?'. Accessed 01.11.2020.
URL: *https://foreignpolicy.com/2017/03/01/why-is-finland-able-to-fend-off-putins-information-war/*

Stenslie, S., Haugom, L. and Vaage, B. H. (2019), 'Ettterretningsanalyse i den digital tid - en innføring'.

Team, R. C. (2019), 'Nigerian elections 2019: The spread of false information'. Accessed 14.09.2020.
URL: *https://www.bbc.com/news/world-africa-47226397*

Trending, B. (2016), 'Who's at the controls of iran's bot army?'. Accessed 19.10.2020.
URL: *https://www.bbc.com/news/blogs-trending-35778645*

Tufekci, Z. (2019), 'The imperfect truth about finding facts in a world of fakes'. Accessed 29.09.2020.
URL: *https://www.wired.com/story/zeynep-tufekci-facts-fake-news-verification/*

Vega, A. F. (2012), 'Spambots driving mexican twitter users crazy ahead of presidential election'. Accessed 28.10.2020.
URL: *https://slate.com/technology/2012/06/spambots-on-twitter-for-mexican-presidential-candidates-like-the-pri-s-enrique-pena-nieto.html*

Vilmer, J.-B. J., Escorcia, A., Guillaume, M. and Herrera, J. (2018), 'Information manipulation - a challange for our democracies'.

Williams, S. (2017), 'Rodrigo duterte's army of online trolls - how authoritarian regimes are winning the social media wars'. Accessed 04.10.2020.
URL: *https://newrepublic.com/article/138952/rodrigo-dutertes-army-online-trolls*

York, J. C. (2011), 'Syria's twitter spambots'. Accessed 30.10.2020.
URL: *https://www.theguardian.com/commentisfree/2011/apr/21/syria-twitter-spambots-pro-revolution*

Yuan, A., Gillani, N., Vosoughi, S., Roy, D. and Saveski, M. (2018), 'Me, my echo chamber, and i: Introspection on social media polarization'. Accessed 18.09.2020.
URL: *https://arxiv.org/pdf/1803.01731.pdf*

Zakem, V., and. Antoun, P. S., Gorenburg, D. and Markowitz, M. (2015), 'Mobilizing compatriots: Russia's strategy, tactics, and influence in the former soviet union'. Accessed 02.11.2020.
URL: *https://www.cna.org/cna_files/pdf/DOP-2015-U-011689-1Rev.pdf*

Zuckerberg, M. (2017), 'Mark zuckerbergs facebook profile'. Accessed 29.10.2020.
URL: *https://www.facebook.com/zuck/posts/10104067130714241*

Chapter 6

Social Media Awareness and Operational Education

By Kristian Tørseth

6.1 Executive summary

In today's societies social media plays a important part in everyday life. Whether this is to stay in touch with friends or keep up with what goes on in the local news, social media is the place to be. When looking on how social media, internet and democratic election work together during election time is something that is some what new to the general population, and is complex as it is unfolding.

Social media manipulation is not a new phenomenon. Methods as "fake news" and propaganda has been using effectively long before the 2016 elections. With the ever-growing population being more and more globally connected with smartphones and almost everyone connected with Facebook or Twitter. Social media has been over the past ten years been a useful tool for group's to spread misinformation and manipulate the average user of social media. Although social media is not the sole reason to the increasing polarization in political group's, it amplifies and gives life to an increasing polarization and social media is the perfect tool to undermine the sole integrity of a democratic election.

This paper will try to give an understanding of how humans have been manipulated by news for hundreds of years. It takes up how "fake news" and propaganda has been used as tools for election manipulations. It will also be presented how social media is helping the increasing polarization in America and display how social media companies are not protecting their users from false or incorrect information on their platforms. The trust in mass media will be questioned and data can be displayed how depending on what political party whom chooses to vote for the trust either deteriorates or improves.

An angle to show how "us" the public should have been thinking and being critical to content seen online before, during and after the election in 2016 is discussed further. How false news manages to spread so fast and widely is not only a problem from the social media platform companies but something that the users of their "tools" should have been very aware off. Then fix the general malpractices of use of social media to share and spread misleading information a set off tools are proposed to be integrated into everyday use to validate and authenticate social media content. Whether this is in the form of validating time, images, or text a set of tools is submitted to be used for everyone.

To finish of this chapter mitigation methods implemented by social media companies mitigating spreading of misinformation of election-oriented content is presented. This also draws similarities to how the COVID-19 situation is handled by social media companies as COVID-19 misinformation should be handled with the same care and delicacy as election manipulated content as content about either of these could potentially be devastation for a multitude of group's. A common defense against the spreading and abuse of misinformation spreading on social media should be high on the agenda for everyone, but this will depend on choices of policy changes by social media companies and the general public's use of social media. The necessity to critically think before sharing and writing incorrect on content that is shared on social media is one of the main educational points and the general public should spend more time thinking like a journalist on social media and be overall more sceptical and not so benevolent when watching media on social media.

6.2 Introduction

Majority of 2016's well-established countries are today using social media as a tool to connect and interact with their friends, community, and family. Social medias like Facebook, Twitter, Reddit and Instagram are just some of the major social media actors in that is used every day. They offer a "free" tool that helps you connect with other people around the globe. Social media is helping families and old friend all around to world to find each other and stay connected. With social medias self-disclosure has become one of the most essential things that is done to maintain and to build new network. People updating profile information, whereabouts, and status updates, uploading videos and pictures, liking and commenting are some examples of everyday use of social medias.

6.2.1 Evolution of fake news

The definition of "fake news" is not a new phenomenon. The term fake news can be traced back to the 13th century BC portraying the battle Kadesh (*Fake news - Wikipedia*, n.d.), and is portrayed in the same manner as when this term became known to every person in the United states of America during and after the 2016 U.S. election. But during the elections of 2016 social media become a powerful tool to spread miss-information. But social media is not he first media type that has been exploited. During the 1900th century many different types of spreading miss-information has been used, and if we look back 70 years we see the efficient usage of propaganda during the war to portrait their enemies as these big bad monsters that is coming to eat our children. Propaganda is focused on playing on people's emotions and thoughts with messages that wasn't always true, and we can say that propaganda is somewhat in the same alleyway as fake news. Another type of this is conspiracy theories. These theories are not something that is bound to modern time only. There is examples from the 1835 a series of articles published by *New York Sun* about how life on the moon had been discovered (*"The Great Moon Hoax" Is published in the "New York Sun" - HISTORY*, n.d.).

Conspiracy theories is not "fake news" because they are difficult to verify either as reliable and true or false and lying. What is also a factor when talking about conspiracy theories is that they usually have a substantial following. We can draw some similarities between conspiracy theories that is about political controversies, and from 6.1 we see a new survey done in 2019 that show how Americans are prone to believing preposterous theories. We can see that 47% of the people asked still believes that Lee Harvey Oswald did not act alone in assassinating JFK.

These are some of the more conspiracy theories that Americans do believe, and the reason why only

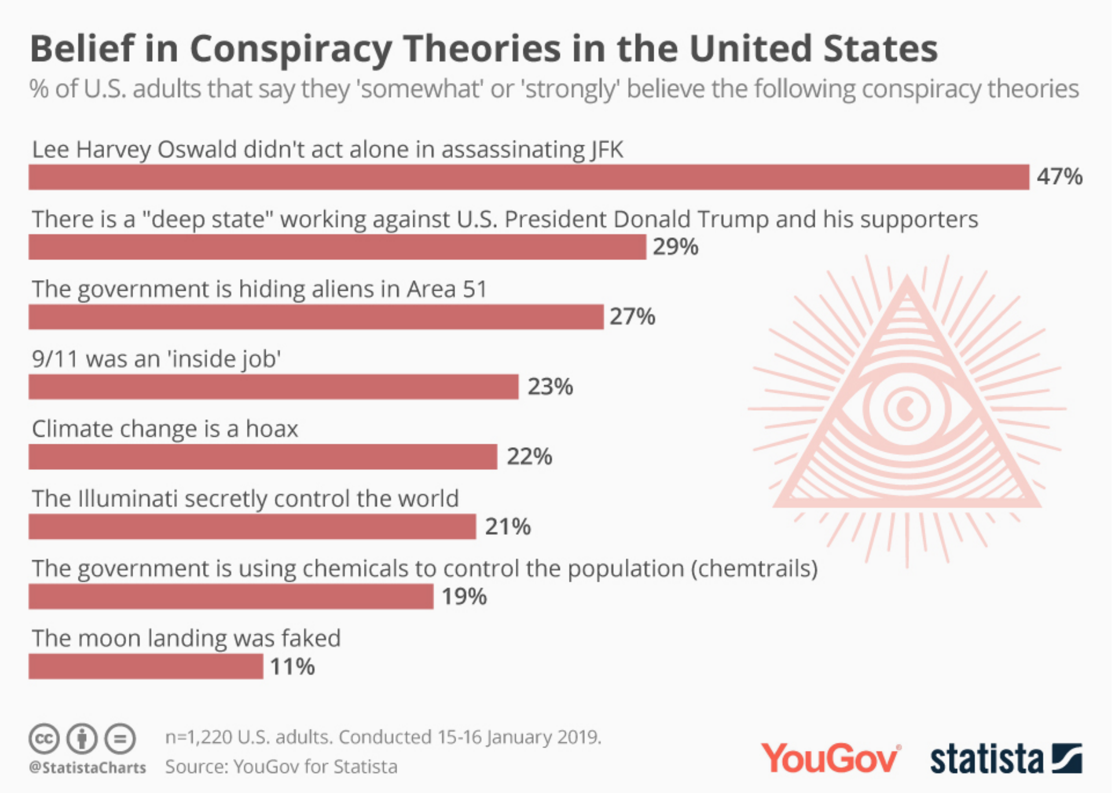


Figure 6.1: Data gathered by YouGov for Statistica shows how many U.S. adult that believes in conspiracy theories in 2019 (Statista, n.d.*a*).

49% believe in the JFK assassination and not 100% is that these hoaxes has been denied and also fact checked by professional journalism that is working in reputable newspapers or tabloids, and this has made sure that these theories are quickly denied and disclosed as hoaxes. Now the barriers of good journalistic has been lowered with the growing usage of web services. It is now easy to even for people with no knowledge on IT and programming to easily set up a website and be able to quickly monetize the site with the help of advertisements.

Also, with the rapidly growth of social media, these new websites have a huge platform to spread their information. Before the 2016 elections Twitter had 325 million monthly active users (*Twitter Users Statistics 2016 Infographics | GMI*, n.d.), Facebook had approximate 1750 million monthly active users (Statista, n.d.*b*) and this gives an indication on how new news sites can use social media as a place to gain traction and "clicks" to their websites. These new news websites also go under names that is similar to what reputable news sites are named. Some examples of this is NationalReport.net and WashingtonPost.com.co, these were all owned by a company called Disinformedia and is a US owned company with 20 - 25 writers that post "fake news". One of these stories for the Denver Guardian about an FBI agent that leaked Hillary Clinton emails had been killed. Stories like this build upon right-wing conspiracy and got approximately 1.6 million clicks in the first ten days (*We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned : All Tech Considered : NPR*, n.d.).

Social media also brings the general trust to established news outlets down, with "fake news" spreading so rapidly on social media platforms reputable news outlets is finding it hard to compete with "click bait" headers and engaging stories. Trusted news outlets are not gaining as much traction in social media for many reasons. One of them is how rapid "fake news" sites are often fist at commenting about a resent news event. But, what they report may not be true at all, but the fact that they were the fist and managed to get good engagement from left/right -oriented people may cause the "fake news" story to be the most shared and the most discussed news. With this said, normally "fake news" will be called out with some time, but when the news are fresh and not been clarified "fake news" sites will win most of the time.

6.2.2 How social media is helping splitting the population

In the United States the general polarization has increased with the rapid sharing of "fake news" that is targeted towards left/right -wing people in the US. With this the overall distance and hatred towards each group's polarization has spiked because of the impact social media had during the 2016 election. About 67% of adult Americans uses a social media platform, such as Facebook, Twitter and Reddit for their news consumption. And about 64% of these gets their news from one site, 26% gets their news from two sites (*News Use Across Social Media Platforms 2016 | Pew Research Center*, n.d.). By only checking less than two sites we can assume that users are prone to be exposed to news sites with a "bias" either towards the right or left -wing in the US. This is also bringing the trust in mainstream media down as people are being more exposed to fake news sites that works as a echo-chamber for the users(Allcott and Gentzkow, n.d.).

The polarization of the US. population is not necessarily only social medias fault, but the trends that we can see from 6.2 that the republicans have lost much of the trust in mainstream media, while the democrats have increased their consumption of mainstream media. The republicans mostly use FOX news as their only mainstream media outlet that they mostly trust. This can show us that republicans are more susceptible to biased news reporting and misinformation in newspapers than their counterpart the democrats. The decline of the republicans trust to mainstream media can be that President Donald Trump accusation towards the media as "the enemy of the poplar" and with this we see from 6.2 the republicans decline around 2016 when President Donald Trump was campaigning for presidency (*Americans' Trust in Mass Media Edges Down to 41%*, n.d.).

6.2.3 Exposure to "fake news"

In the run up to the 2016 US. presidential election Facebook was a prime location where "fake news" was shared and in heavy circulation. pro-Trump supporters with a conservative online history was the most targeted group for "fake news" on Facebook. Visits to Trump favored "fake news" sites were increased during president trumps campaign for both Clinton supporters with a increase from 2.8% to 16.1% to pro-Trump "fake news" sites. The most drastic increase was with Trump supporters who increased from 16.3% to over 63%. Similar patterns can also be seen for Clinton supporters towards pro-Clinton "fake news" in the same time frame and all the majority comes from news shared on Facebook (Guess et al., 2018).

6.3 Trust in mass media in an era of social media

After the 2016 presidential election a wider focus on detecting "fake news" has become something the public is interested in understanding and be able to detect for themselves. Authentic news is not a science and news are built up differently and reported differently. The balance between being first with the news or having the facts right before publishing is something journalist have to fine tune to find a balance. So, for the users reading there is always the need to be critical to what is read. The use of critical thinking to detect news that has been fabricated and tampered with is of importance. This critical thinking is something that everyone uses in their everyday life, but before the 2016 the general public wasn't using this to its full intent when browsing and reading news on the online, and with the growth in usage of social media the public was acceptable to news without using critical thinking.

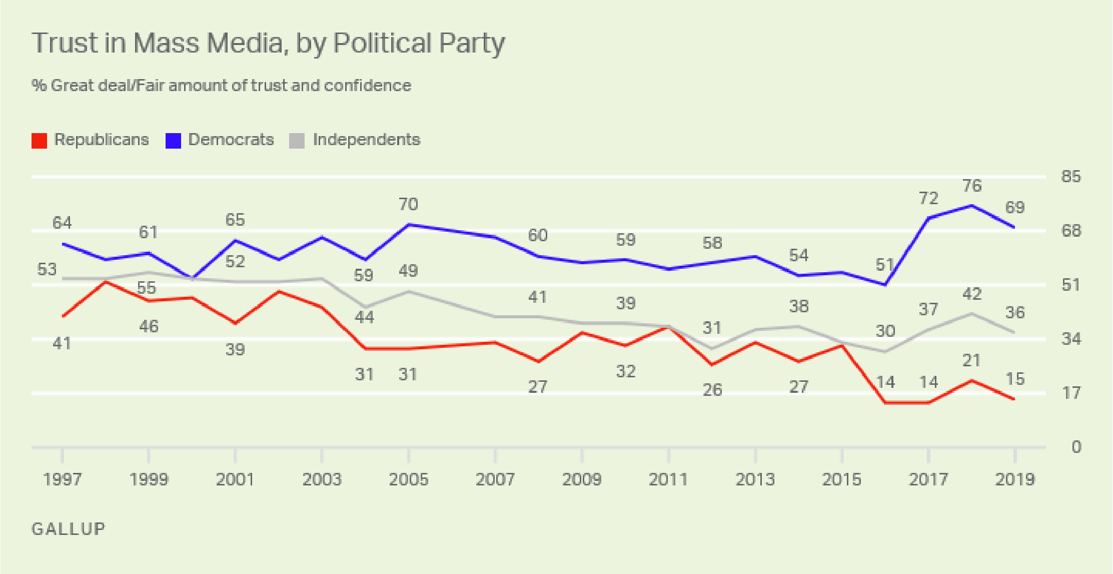


Figure 6.2: Graph showing the increase and decline in trust towards mainstream media(*Americans' Trust in Mass Media Edges Down to 41%*, n.d.).

This misinformation disguised as real news has been detected during the French, German and the U.S election. In Mexico 2017 social media as well as mass media covered about a girl (#FridaSofía) that was trapped after a earthquake had struck Mexico. This was "fake news" and there was no schoolgirl trapped named Frida or Sofía. This "fake news" story had gain so much traction on social media that it seemed creditable. In this case this was not a malicious act of spreading misinformation, but the public as well as the responsible journalists must be of that mistakes and counterfeit information may be present in on going news stories. This reflects back onto social media and people sharing using social media as a tool. Although it is easy to spot news that fails to deliver hard facts such as polar bear spotted in Berlin. This we can clearly see it is fraudulent and deceptive and this fails check of as logical for the public. There is a difference between fraudulent writing of news and sloppy reporting, and while election-oriented news often is written with an agenda there could be a possibility of sloppy reporting. With young people getting their news on their phones via social media application, while on some social media applications there is no distinguishing by the application between reputable news reporting versus what is armature reporting (Cherilyn Ireton and Julie Posetti, 2018).

6.4 Social media verification

The rapid growth of visuals in form of photos, videos and GIFs gives the users of social media an enormous exposure to visual information. With 14.2 billion phones (Statista, 2020) where most of them have a camera embedded, media (photos and video) have never been captured and shared online as fast as we see today. In some countries mobile data is also almost free for the users to use, and this brings many of the users to social media where they share and update. In many cases a status, photo or video shared on social media is often where a "fake news" story is first shared. This can be a disaster, attack or something a politician said. These stories are often to be shared by someone that was close to the origin of the story with a mobile phone. These status updates or story origins may be affected by the persons personal biases and could contain unwillingly false information that was "lost in translation".

So, when sharing a post some key steps should be taken by the person sharing. The person who is going to share need to check and be able to identify the originator of the information shared as well as images used in the status or story. This check is the same check that a professional journalist would do to its sources, and are important when sharing to multiple people on social media. Using normal logic and trusting instinct is also a guidance tool that must not be under prioritised and used. Conclusions should be made about the confirmations of a source even though source cannot be accessed in person. This verification of the source and origin is important and especially with visual content. For malicious actors it is simple to create and share fake photos and stories.

With the high volume of visual media available there can be a certainty that photos will be recycled and used in different contexts in social media post, news, or election material. This is something that even politicians and professionals gets fooled alike. Editing and sharing with scepticism is a vital tool in battling "fake news" spreading on social media. People who report "on-site" and watching a event unfold are less likely to deceive and are just reporting based on what they are experiencing. If there is misinformation in what the person reported on their social media, then it may not be of malicious intent, it could rather be the possibility that the person writing on his/her social media just had a different view on the event, and remembers it differently. Although difficult to distinguish between this and "troll" behavior "fake news" sites sets out in motion there are some checkpoints that can be done to verify the post, news or image a person would like to share on social media (Cherilyn Ireton and Julie Posetti, 2018).

- Is the content of the post original.
- Is the media manipulated in some way.
- Is it possible to confirm the time and place of the media using metadata.
- Is it possible to confirm the time and place using visual clues in the media shared.

These are some of the clues that needs to be understood and verified before sharing on social media, and if this was followed by the general public prior to the 2016 US. election there would possible have been less circulation of “fake news” on social media. This also applies to post that is not political oriented, but could instead be targeting a vulnerable group in society. With these verification clues in mind here are some of the most common types of manipulation of visual content on social media.

- Wrong time and place.
- Manipulated content.
- Staged content.

Example of manipulation of time and place is a video that emerged in 2017 that he Begaluru airport in India had been flooded. This footage was uploaded to YouTube by Amazing Original Videos but showed a flood that took place which is 14,000 km away(Arun Dev, 2017). Survivors of the school shooting at Marjory Stone Duglas High School in Florida was a target of media that were manipulated against their cause. GIF of one of the survivors ripping the United States Constitution was a fake and the original poster “Gab” said “obviously satire”. This was shared mostly on social media by right oriented people where many of them believed the GIF was real. Reports from right-oriented news outlets targeted David Hogg another school shooting survivor with claims that he did a Nazi salute at the end of one of his speeches, another example of manipulated media was a photo claiming to be Emma Gonzales attacking a 2ND amendment supporter’s car when in reality this was a recycled images of Britney Spears from 2007 (Lytvynenko Jane, 2018). Stage content might be the most dangerous type of manipulation. With the rising use of artificial intelligence to make fake footage that looks to the naked eye like they are 100% real. This fake video of President Barrack Obama `https://www.youtube.com/watch?v=AmUC4m6w1wo` shows how biometrics can be used to make deep fakes (BBC News, 2017).

6.4.1 Tools: Verification social media content

On social media there is still no good option to be able to verify and authenticate images or videos. There some free and some payed options to be able to do analysis of content on social media. To do analysis of Facebook accounts a payed OSINT service called Intel Techniques that analyses and cross references public information available through Facebook with what is found online(Michael Bazzell, 2017). The Facebook tool is a payed option and might not be “reasonably” to use for the average users. What a “normal” user might want to use are free options. Free options like reverse image search. Services like Google reverse image search, tineye and reveye are options that is easily available. Googles reverse images search lets the users search in googles DB for visually similar images or even the same images. This can help to verify facts and images that is used out of its original context. Images can be verified by looking at previous images and look for time stamps, if the time stamps found on images when doing a reverse image search don’t correlate with picture in social media post a huge red flag should be put on the post and the post will most likely be a fake and a fraud. When doing a reverse image search and no hits for that image is presented. This does not mean that the image is original and further checks should be done on the media (Cherilyn Ireton and Julie Posetti, 2018).

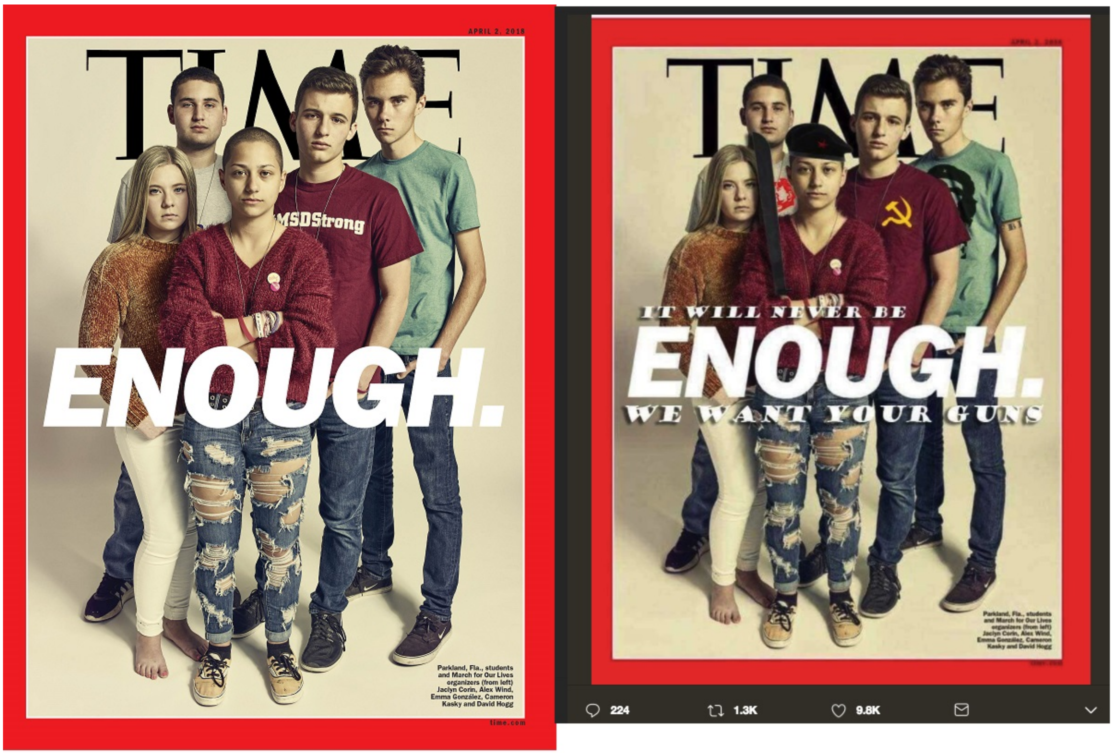

APRIL 2, 2018
TIME
MSDStrong
ENOUGH.
Parkland, Fla., students and March for Our Lives organizers (from left) Jaclyn Corin, Alex Wind, Emma González, Cameron Kasky and David Hogg
time.com
IT WILL NEVER BE
ENOUGH.
WE WANT YOUR GUNS
224
1.3K
9.8K

Figure 6.3: Image to the left is the original image posted on the front page of Time magazine with the caption “Enough”. Image on the right is a **manipulated** photo spread on social media by gun activists. This image is playing on people’s emotions and includes communist symbols as well as an altered and manipulative caption “It will never be enough. We want your guns” (Lytvynenko Jane, 2018).

Video detection is a different story. Usually the public do not have access to “reverse video search”, but this is integrated in YouTubes video algorithm that recommend videos. There are some tools that take the thumbnail and does a reverse image search of the thumbnail. Tools like Amnesty’s YouTube data viewer and NewsCheck are two options. Videos on YouTube, especially news videos often try to “hide” from the detection algorithm of YouTube by scaling videos differently, adding particles like snow or spinning a globe in reverse/ mirroring object in the video. YouTubes own algorithm to check for duplicate videos is discussed elsewhere in this book and not a subject for this chapter. Another way to verify media on social media is to look at metadata attached to the content. This will include data about what digital camera took the media, or if it was a phone that captured the media. It will include important information about data, location, time, light settings which can be beneficial to know when trying to verify if media used in content on social media is true or not. A limitation done by the social media companies strips most of the content shared on social media platform of its metadata when uploaded. This means that media that is shared on Twitter and Facebook don’t include metadata, and this is bad for the verification process, but the protection of data done by the companies can be seen as a way to protect the users public data. Meta data used in media content has another downside and that is that metadata easily can be modified and edited, so relaying only on meta data cannot be the only option in a verification process, but can be used as one of many sources for verification. Often when looking at meta data geolocation can be found to determine where the photo/video/media was captured. Often social media such as Instagram, Facebook and Snapchat uses this geolocation or geotagging as often used. If geolocation is present the possibility to cross reference visuals in the photo/content with a google search can authenticate if fake or not. Checking weather and light conditions can be used also as indicators if media is manipulated. In social media photo manipulation is often used to enhance body features. Some of the more common once's are not directly dangerous to elections this show how widespread manipulations of photos are on social media. manipulation of that makes eyes, lips and muscle and making other body parts look smaller are some of the more commonly used manipulation on social media. For the more interested parties a digital image forensics tool such as Forensically and Photo Forensics can be used to investigate and check the validity of images metadata (Cherilyn Ireton and Julie Posetti, 2018).

6.5 Social media stepping up

6.5.1 Battling COVID-19 misinformation on Twitter

In May 2020 Twitter came out and updated their policy's. The ongoing pandemic COVID-19 got a lot of traction on social media and all sorts of miss leading information was spreading. This was a serious thing going on in the world, and Twitter could not stand by while their platform was used to spread misleading information about COVID-19. Second this was also a prevention system for political miss information that could be spread during the buildup and during the 2020 US. election campaigns. These new labels and warnings on tweets are to protect the public from risk or harm that could be associated with the tweets. These warning was as said first only used to address content that was directly against COVID-19 guidance's from global health, national health, and local health authorities. This was added by Twitter so when users had a dispute on their platform, actual sources could come and back up their COVID-19 claims. Examples of this marking of messages happen when users gave out incorrect numbers of actual infected people in the US. This was a means to add security to information about the pandemic on Twitter. These labels will not only be applied for new tweets tweeted after this new policy was implemented, but also be applied to tweets before policy was. Tweets with manipulative and harmful messages will be labeled as potentially fakes. A warning may also be applied on the tweet depending on the level of harm level of the tweet. This warning label will inform people who wants to watch the tweet that information that's in the tweet might be in direct conflict with guidelines given about COVID-19 by health authorities (Roth and Pickles, 2020).

Three distinct categories were made by Twitter to distinguish between "real" and "fake" information. Misleading information is the first category, and this is labeled by Twitter when information in a tweet contains information that has been confirmed misleading or false by exports, such as public health authorities. Second, is the disputed claim category. This is statements where the truthfulness and accuracy in the tweet are contested or unknown. Last, claims that are unverified. These are claims that could either be true or false, but at the point of time are unconfirmed. By marking unverified claims Twitter is trying to shut a lid on spreading of hoaxes and false information over their platform. Although they will also mark "real" tweets as potentially fake once, they preserve integrity by marking false information right away (BBC News, 2020).

The Twitter team is monitoring contents related to COVID-19 using their own internal systems. These systems are there to make sure that tweets with warnings or labels don't get amplified in their systems, this also monitors tweets that are getting majority of views on Twitter (Roth and Pickles, 2020).

6.5.2 Battling 2020 US election misinformation on Twitter

October 9Th 2020 Twitter implemented another policy change to protect their users from spreading misinformation on Twitter. Twitter is acknowledging their position as a social media platform that is heavily used during election time, and with this new policy change Twitter is holding users of twitter more accountable for their tweets. Twitter understands that tweets can facilitate to democratic conversations, having political debate and enabling the users to hold people of power accountable for their actions. This means that twitter needs to protect their platform from people that will try to undermine this democratic process. Twitter is from November 9Th 2020 not allowing anyone to use their social media platform to manipulate or interfere with an election. This indicates a stricter rule to label and set warning on tweets. Some rules about what can be said on Twitter is also ruled out. It is now not allowed on Twitter to claim that an election has been won before official authorities has concluded a victor of the election. These tweets will be marked as misinformation like the COVID-19 tweets. The same goes for tweet that could interfere with the election process. Tweets that would suggest to violence or scare voters will be removed (Gadde and Beykpour, 2020).

6.6 Conclusion

There is a growing concern with the usage of social media, and from this paper there is clear examples on how social media is contributing negatively right now to spread misinformation. From 2016 there is a miss trust in well-established media outlets and the rise of “fake” news sites has grown in popularity due to people being little or non- critical to what they read online. Now in 2020 mass media and journalist has a golden opportunity to gain back the trust from the public after the public now has “awoken” and started to be more critical to what they read and share. But not only media outlets has to take on the responsibility to share good real news on social media.

The users of social media need to be better educated and protected from sharing fake news stories on social media. This is a “two-man job” as on one side there is us the users who has to be more critical and sceptic to what we share, read and talk about on social media. The users should use tools that journalist would use to confirm sources or do reverse images search to confront sites that are only looking for likes and clicks. The other side is the responsibility the social media platforms must not amplify the spread of misinformation on their platforms. Some social media platforms have started to take steps in the right direction but others are neglecting their users and by the looks of it has their own view on what misinformation is. With the ever-increasing usage of social media not only are democratic elections threatened but also our youngest and brightest. That social medias want the most engagement on their sites is not always a good thing. We see examples of young teenage girls getting recommended anorexia videos on YouTube (Syed-Abdul et al., 2013) and this is because these are the videos that get the most engagement when teenage girls search for diet videos on YouTube. And, this is also true for when misinformation is shared on social media. If the misinformation is the most engaging story then that is what the social media algorithm for recommendations will recommend to their users.

Bibliography

Allcott, H. and Gentzkow, M. (n.d.), ‘Social Media and Fake News in the 2016 Election’.
URL: *https://doi.org/10.1257/jep.31.2.211*

Americans' Trust in Mass Media Edges Down to 41% (n.d.).
URL: *https://news.gallup.com/poll/267047/americans-trust-mass-media-edges-down.aspx*

Arun Dev (2017), 'Fake Video Claiming Bengaluru Airport Was Flooded, Is From Mexico'.
URL: *https://www.thequint.com/news/webqoof/fake-video-claiming-bengaluru-airport-was-flooded-is-from-mexico*

BBC News (2017), 'Fake Obama created using AI video tool - BBC News - YouTube'.
URL: *https://www.youtube.com/watch?v=AmUC4m6w1wo*

BBC News (2020), 'Twitter bans David Icke over Covid misinformation - BBC News'.
URL: *https://www.bbc.com/news/technology-54804240*

Cherilyn Ireton and Julie Posetti (2018), *Handbook for Journalism Education and Training UNESCO Series on Journalism Education*.
URL: *http://www.unesco.org/open-access/terms-use-ccbysa-en*

Fake news - Wikipedia (n.d.).
URL: *https://en.wikipedia.org/wiki/Fake_news*

Gadde, V. and Beykpour, K. (2020), 'Additional steps we're taking ahead of the 2020 US Election'.
URL: *https://blog.twitter.com/en_us/topics/company/2020/2020-election-changes.html*

Guess, A., Nyhan, B. and Reifler, J. (2018), 'Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign'.
URL: *https://about.fb.com/wp-content/uploads/2018/01/fake-news-2016.pdf*

Lytvynenko Jane (2018), 'Here Are The Hoaxes And Conspiracies Still Going Around About The Parkland Students'.
URL: *https://www.buzzfeednews.com/article/janelytvynenko/here-are-the-hoaxes-and-conspiracies-still-going-around#.jhe2YvV44*

Michael Bazzell (2017), 'IntelTechniques.com | OSINT & Privacy Services by Michael Bazzell | Open Source Intelligence'.
URL: *https://inteltechniques.com/contact.html*

News Use Across Social Media Platforms 2016 | Pew Research Center (n.d.).
URL: *https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/*

Roth, Y. and Pickles, N. (2020), 'Updating our approach to misleading information'.
URL: *https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html*

Statista (2020), '• Number of mobile devices worldwide 2020-2024 | Statista'.
URL: *https://www.statista.com/statistics/245501/multiple-mobile-device-ownership-worldwide/*

Statista (n.d.*a*), 'Chart: Belief in Conspiracy Theories in the United States | Statista'.
URL: *https://www.statista.com/chart/18196/belief-in-conspiracy-theories-in-the-united-states/*

Statista (n.d.*b*), 'Facebook: mobile monthly active users 2016 | Statista'.

URL: https://www.statista.com/statistics/277958/number-of-mobile-active-facebook-users-worldwide/

Syed-Abdul, S., Fernandez-Luque, L., Jian, W. S., Li, Y. C., Crain, S., Hsu, M. H., Wang, Y. C., Khandregzen, D., Chuluunbaatar, E., Nguyen, P. A. and Liou, D. M. (2013), 'Misleading health-related information promoted through video-based social media: Anorexia on youtube', *Journal of Medical Internet Research* **15**(2).

"The Great Moon Hoax" Is published in the "New York Sun" - HISTORY (n.d.).
URL: https://www.history.com/this-day-in-history/the-great-moon-hoax

Twitter Users Statistics 2016 Infographics | GMI (n.d.).
URL: https://www.globalmediainsight.com/blog/twitter-users-statistics/

We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned : All Tech Considered : NPR (n.d.).
URL: https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs?t=1604837116519

Chapter 7

Disclosure and Mitigation of SMME – Technical Solutions, Limitations and Challenges

By Kristian Havstein

7.1 Executive summary

This paper examines the current automation techniques and tools associated with social media manipulation of elections (SMME) by coordinated inauthentic behavior (CIB). It gives an overview of current problems and solutions for detection and mitigation on social media platforms, particularly by machine learning. Limitations in these tools are described. The paper looks at future trends in tactics, machine learning advances and synthetic media creation (deepfakes) through use of neural networks. Methods used consists of systematic literature review of western peer-reviewed papers, primary sources such as news sites, in addition to government- and NGO-reports.

The need for primarily human detection (classification) of many content features – such as truth and social context persists and will not change in the foreseeable future. Recent advances in machine learning show promising results in classification of additional feature types and a significatn increase in accuracy. Breakthroughs in creation of part- or wholly synthetic media by neural network-based machine learning is rapidly becoming democratized and is poised for increased utilization by cyber troops, advertisement firms and regular hobbyists. Publicly available examples of this technology are already advanced enough to consistently deceive more than half of human observers when examining short text and can generate quite believable pictures of faces, bodies, video and voices. Machine learning detection of synthetic media does not show promising accuracy, either. There are examples of heads of state promoting deepfakes of political opponents in social media during the 2020 US election. New initiatives for detection have been proposed by Facebook, Microsoft and Adobe.

The use of automation for Coordinated inauthentic behavior (CIB) and SMME by government cyber troops and private firms is increasing in scope and technical abilities. Only a handful of nations and states have enacted regulations of social media companies and synthetic media.

There is an urgent need for:

- Accurate automated detection and mitigation of coordinated inauthentic behavior and synthetic media
- Investment in human moderation by social media companies to backstop current trends
- Government regulation of social media companies and mandates that ensure transparency

The lack of transparency by social media companies towards the general public and researchers leave only inferences when evaluating research results that have some degree of transparency, research based on publicly available sources and empirical, often anecdotal observations.

7.2 Preface

This paper will look at techniques used to manipulate social media platforms through CIB (Gleicher, 2018), to accomplish SMME. Associated techniques for technical detection and mitigation will also be presented. Further, it will describe the evolution of both attacks and defenses, including the increasing capabilities for creation of synthetic media, and what the future might have in store for us.

This paper will not look at technical aspects where authorities outright block internet resources, social media platforms or internet access altogether.

7.3 Introduction

The use of social media platform infrastructure to influence public political opinion has global reach and can now be accomplished from anywhere in the world, directed at any community in the world. Advertisings targeting-problem has largely been solved by advances in data mining of user content and behavior in social media platforms. The cost and difficulty of quickly reaching large and specific segments of a population with targeted political messaging has declined dramatically (Bradshaw and Howard, 2018).

In the same time frame, automated political communication, defined by Howard, Woolley and Calo (2018) as "[...] the creation, transmission, and controlled mutation of significant political symbols over expansive social networks" have been refined and expanded to accomplish SMME. They identify this relatively new political trend as: "..among the most important recent innovations in political strategy and communication strategy" and conclude that: "[...] the prevalence, variety, and influence of computational propaganda on political communication will only grow"(Calo et al., 2018).

The cost and technical prowess needed to wield these tools has declined and is now within reach for small countries and local governments (Freedom House, 2019). Services providing this are also offered by political advertisement firms that leverage big data analytics, and politicians are increasingly putting these tools to use (Bradshaw and Howard, 2018). Government authorities are expanding their capacity to shape public opinion in this space by deploying cyber troops (Bradshaw and Howard, 2018) that employ both humans and automation to accomplish their goals. For some cyber troops these goals include creation and distribution of disinformation, fomenting civil unrest, sowing distrust in democratic institutions and authorities, and outright election manipulation in democratic countries.

With this backdrop, there have been several breakthroughs in the creation of synthetic media by ML, and neural networks in particular. Human believable and machine undetectable fully synthetic as well as manipulated media, so-called - deepfakes - is poised to enter social media in force.

7.4 Description of techniques and tools of manipulation

Large scale campaigns for SMME are carried out by domestic or foreign cyber troops and political advertisement firms which sell services for online campaigns using big-data analytics. There is also a growing subset of these businesses that use techniques such as bot farms, in breach of social media platforms' acceptable use policies (Bradshaw and Howard, 2018).

Cyber troops are defined by Bradshaw and Howard (2018) as government or political party actors tasked with manipulating public opinion online. These can consist of professionals, private firms, volunteers, or all of the above. Some governments have established defensive cyber troops to combat social media manipulation operations by foreign adversaries. Bradshaw and Howard (2018) document how troop numbers and their use is increasing throughout the world. Both for offensive and defensive operations. They also note that capacities established for defensive operations today, can later be used offensively.

There are several types of information that can be spread to manipulate social media platforms. Freedom House identified five methods of content manipulation in their 2019 Freedom on the net report (Freedom House, 2019):

- Propagandistic news
- Outright fake news
- Paid commentators
- Bots (automated accounts)
- Hijacking of real social media accounts

The content can be divided into text or multi-media type.

There are different ways of effectively propagating content on social media platforms. These ranges from legitimate use of platforms to outright misappropriation by manipulation. All methods leverage social media algorithm optimization or manipulation for better visibility and propagation. The algorithms are trade secrets and social media companies have not been open with researchers in this regard (Calo et al., 2018). Publication of these algorithms may make platforms that rely on them more vulnerable to manipulation by malicious actors. Aral et al. (2018) found that political false news is diffused in social media platforms faster than any other type of content. Typical factors like account age, network size, activity and so forth showed that this contents diffusion happened despite low scores in otherwise crucial metrics (Aral et al., 2018).

The legitimate use of political PR often involves use of partisan or paid commentators and advertisement. Paid commentators can be paid secretly, or through arrangements such as think tanks. Advertisements in social media can be highly targeted if the advertiser has enough data to select targets. Data for accurate targeting can also be a feature of the social media platform.

Trolls can be used to harass or intimidate legitimate voices in the political discourse (Bradshaw and Howard, 2018). Hacking of real social media accounts is a phenomenon, but the literature on the effectiveness of such tactics is limited. One can speculate that this is not very effective. As the validity of the messages are undermined as soon as the hacking is publicly known. Both the fake message(s) and hacking can be characterized as novel and newsworthy information, and as such should propagate with the same efficiency.

Creating fake accounts are the primary tactics used to manipulate social media. These accounts can be automated, human, or use a hybrid approach. The human-controlled fake accounts are primarily used to post a novel message. These are in turn amplified by bots to game the social media algorithms. Typically, through re-posting and re-tweeting, boost follower count, as well as posting keywords and hashtags to create a false sense of popularity – so-called astroturfing.

Bots can also be used to game other functionality in the social media platforms, such as reporting legitimate users. These user accounts are often suspended, or their content removed pending human review. They are therefore effectively blocked from the conversation and opportunity to present counterviews (Bradshaw and Howard, 2018).

There is a hybrid approach where the fake account is operated by humans and automated. This has several benefits, where the automation gives speed and volume, and the human supplies features in the content that is indistinguishable from legitimate messaging. This approach has been detected, but to a minor degree compared with full automation or human operation. The researchers speculate that this might be because hybrid accounts are harder to detect (Bradshaw and Howard, 2018).

There has been an increase in the use of chat and private messaging applications to target users of social media platforms (Bradshaw and Howard, 2018). This method is largely carried out by cyber troops contacting legitimate users directly and should therefore be harder to detect by ML, which is often trained in automated account pattern recognition.

7.5 Overview of technical solutions for detection and mitigation

Detecting cyber propaganda on social media platforms entails detection of outright false information, propaganda or hyper-partisan messaging by features associated with the use of the platform. These features can be; social and cultural context, content propagation, text-, image- and user content.

The automated techniques used for detection (classification) of social media manipulation can largely be categorized as ML, a sub-field of artificial intelligence. This entails extracting user features into a labeled dataset, training one or more ML models with the labeled data, and then giving the model features that match the training data. These datasets can be homogenous, a single type of data, or heterogenous. Heterogenous datasets are derived from more than one social media platform.

Classification of social media content can to a varying degree be automated, as described in chapter 7.6 Limitations and challenges of today's approaches.

7.5.1 Technologies for detection and identification

Maseri et al. (2020) general impression during the process of article selection, were that:

> "Most of the mitigation efforts use data mining, machine learning and sentiment analysis as an approach to mitigate cyber propaganda" (Maseri et al., 2020).

The majority of research effort in the selected articles on mitigating cyber-propaganda (2014-2018), focused on three aspects. Detection, identification, and comprehension which deals with how rumors happen and propagate in social networks.

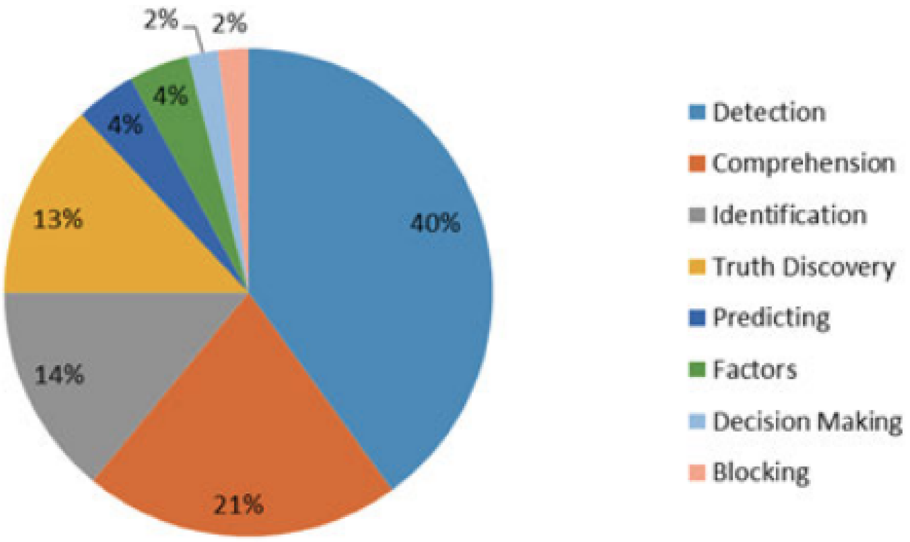


Figure 7.1: Composition of Effort in Mitigating Cyber-propaganda (Maseri et al., 2020).

Automatic detection is based on the extraction of features and subsequent application of ML models to classify the data. Due to the somewhat complex nature and technical knowledge needed to work with ML, this paper will not provide an in-depth description of the underlying concepts.

Social media platforms perform automatic detection of CIB by giving machine learning detection algorithms access to full datasets in real-time. There is little or no public information about the structure and capabilities of deployed systems.

Features

Content features are extracted and combined into datasets. Some of these have been made available to researchers by several social media platforms. A few are open for everyone to download (Maseri et al., 2020). Open datasets usually do not contain regular user data or content.

User features typically refer to user profile content and profile metadata such as profile pictures, age of account, likes, behavior, associated networks, and patterns of use.

Text & visual content are features associated with what the user links to, the credibility of sources, diversity, visual profile, and the veracity of the content.

Social & cultural context are features that are associated with larger parts, or the whole of a social media platforms' content. The cultural context can also reside outside the platform or data available in the set.

Propagation features look at how information is propagated through the network, typically patterns of re-tweets and shares, patterns in interactions with other users, and amount of followers.

Classification

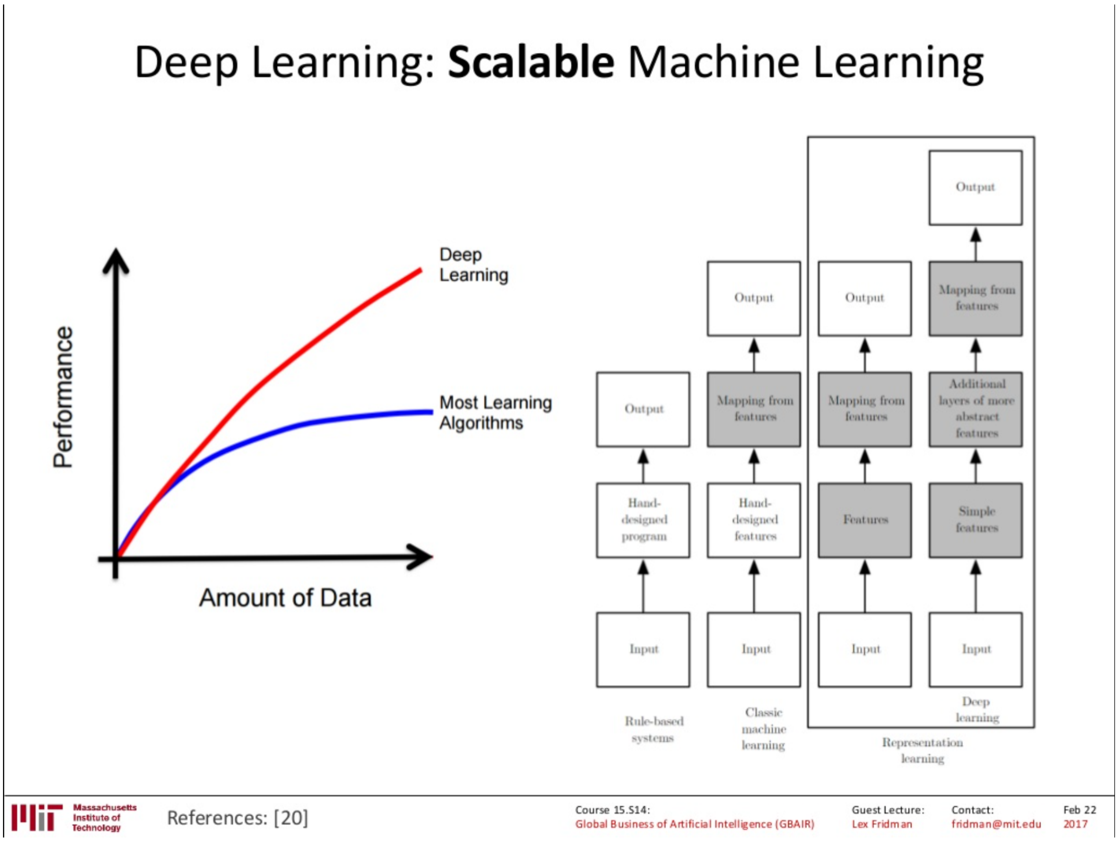
> "Machine learning is concerned with building systems that improve their performance on a task when given examples of ideal performance on the task, or improve their performance with repeated experience on the task" (Bringsjord and Govindarajulu, 2020).

There are three types of AI – logicist, probabilistic and neurocomputational. Probabilistic and neurocomputational ML have been applied to social media content classification in the studies evaluated by Maseri et al. (2020). The automatic classification of data is accomplished using ML. Several techniques are often combined for higher accuracy and\or comparison of performance in studies. ML allows the computer to break down complex tasks into several subtasks, learn from observational data and solve the problem.

Machine learning algorithms can be trained by supervised, unsupervised, and reinforcement learning. Both the probabilistic and neurocomputational algorithms identified Maseri et al. (2020) utilized supervised learning. Supervised learning requires labeled datasets to train the ML algorithm. Labeled datasets are data where features has been identified by humans, and the correct answer for a given input have been manually classified by humans. Learning consists of the algorithms making a classification from a given input, comparing this classification to the correct answer given in the training dataset and adjusting accordingly. Traditional machine learning algorithms learn in a supervised way. SVM, Adaboost\XGboost, and DT\RF were the most common traditional ML-algorithms applied to cyber propaganda detection in the studies identified by Maseri et al. (2020). Compared to deep neural networks, traditional ML typically require less data to achieve a reasonable level of performance. Some types of neural networks (GAN) can be trained in a semi- or unsupervised manner using unstructured data.

Training deep neural networks is achieved through the use of the backpropagation algorithm. Each node in a neural network is associated with a layer of nodes (neurons), an associated weight and an activation threshold. Layers are typically assigned a sub-task to solve. The individual node weights are adjusted by mapping inputs to outputs, such that the prediction is optimized for the expected classification for a given input in the labeled training dataset. The backpropagation algorithm optimizes by revealing the error rate associated with each neuron and adjusting for optimization.

Neural networks are good at generalizing through inferences of relationships in data. This is not dependent on the researcher even recognizing or optimizing for these in advance. The promise of neural networks are in the more complex classification tasks where many sub-tasks are required to achieve excellent performance. Deep neural networks solve this problem by implementing hidden layers for each sub-task. Deep neural networks excel at classifying unstructured data, such as images (video), sound and text to a degree. It typically requires more data to improve accuracy compared to traditional ML, but with large datasets it can achieve higher classification performance (Kavlakoglu, 2020).



Deep Learning: **Scalable** Machine Learning

Massachusetts Institute of Technology References: [20] Course 15.S14: Global Business of Artificial Intelligence (GBAIR) Guest Lecture: Lex Fridman Contact: fridman@mit.edu Feb 22 2017

Figure 7.2: Deep learning: Scalable Machine Learning (Goodfellow et al., 2017).

Some neural networks have short or long term "memory" of previously processed data (RNN) – which can be useful in tasks such as natural language processing and sound classification. Evaluation of ML performance at a given task is achieved by a set of performance metrics.

Recent advances in picture recognition and semantics have also spurred research into the detection of CIB by classifying profile pictures (Maseri et al., 2020). Some features in the text have until recently been extracted manually, imposing a human limitation. Especially the credibility of linked content, the social context and veracity of text have been hard problems for machine learning to automatically classify with good performance. Recent advances in the use of deep neural networks have successfully been applied to classify complex features such as social context and pattern recognition in word placement (Ma et al., 2016). Jin et al. (2017), also showed promising performance results in rumor detection by use of recurrent neural networks (RNN\at-RNN) (Jin et al., 2017). They examined text, visual, and context-based features.

7.5.2 Avenues for mitigation

Permanent detection and mitigation of CIB, especially human-produced CIB, in social media seem like a utopian goal.

Of curious note and importance for understanding the problem at hand is a finding by MIT researchers

Aral et al. (2018) published in Science (Aral et al., 2018). They show that false stories spread faster – by up to six times, reach more people and are shared by as much as an order of magnitude more than true stories. Another finding is that false political news reached people at a rate of 3 times faster, compared to other false news (rumors). The authors speculate that there are emotional and novelty factors at play, where sharing and receiving new information is inherently desirable by humans.

The systematic literature mapping by Maseri et al. (2020) identified only two studies (2%) in their selection of studies from 2014-2018 that looked at automatic blocking of content. A single study looked at blocking propagation of rumors in networks by introducing time delay to posts classified as rumors, without diminishing the value of network effects for legitimate users (Wang et al., 2017).

Traditional techniques for blocking malicious user activity on social platforms typically target IP addresses or soft, level 1 identification of users in the signup process. These can effectively be faked or masked by techniques such as the use of VPN, TOR network, and anonymous email providers.

Twitter uses algorithms to skew the visibility of accounts that tend to generate reports of abuse (Serving healthy conversation, 2018). Although this has some effect, is unlikely to greatly impact CIB. They have implemented policies to limit the spread of hacked materials as well (Distribution of hacked materials policy, 2020), which were recently adjusted (Brodkin, 2020). Some platforms use shadow banning or limiting the use of accounts to reduce trolling and other malicious or un-civic behavior. Twitter does not shadow ban (Setting the record straight on shadow banning, 2018), but use limiting of accounts when outright violations of policy are detected. Facebook was granted a patent in 2019 (filed in 2015) for automatic shadow banning of users (Kanter et al., 2019). They have not shared information about the implementation of such techniques.

Another possible solution to the CIB problem can be stronger identification of users at account creation. This has obvious limitations, where some countries and demographics are limited in their ability to provide strong identification, especially online. Further, the strong identification of users can have free speech implications in non-democratic or flawed democracies. Even if the user handle is not required to be a real name, the storage of the identity in the service, in case of compromise, can have serious ramifications for users in totalitarian or authoritarian countries, thus leading to a chilling effect on free speech. One of the few still proposed redeeming qualities of social media.

Some social media platforms do require the use of real names, like Facebook (What names are allowed on Facebook?, n.d.)). This is enforced by attempts at algorithmic detection on the name itself, leading to various controversies (Wikipedia Contributors, 2019). Twitter, however, does not require any identification at signup (*Guidelines for law enforcement*, n.d.).

Flagging of user content identified as untrue or misleading by human fact-checkers is another technique that has been introduced. Twitter recently marked a post by the US president (Isaac and Kang, 2020) as a violation of their election integrity policy (*Civic integrity policy*, n.d.). Flagging of content gives users the possibility to easily fact-check statements, but this might not have the desired effect due to human psychological limitations in interpreting facts that contradict our pre-existing beliefs (Washburn and Skitka, 2017).

Twitter and Facebook implemented outright blocking of posts linking to external content that had not been independently fact-checked or corroborated by other credible news media during the runup to the US 2020 election. This content was probably flagged by human moderators (Senate Republicans, 2020), (Andy Stone, 2020).

An automated classification of truthfulness in user-generated content is still far from current capabilities in ML, as we will discuss in the following chapter.

7.6 Limitations and challenges of today's approaches

Social media platforms are not open about their deployed classification techniques for potential CIB. Some platforms give updates on their successes and take downs when combating this problem. Relatively recent and prominent examples of CIB on Twitter and Facebook can be interpreted as these platforms losing the battle against CIB and election manipulation on two fronts. Both advertisement firms leveraging data analytics and cyber troops have caused several large scandals (Graham-Harrison and Cadwalladr, 2018), (*Committee sensitive*, 2019), (Varol et al., 2017).

Currently, there is only one systematic mapping (Maseri et al., 2020) conducted in this area of study. There are no systematic literature reviews available. The systematic literature mapping by Maseri et al. (2020) problematizes the mixed type of dataset applied, ML classification models and performance metrics when evaluating prediction strength. Some studies did not reveal which social media platform their datasets were from. This makes comparison and review of efficacy of different approaches difficult. Further, most of the selected research (2014-2018) was carried out on datasets from Twitter and Facebook. This can introduce bias in our current understanding of the issues. There is however a recommendation for future use of performance metrics that are resistant to inherent differences in homogeneous and heterogeneous datasets, where data from different services are mixed.

7.6.1 ML & AI – determining truth and social context

From a cursory understanding of the classification techniques researched and employed to date, user, text (partly) and propagation features seem the easiest to classify in the near term. Effective automatic classification of social, cultural context and truth might be hard or impossible using ML in the near to medium term.

Truth discovery is the problem of detecting "true" facts from multiple conflicting sources. Today's algorithms try to solve this problem by examining the source(s) of the news, their credibility and conflicting views (Shu et al., 2017). There are several obvious limitations to this approach. The method requires several credible sources to validate "truth". The factual claims themselves are not evaluated. The news item must be reported widely, making novelty news items potential casualties. This approach might diminish the value proposition of the platform for users. Having a potential impact on the business of the social media platforms parent company.

Determining factual truth is a huge problem, even for humans (Washburn and Skitka, 2017). Especially when classifying political claims that might be novel. It depends on several factors such as social context, culture, sources, and individual bias - both recognized and unconscious. Humans use extrinsic information that an AI is typically not presented with or trained on. Maseri et al. (2020) did not identify any studies contributing to advancements in techniques for truth discovery. There is also a potential bias in the labeled dataset that the AI is trained on. The algorithms lack a deeper understanding of undesirable biases that might be present in the dataset (Mehrabi et al., 2019). The classifier might therefore reproduce unwanted biases(Dastin, 2018), (Obermeyer et al., 2019). Unlabeled datasets can also contain biases.

AI in its current, and foreseeable future cannot understand the concept, nor determine factual truth. This problem might require strong AI with heterogeneous input data from a wide set of sources. We currently have weak AI that is trained on narrow problems and are successful to some degree.

7.6.2 The bottom line

Many of today's techniques for combating CIB require human intervention. Especially when classifying social context and factual truth. This in turn requires large budgets and personnel. It is difficult to assess the willingness to make these investments by the social media companies. It might be argued successfully that some companies have had ample financial earnings to scale up investment in both human and automated moderators (Facebook, 2020*b*), (Alphabet, 2020).

Another aspect is the initial unwillingness of social media companies to moderate, much less be arbiters of truth. This can partly be explained by immunity afforded by US laws. Many of the biggest social media companies are based in the US and therefore protected from liability from third party content on their platforms. This has hinged on a controversial piece of internet legislation commonly known as section 230 of the Communications Decency Act of 1996. The platforms' increase in both human and automated moderation of content has been followed by claims of bias (NW et al., 2020), (Denham, 2019), (Senate Republicans, 2020). US government officials and US lawmakers are now considering changes that would limit social media company liability protection significantly (Office of Chairman Pai, 2020). Potential changes in regulation, algorithmic- and business model transparency has a very real possibility of curtailing earnings and increasing liability for social media companies.

7.7 Evolution - A look at a likely future

The use of cyber troops, and their associated budgets, are expanding in many nations throughout the world (Bradshaw and Howard, 2018). It is reasonable to assume that an expansion of automated cyber capabilities will follow. It will become harder to determine illegitimate use through technical means when the adversary increases the use of humans and hybrid accounts. Particularly the use of hybrid accounts is emphasized by Bradshaw and Howard (2018) as especially effective. This technique gives the volume and speed of a bot and the innocuousness of a real human that can confuse classification algorithms.

The literature mapping (Maseri et al., 2020) show that traditional ML algorithms have been favored over neural networks for automatic detection of CIB. This probably has historical reasons, where traditional machine-learning algorithms have been in use over a longer period. The utilization of different types of deep neural networks to classify complex sets of features will probably increase. Different deep neural networks can also be used in cascade to tackle very complex classification problems, where one neural network would be too deep-, or the available datasets too small to effectively train a single model.

There is a general increase in both public and private datasets in a wide variety of fields. This bodes well for training deep neural networks that require large training sets to effectively generalize classification problems. Some data types, if there is a limited amount of available data, can benefit from data augmentation techniques. GANs can also be used to increase the size of datasets, where a generator neural network creates new artificial data based on a smaller labeled dataset, and a discriminator neural network that classifies the output of the generator combined with real data from the dataset. GANs were proposed in 2014 (Goodfellow et al., 2014) and can supplement both supervised and semi-supervised, as well as generate unsupervised training data. Their use is likely to increase, as they have useful properties in several areas (Salimans et al., 2016).

It is reasonable to expect that recent and future advances in ML will be utilized both by the social media companies for detection and mitigation (Facebook, 2017*b*), (Facebook, 2017*a*). Their adversaries are likely to increase utilization of ML in their pursuit of SMME. Text, voice, and video can now be created or manipulated by ML.

7.7.1 Artificially generated text

Cyber troops are increasing their use of direct messaging (Bradshaw and Howard, 2018). Soon, more advanced adversaries are likely to utilize recent breakthroughs in ML generated text for both mass private messaging of users and posting of public content.

Real-world examples of such attacks can be seen in the cruder bot and fake text attacks against the US public comment system associated with a net neutrality regulation proposal in 2017. Some 22 million comments were received. 99,7% of these were later classified as fake(Singel, 2018).

There were months-long internal deliberation and doubts in the OpenAI foundation (Hern, 2019) before they grudgingly released the full version of the GPT2 model for ML-based text generation. Among their chief concerns were:

> "[...] generating misleading news articles, impersonating others online, automating the production of abusive or fake content for social media, and automating the creation of spam and phishing content" (Tung, 2019).

A modified version of the limited GPT2 model was used to submit fake public comments to the Idaho Medicaid Reform Waiver proposal. These comments were then evaluated by humans in a Turing test. Comments were correctly classified approximately half the time (Weiss, 2019). Similarly, FireEye achieved very believable fake text results by re-purposing social media posts linked to Russia's IRA troll factory as input to the GPT-2 model (Tully and Foster, 2020). Google recently made serious breakthroughs in natural language processing with the BERT technique (Devlin et al., 2019).

This shows that a simplified model of easily re-purposed, publicly available technology can pass the Turing test. The pace of progress makes text a likely candidate for near-future weaponization by cyber troops and other actors.

7.7.2 Artificially generated images, voice, and video

Artificial neural net-based creation or manipulation of still images and video for mimicry of real people, so-called "deepfakes" entered mainstream public consciousness in 2017 (USA Today, 2019). Manipulation of video content with very believable results – both picture and audio – is now well within the abilities of AI researchers all over the world. Recent breakthroughs are so realistic that it is hard to distinguish between real and fake content (Naruniec et al., 2020). The synthetic creation of portrait photos at thispersondoesnotexist.com utilizes a GAN with an output that is in most cases indistinguishable from real humans(*This Person Does Not Exist*, 2018).

The technology and subsequent democratization of it have made serious advances in the last three years. It is now even utilized by advertisers. Notable examples are a video ad for the 2020 Netflix documentary "The last dance" where a 90's sports anchor Kenny Mayne makes surprisingly accurate predictions about the future documentary and how it's going to be "lit" (Insurance, 2020), although we "[...] don't know what that means yet".

FireEye showed that hobbyists can use publicly available pre-trained neural networks for generation of fairly believable text, image and voice (Tully and Foster, 2020). Youtube users are now significantly improving on professional FX-studio work (Shamook, 2020). Whole head and body swaps with mannerisms are being explored as the next steps by researchers.

The use of deepfakes for political purposes in social media have already begun. US President Donald Trump's retweet of a Joe Biden deepfake video in April 2020 can be interpreted as a harbinger of the future (Frum, 2020).

The arms race to detect deep fakes by ML has started. Facebook arranged the "Deepfake detection challenge" in 2020. So far, the results for classification performance are depressing. The best model in this contest achieved a 65% detection rate on novel real-world examples that were not in the training dataset(Facebook, 2020*a*). Microsoft recently released their "Newsguard Video Authenticator" for detection of artificial manipulation in images and video (Microsoft, 2020), as a bid to combat deepfakes in the 2020 US election.

Neural network-based classifiers suffer from a robustness problem. Adversarial inputs can effectively fool detection models. This can be achieved by introducing noise in input or training data to confuse ML pattern recognition. Adversaries with access to training datasets and knowledge of the specific model can also perform poisoning attacks, where fake samples are introduced to lessen model accuracy. These attacks have also been successfully applied to some traditional ML algorithms such as SVM. Another type of attack is the evasion type, where input data is modified in a way that is often invisible to humans and makes the classifier mispredict (Szegedy et al., 2014), (Nguyen et al., 2015).

Worse, a model trained for generating adversarial inputs in a specific neural network model seems to have properties that make them effective on other unseen classification models of the same type – so-called "transferability". Black-box attacks are therefore effective and achievable (Xu et al., 2020). Adversarial attacks are known in both text, audio and video domains. There is currently a cat and mouse game, where new successive attacks are detected, and new attacks to defeat these are devised. There is no panacea for mitigating these attacks. Currently, known defenses probably reduce the classifier accuracy on real data (Xu et al., 2020).

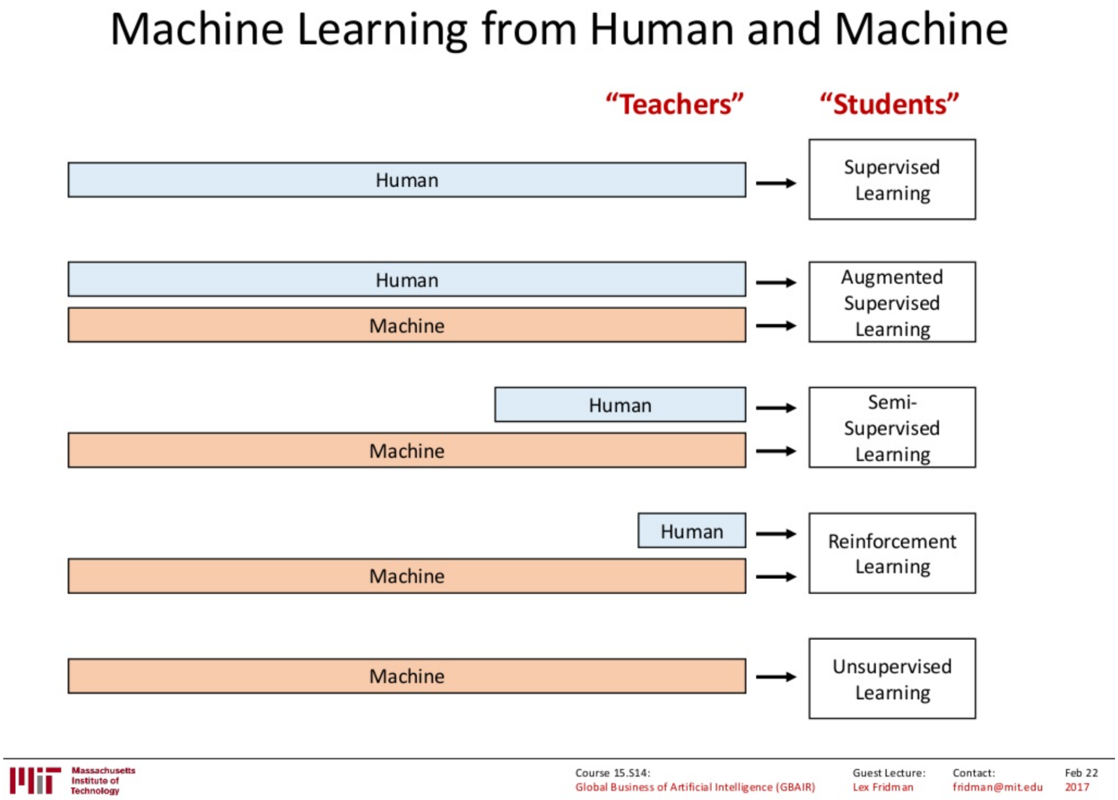
Video, audio, and images can be post-processed to remove telltale signs, confusing both classification algorithms and humans. Some researchers believe that ML algorithms for generating fake videos will become too good to effectively detect using ML and that other novel techniques for authentication of content is needed (Agarwal et al., 2019), (Microsoft, 2020). Adobe, Twitter, and The New York Times recently revealed their "Content Authenticity Initiative" to provide content attribution (Adobe Communications Team, 2019).

The asymmetric nature of deepfake attacks, when applied in SMME, makes the damage potential high. A single viral video on the morning of an election could influence the outcome, or at the very least cause chaos and confusion. Possibly inflicting permanent long term damage to public trust in institutions.

Currently, the generative models have a clear upper hand compared to automated detection models in the video and audio domains. The best detection model will correctly classify 2/3 of deepfake videos. With human discrimination, most videos can still be correctly classified, but even today it can be hard to tell on smaller screens. Synthetic profile pictures are now believable enough to pass human discrimination in most cases. The rate of creation for short text by ML algorithms far outpace human capacity to evaluate and has passed the Turing test. It can therefore be argued that we have already lost the ability to effectively distinguish between real and fake short text.

7.7.3 Removing some of the human limitations in machine learning

Researchers are hopeful that future breakthroughs in machine learning will yield further advances in semi-supervised, reinforcement, and unsupervised learning. This might alleviate some of the human limitation in machine learning, where creating large labeled datasets have a high human cost.



Machine Learning from Human and Machine

"Teachers" "Students"

Human → Supervised Learning

Human → Machine → Augmented Supervised Learning

Human → Machine → Semi-Supervised Learning

Human → Machine → Reinforcement Learning

Machine → Unsupervised Learning

Massachusetts Institute of Technology

Course 15.S14: Global Business of Artificial Intelligence (GBAIR)

Guest Lecture: Lex Fridman

Contact: fridman@mit.edu

Feb 22 2017

Figure 7.3: Machine Learning from Human and Machine (Fridman, 2017).

GANs show promise in improving classification performance in supervised- and semi-supervised learning today. Neural networks are inefficient learners that require large and varied datasets to improve and generalize performance. GANs can provide this through generating artificial data input by introducing small variations in the real dataset, thereby expanding the total dataset and its diversity (Salimans et al., 2016). They have successfully been applied to achieve previously unseen levels of realism in high-resolution face-swapping models (Naruniec et al., 2020).

The observant reader might already have concluded from the previous example that this type of neural network can be used to improve synthetic media realism, and for creating adversarial content that will fool the detection classifiers. Promisingly, GAN's have been utilized to improve the detection of adversarial inputs (Xu et al., 2020), which can be very useful for detecting synthetic media with adversarial content in social media.

7.7.4 PR companies, big data, and governments

Bradshaw and Howard (2018) identify that the use of PR companies and big data analytics are likely to increase (Bradshaw and Howard, 2018). They are poised for an even more central role in future elections, including illegitimate companies that employ automatic CIB.

Very few governments have proposed and enacted laws to increase transparency and outlaw CIB, deepfakes and the use of political bots (Bradshaw et al., 2019). Only the most high profile scandals have led to some degree of voluntary transparency regarding user data collection and sharing by social media companies (Facebook, 2018). Researchers at Oxford University have called on politicians to enact laws requiring higher levels of cooperation and transparency from social media companies (Bradshaw et al., 2019).

Russian Cyber troops have started to outsource their account management to human third parties, that are blind to the real objectives. This is exemplified in operation "Double Deceit" where Ghanese employees were handed fake IRA content and duped to perform CIB targeting black communities in the US (Nimmo et al., 2020). China and other actors have not been as prolific or successful as Russian operators in their automation, believability, or persistence in this space (Twitter, 2020). This can have historical explanations, where Russia has decades of experience in disinformation campaigns (University, 2014). There is little doubt however that other, very technically competent nations like China are playing catch-up and experimenting with CIB for information wars – seen recently in their targeting of Hong Kong residents (Twitter, 2020).

7.8 Conclusion

For many years there has existed a cybersecurity asymmetry between defenders and attackers. Defending social media platforms from CIB performed for SMME is no exception. The causes are manifold.

The old cybersecurity trope that "defenders must be right every time, attackers only have to be right once", is as true here as elsewhere. A knockout punch is not needed to inflict permanent damage to democratic institutions. The most potent current technical defensive capabilities have side effects that can be harmful to free speech and are often politically infeasible for democracies to implement.

This asymmetry can also be spotted in the lightning-fast advance of technical capabilities compared to traditional political processes, which often take years. Several crucial technical areas like machine learning based classification of truth, which is an important problem today, and effective classification of deepfakes, for tomorrow, are simply out of automated technical reach. Generating believable fake text, video, and audio increasingly is.

Russia has a long track record of CIB for SMME, which includes the use of large botnets. They (and others) have recently displayed a change in tactics, where they outsource to third parties, increase the use of human operators and directly message users. Possibly evading existing technical detection measures. Other nations are trying to replicate their tactics and use of technology for similar purposes. Worryingly, nations that possess advanced cyber capabilities, like China, have recently shown an increased interest in this domain.

The political class in democratic countries have only just begun to enact new regulations in areas such as political campaigns, use of political advertisement companies, social media transparency, CIB and automated computational propaganda. But they are waking up.

The 2020 US presidential election is nigh. A curse often wrongly attributed to the Chinese is "May you live in interesting times". Even if detection of SMME is achieved, attribution and effective automatic mitigation within a usable time frame is not technically feasible. At this time, we are left with post-mortem analysis of election interference and possible skew in election outcomes.

Bibliography

Adobe Communications Team (2019), 'Introducing the content authenticity initiative'. Accessed 01.11.2020.
URL: *https://blog.adobe.com/en/publish/2019/11/04/content-authenticity-initiative.html*

Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K. and Li, H. (2019), 'Protecting world leaders against deep fakes'. Accessed 01.11.2020.
URL: *http://www.hao-li.com/publications/papers/cvpr2019workshopsPWLADF.pdf*

Alphabet (2020), 'Alphabet announces second quarter 2020 results'. Accessed 01.11.2020.
URL: *https://abc.xyz/investor/static/pdf/2020Q2_alphabet_earnings_release.pdf*

Andy Stone (2020), 'https://twitter.com/andymstone/status/1316423671314026496'. Accessed 01.11.2020.
URL: *https://twitter.com/andymstone/status/1316423671314026496*

Aral, S., Vosoughi, S. and Roy, D. (2018), 'The spread of true and false news online', *Science, DOI: 10.1126/science.aap9559* **359**, 1146–1151. Accessed 01.11.2020.
URL: *DOI: 10.1126/ science.aap9559*

Bradshaw, S. and Howard, P. (2018), 'Challenging truth and trust: A global inventory of organized social media manipulation'. Accessed 01.11.2020.
URL: *http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf*

Bradshaw, S., Neudert, L.-M. and Howard, P. (2019), 'Countering the malicious use of social media government responses to malicious use of social media'.
URL: *https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Nato-Report.pdf*

Bringsjord, S. and Govindarajulu, N. S. (2020), 'Artificial intelligence'. Accessed 01.11.2020.
URL: *https://plato.stanford.edu/entries/artificial-intelligence/#BlooMachLear*

Brodkin, J. (2020), 'Twitter abruptly changes hacked-materials policy after blocking biden story'. Accessed 01.11.2020.
URL: *https://arstechnica.com/tech-policy/2020/10/twitter-lifts-ban-on-sharing-hacked-materials-after-blocking-ny-post-story/*

Calo, R., Howard, P. and Woolley, S. (2018), 'Algorithms, bots, and political communication in the us 2016 election: The challenge of automated political communication for election law and administration', *Journal of Information Technology & Politics* **15**, 81–93.

Civic integrity policy (n.d.). Accessed 01.11.2020.
URL: *https://help.twitter.com/en/rules-and-policies/election-integrity-policy*

Committee sensitive (2019).
URL: *https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf*

Dastin, J. (2018), 'Amazon scraps secret ai recruiting tool that showed bias against women'.
URL: *https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G*

Denham, H. (2019), 'Trump accuses google of anti-conservative bias without providing evidence', *Washington Post* . Accessed 01.11.2020.
URL: *https://www.washingtonpost.com/technology/2019/08/06/trump-accuses-google-anti-conservative-bias-without-providing-evidence/*

Devlin, J., Chang, M.-W., Lee, K., Google, K. and Language, A. (2019), 'Bert: Pre-training of deep bidirectional transformers for language understanding'. Accessed 01.11.2020.
URL: *https://arxiv.org/pdf/1810.04805.pdf*

Distribution of hacked materials policy (2020), 'Distribution of hacked materials policy'. Accessed 01.11.2020.
URL: *https://help.twitter.com/en/rules-and-policies/hacked-materials*

Facebook (2017*a*), 'Getting our community help in real time'. Accessed 01.11.2020.
URL: *https://about.fb.com/news/2017/11/getting-our-community-help-in-real-time/*

Facebook (2017*b*), 'Hard questions: Are we winning the war on terrorism online?'.
URL: *https://about.fb.com/news/2017/11/hard-questions-are-we-winning-the-war-on-terrorism-online/*

Facebook (2018), 'An update on our plans to restrict data access on facebook'. Accessed 01.11.2020.
URL: *https://about.fb.com/news/2018/04/restricting-data-access/*

Facebook (2020*a*), 'Deepfake detection challenge dataset'. Accessed 01.11.2020.
URL: *https://ai.facebook.com/datasets/dfdc/*

Facebook (2020*b*), 'Facebook reports second quarter 2020 results'.
URL: *https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Second-Quarter-2020-Results/default.aspx*

Freedom House (2019), 'Freedom on the net 2019: The crisis of social media'. Accessed 01.11.2020.
URL: *https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf*

Fridman, L. (2017), 'Mit sloan: Intro to machine learning'. Accessed 01.11.2020.
URL: *https://www.slideshare.net/lexfridman/mit-sloan-intro-to-machine-learning-in-360vr*

Frum, D. (2020), 'The very real threat of trump's deepfake'. Accessed 01.11.2020.
URL: *https://www.theatlantic.com/ideas/archive/2020/04/trumps-first-deepfake/610750/*

Gleicher, N. (2018), 'Coordinated inauthentic behavior explained'. (Accessed: 01.09.2020).
URL: *https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/*

Goodfellow, I., Courville, Y., Courville, A. and Courville, B. (2017), '"deep learning." an mit press book in preparation'. Accessed 01.11.2020.
URL: *http://www.deeplearningbook.org/*

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014), 'Generative adversarial nets'. Accessed 01.11.2020.
URL: *https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf*

Graham-Harrison, E. and Cadwalladr, C. (2018), 'Revealed: 50 million facebook profiles harvested for cambridge analytica in major data breach'.
URL: *https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election*

Guidelines for law enforcement (n.d.). Accessed 01.11.2020.
URL: *https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support#:~:text=Twitter%20doesn*

Hern, A. (2019), 'New ai fake text generator may be too dangerous to release, say creators'. Accessed 01.11.2020.
URL: *https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction*

Insurance, S. F. (2020), 'Predictions | state farm® + espn commercial (featuring kenny mayne)'. Accessed 01.11.2020.
URL: *https://www.youtube.com/watch?v=FzOVqClci_s*

Isaac, M. and Kang, C. (2020), 'While twitter confronts trump, zuckerberg keeps facebook out of it', *The New York Times* . Accessed 01.11.2020.
URL: *https://www.nytimes.com/2020/05/29/technology/twitter-facebook-zuckerberg-trump.html*

Jin, Z., Cao, J., Guo, H., Zhang, Y. and Luo, J. (2017), 'Multimodal fusion with recurrent neural networks for rumor detection on microblogs', *Proceedings of the 2017 ACM on Multimedia Conference - MM '17* . Accessed 01.11.2020.

Kanter, J. A., Singh, M. and Muriello, D. G. (2019), 'United states patent: 10356024 - moderating content in an online forum'. Accessed 01.11.2020.
URL: *http://patft.uspto.gov/netacgi/nph-Parser?Sect2=PTO1&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&d=PALL&RefSrch=yes&Query=PN/10356024*

Kavlakoglu, E. (2020), 'Ai vs. machine learning vs. deep learning vs. neural networks: What's the difference?'. Accessed 01.11.2020.
URL: *https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks*

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B., Wong, K.-F. and Cha, M. (2016), 'Detecting rumors from microblogs with recurrent neural networks'. Accessed 01.11.2020.
URL: *https://www.ijcai.org/Proceedings/16/Papers/537.pdf*

Maseri, A. N., Norman, A. A., Eke, C. I., Ahmad, A. and Molok, N. N. A. (2020), 'Socio-Technical Mitigation Effort to Combat Cyber Propaganda: A Systematic Literature Mapping', *IEEE Access* **8**, 92929–92944. Accessed 17.11.2020.
URL: *https://ieeexplore.ieee.org/document/9093872/*

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2019), 'A survey on bias and fairness in machine learning'. Accessed 01.11.2020.
URL: *https://arxiv.org/pdf/1908.09635.pdf*

Microsoft (2020), 'New steps to combat disinformation'. Accessed 15.09.2020.
URL: *https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/*

Naruniec, J., Helminger, L., Schroers, C. and Weber, R. (2020), 'High-resolution neural face swapping for visual effects', *Computer Graphics Forum* **39**, 173–184. Accessed 01.11.2020.
URL: *https://s3.amazonaws.com/disney-research-data/wp-content/uploads/2020/06/18013325/High-Resolution-Neural-Face-Swapping-for-Visual-Effects.pdf*

Nguyen, A., Yosinski, J. and Clune, J. (2015), 'Deep neural networks are easily fooled: High confidence predictions for unrecognizable images'. Accessed 01.11.2020.
URL: *https://arxiv.org/pdf/1412.1897.pdf*

Nimmo, B., François, C., Eib, S., Ronzaud, L., Smith, M., Lederer, T. and Carter, J. (2020), 'Ira in ghana: Double deceit'. Accessed 01.11.2020.
URL: *https://public-assets.graphika.com/reports/graphika_report_ira_in_ghana_double_deceit.pdf*

NW, . L. S., 800Washington, S. and Inquiries, D. U.-.-. . M.-.-. . F.-.-. . M. (2020), 'Most americans think social media sites censor political viewpoints'. Accessed 01.11.2020.
URL: *https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints/*

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019), 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science* **366**, 447–453. Accessed 01.11.2020.
URL: *https://science.sciencemag.org/content/366/6464/447.full*

Office of Chairman Pai (2020), 'Statement of chairman pai on section 230'.
URL: *https://docs.fcc.gov/public/attachments/DOC-367567A1.pdf*

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A. and Chen, X. (2016), 'Improved techniques for training gans'. Accessed 01.11.2020.
URL: *https://arxiv.org/pdf/1606.03498*

Senate Republicans (2020), 'https://twitter.com/senategop/status/1316495807286304771'.
URL: *https://twitter.com/SenateGOP/status/1316495807286304771*

Serving healthy conversation (2018), 'Serving healthy conversation'. Accessed 01.11.2020.
URL: *https://blog.twitter.com/official/en_us/topics/product/2018/Serving_Healthy_Conversation.html*

Setting the record straight on shadow banning (2018), 'Setting the record straight on shadow banning'. Accessed 01.11.2020.
URL: *https://blog.twitter.com/official/en_us/topics/company/2018/Setting-the-record-straight-on-shadow-banning.html*

Shamook (2020), 'De-aging robert deniro in the irishman [deepfake]'. Accessed 01.11.2020.
URL: *https://www.youtube.com/watch?v=dHSTWepkp_M&ab_channel=Shamook*

Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. (2017), 'Fake news detection on social media', *ACM SIGKDD Explorations Newsletter* **19**, 22–36.
URL: *https://dl.acm.org/citation.cfm?id=3137600*

Singel, R. (2018), 'Filtering out the bots: What americans actually told the fcc about net neutrality repeal'. Accessed 01.11.2020.
URL: *http://cyberlaw.stanford.edu/files/blogs/FilteringOutTheBotsUnique2017NetNeutralityComments1024Update.pdf*

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. and Fergus, R. (2014),

‘Intriguing properties of neural networks’. Accessed 01.11.2020.
URL: *https://arxiv.org/pdf/1312.6199.pdf*

This Person Does Not Exist (2018). Accessed 01.11.2020.
URL: *https://thispersondoesnotexist.com/*

Tully, P. and Foster, L. (2020), ‘Repurposing neural networks to generate synthetic media for information operations’.
URL: *https://www.fireeye.com/blog/threat-research/2020/08/repurposing-neural-networks-to-generate-synthetic-media-for-information-operations.html*

Tung, L. (2019), ‘Openai’s ’dangerous’ ai text generator is out: People find gpt-2’s words ’convincing”. Accessed 01.11.2020.
URL: *https://www.zdnet.com/article/openais-dangerous-ai-text-generator-is-out-people-find-gpt-2s-words-convincing/*

Twitter (2020), ‘Disclosing networks of state-linked information operations we’ve removed’. Accessed 01.11.2020.
URL: *https://blog.twitter.com/en_us/topics/company/2020/information-operations-june-2020.html*

University, C. (2014), ‘Mitrokhin’s kgb archive opens – churchill college’. Accessed 01.11.2020.
URL: *https://www.chu.cam.ac.uk/news/2014/jul/7/mitrokhins-kgb-archive-opens/*

USA Today (2019), ‘Tricked by the fake obama video? deepfake technology, explained | usa today’. Accessed 01.11.2020.
URL: *https://www.youtube.com/watch?v=EtEPE859w94*

Varol, O., Ferrara, E., Davis, C., Menczer, F. and Flammini, A. (2017), ‘Online human-bot interactions: Detection, estimation, and characterization’. Accessed 01.11.2020.
URL: *https://arxiv.org/pdf/1703.03107.pdf*

Wang, B., Chen, G., Fu, L., Song, L. and Wang, X. (2017), ‘Drimux: Dynamic rumor influence minimization with user experience in social networks’, *IEEE Transactions on Knowledge and Data Engineering* **29**, 2168–2181. Accessed 01.11.2020.

Washburn, A. N. and Skitka, L. J. (2017), ‘Science denial across the political divide: Liberals and conservatives are similarly motivated to deny attitude-inconsistent science’, *Social Psychological and Personality Science* **9**, 972–980. Accessed 01.11.2020.

Weiss, M. (2019), ‘Deepfake bot submissions to federal public comment websites cannot be distinguished from human submissions’, *Technology Science* . Accessed 01.11.2020.
URL: *https://techscience.org/a/2019121801/*

What names are allowed on Facebook? (n.d.), ‘What names are allowed on facebook? | facebook help center | facebook’. Accessed 01.11.2020.
URL: *https://www.facebook.com/help/112146705538576?helpref=uf_permalink*

Wikipedia Contributors (2019), ‘Facebook real-name policy controversy’. Accessed 01.11.2020.
URL: *https://en.wikipedia.org/wiki/Facebook_real-name_policy_controversy*

Xu, H., Ma, Y., Liu, H.-C., Deb, D., Liu, H., Tang, J.-L. and Jain, A. K. (2020), ‘Adversarial attacks and defenses in images, graphs and text: A review’, *International Journal of Automation and Computing* **17**, 151–178. Accessed 01.11.2020.

Chapter 8

The Russian Internet Research Agency's Interference in the 2016 US Election

By Ketil Østnor

8.1 Executive summary

The Russian government tried to manipulate the 2016 U.S Presidential election using Coordinated Inauthentic Behaviour (CIB) as a method to do so. The Russian government acted through The Internet Research Agency (IRA), also known as Troll factory, to spread Russian propaganda on social media platforms. The Russian government wanted to manipulate the election in favour of President Donald Trump. Special investigator Robert S. Muller, who led the official investigation into Russia's involvement in the 2016 US Presidential election, concluded that there was not enough information to state that Russia was successful in swaying the election. The paper discuss the tactics employed by the IRA, and answers the research question "what was IRA's most successful CIB-tactic when trying to sway the 2016 US Presidential election".

The IRA had several tactics within their CIB-operation. The paper explain some of the key tactics. The tactics discussed in the report are the use of ads, cross-platform branding, the spread of propaganda through memes and amplifying conspirational narratives.

The answer to the research question is that there is not enough information available to give a definitive answer. There are data that shows the propagation of IRA's tactics and one can only assume the effects it may have had. One thing is clear; IRA's operation had great planning and great execution. The combined effect of the tactics shows that they may have been successful in manipulation voters to vote for President Trump, third-party elective or not to vote at all. There is not enough information to say that one tactic was more successful than others were.

The paper shows how other nations may have motives to manipulate democratic election. The Russian operation are not the last of covert operation we will see. In the future, there will be attempts from several nations to manipulate democratic elections. Social media platforms must build systems to

identify attempts of Coordinated Inauthentic Behaviour. People need to be safe from manipulating when they are engaging in activities on social media. Not only for their personal liberty, but also to protect the democracy.

8.2 Introduction

The topic of this paper is Coordinated Inauthentic Behaviour. The paper is a case study of the Russian governments' involvement in the 2016 US Presidential election. Research on existing literature on the topic has provided information for this case study. The purpose of a case study is to develop a holistic knowledge of the event that is being studied (Wæhle et al., 2020, para. 4). The focus of this paper is on the use of Coordinated Inauthentic Behaviour for targeting US citizens on social media to manipulate public opinion and sow discord between different groups. This paper will explain the tactics employed by the IRA, the propagation of their tactics and the impact it may have had. Facebook, Twitter, Alphabet and other social media platforms have provided the data set in this case study. The United States Senate Select Committee on Intelligence (SSCI) and Special investigator Robert Muller used the same data set for their investigation into Russian involvement in the 2016 US presidential election. Their findings will be the basis of this case study.

The Russian covert operation into the 2016 Presidential election was complex. The Russian operation used other tactics than CIB, such as hacking E-mails of the members of the democratic party, leaking E-mails from Clinton to WikiLeaks and more. It was predominantly two methods that the Russian government used in their operation. One method was a hacking operation targeting the Democratic party and its employees, and another was their operation on social media, using CIB. The research question of this paper is "what was IRA's most successful CIB-tactic when trying to sway the 2016 US Presidential election?" To discuss which tactic was the most successful, the propagation and created engagement on their content are variables used for the discussion. A disclaimer, the official investigation concluded that there is not sufficient evidence to say that the Russian covert operation was successful in swaying the election. It can still be an excellent discussion to investigate which of their tactics was most successful. Information about foreign nations tactics can create awareness of how sophisticated it can be. Democratic nations must guard themselves against foreign involvement in their democratic processes.

The following sections will describe what the IRA is, and what Coordinated Inauthentic Behaviour is. Furthermore, the paper will describe some of the key tactics employed by the IRA to manipulate public opinion and sow discord between different groups of people. The IRA used many different tactics, but this paper will discuss some of the most important. The delimitation of the number of tactics explained in this paper is set to four key tactics. The use of advertisement, cross-platform branding, creating and spreading memes, and amplify conspirational narratives. The last section will be a discussion on which tactic was their best and how successful the IRA was in manipulating voters to change the election outcome.

8.3 Russian internet research Agency (IRA) and coordinated inauthentic behaviour (CIB)

8.3.1 The Russian internet research agency

The IRA is a Russian company based in St. Petersburg. They are known as the Russian Troll Farm or a Troll Factory. Russian intelligence is closely linked to IRA who engages in online influence operations on business and political issues (Internet Research Agency, 2019). The IRA started in mid-2013 and is run like a market agency. The IRA has trained thousands of employees to engage in influence operations. In the beginning, they targeting Russian and Ukraine Citizens. Now they are engaging in activities in many different countries. The IRA gets funding from Yevgeniy Viktorovich Prigozhin and the companies he controls, collectively known as Concord (Muller, 2019).

In the report from Office of the director of national intelligence - Assessing Russian Activities and Intentions in Recent US Elections, they concluded that the Russian president Vladimir Putin had ordered a campaign to influence the 2016 presidential election. They also concluded that they were in favour of president elect-Trump and aimed their campaign at discrediting Secretary Hillary Clinton and undermining the liberal democratic order (Exposing Russia's Effort to Sow Discord Online, 2017). The IRA used various social media platforms to discourage voter turnout. In some cases, the IRA tried to mislead people into texting their votes. The IRA also encouraged left-oriented Americans to vote for third-party candidates like Jill Stein or not to vote at all (Thompson and Lapowsky, 2019).

The IRA began their campaign to manipulate citizens in the United States as early as 2014. The IRA created fake social media accounts on various platforms, pretending to be American citizens and group pages designed to attract an American audience. The IRA even communicated with people in the Trump campaign without exposing their Russian connections. By the end of the 2016 election, the IRA had the opportunity to reach millions of Americans through their social media platforms. The IRA had hundreds of thousands of followers across their many Facebook groups and Instagram accounts. Political figures even retweeted content posted by Twitter accounts linked to the IRA. There was no evidence of people willingly or knowingly aided this Russian operation (Muller, 2019).

8.3.2 Coordinated inauthentic behaviour (CIB)

Facebook describes CIB as when groups of pages and accounts work together to mislead others about who they are and what they are doing. These pages and accounts work together in a network. Their content may or may not be false, but they are deceptive about who they are. One example of this is several accounts linked together that make it appear that they are run from one part of the world when they operate from another part. CIB may be motivated by ideological, financial or political purposes (Gleicher, 2018).

Facebook have implemented rules and standards against CIB. They want the users to feel safe and trust the accounts they are interacting with and pages that they are visiting. One of the community standards on Facebook regarding CIB is that one must not "Engage in or claim to engage in Foreign or Government Interference, which is Coordinated Inauthentic Behavior conducted on behalf of a foreign or government actor".This community standard is fitting when discussing the Russian covert operation to sway the 2016 presidential election (Facebook Community Standards, 2020).

The accounts and pages linked with the IRA's CIB-operation were deceiving people about who they were and their agenda. The IRA wanted their accounts to appear as if they were Americans, and their agenda was to spread propaganda to the American public. IRA's accounts can be divided into three different categories. Bots, humans and cyborg. "Bots are highly automated accounts designed to mimic human behaviour online" (Bradshaw and Howard, 2019, p. 11). Human accounts are, as the name implies, run by humans. They engage in live interactions with other accounts. Cyborg accounts is a blend of bots and humans (Bradshaw and Howard, 2019).

8.3.3 Tactics employed by the IRA

IRA employees who were responsible for operating several social media accounts were called specialists. They started their campaign in 2014 where the IRA sent employees to the US on an intelligence-gathering mission (Muller, 2019). The IRA waged in a propaganda war with the US. They wanted to manipulate the citizens through social media to influence American culture and politics. The IRA targeted different groups of citizens. They were ranging from groups on the right-wing of the political spectre, liberal-oriented voters, religious groups, LGBTQ community and the African-American community. The IRA created different pages an platforms to communicate and to create a community within each sub-community. They aimed to reinforce a sense of belonging and loyalty within the targeted community. In this way, they were able to influence the beliefs and actions of the members (Diresta et al., 2019).

IRA's methods were not only automated bots who created and posted content for their social media accounts. Many of IRA's social media accounts were operated by humans who interacted live with those they were trying to influence. They created original content, engaging in live activities and also retweeted and shared content from people of influence. In this way, they were creating influencers and avoided getting flagged as bots or linked to Russian covert operations. The IRAs influencers created content that was shared by political figures (Muller, 2019).

The social media accounts and pages controlled by the IRA reached tens of millions of US citizens. According to Facebook, when they deactivated the accounts controlled by the IRA in 2017, they had over 80 000 posts. In that time, they had reached at least 29 million people, and Facebook estimates that they might have reached as many as 126 million people. Some of their most successful accounts alone had a large following. The Facebook group, "United Muslims of America", had over 300 000 followers. The "Don't Shoot Us" group had 250 000 followers, and the group "Being Patriotic" had over 200 000 followers (Muller, 2019).

The dataset provided by Facebook showed that there were 81 pages and that 33 of these had over 1000 followers. Fourteen pages focused on the African-American population, thirteen pages targeted right-wing audience and five pages were targeting left-wing audience. Across all the pages targeting the African-American community, there were 1,187,810 followers, across all right-wing focused pages, there were 1,446,588 followers, and across all left-wing pages, there were 689,045 followers. Across all pages, Facebook reported 76,5 million engagements. 30,4 million of the engagements were shares, 37,6 million were likes and 3,3 million comments. The data set from Facebook did not include information about "fake likes", therefore the assumption is that all engagements were from real accounts and that the content was distributed into the newsfeeds of the accounts friend. The dataset from Facebook illustrates the reach of the content from IRA controlled pages on Facebook (Diresta et al., 2019).

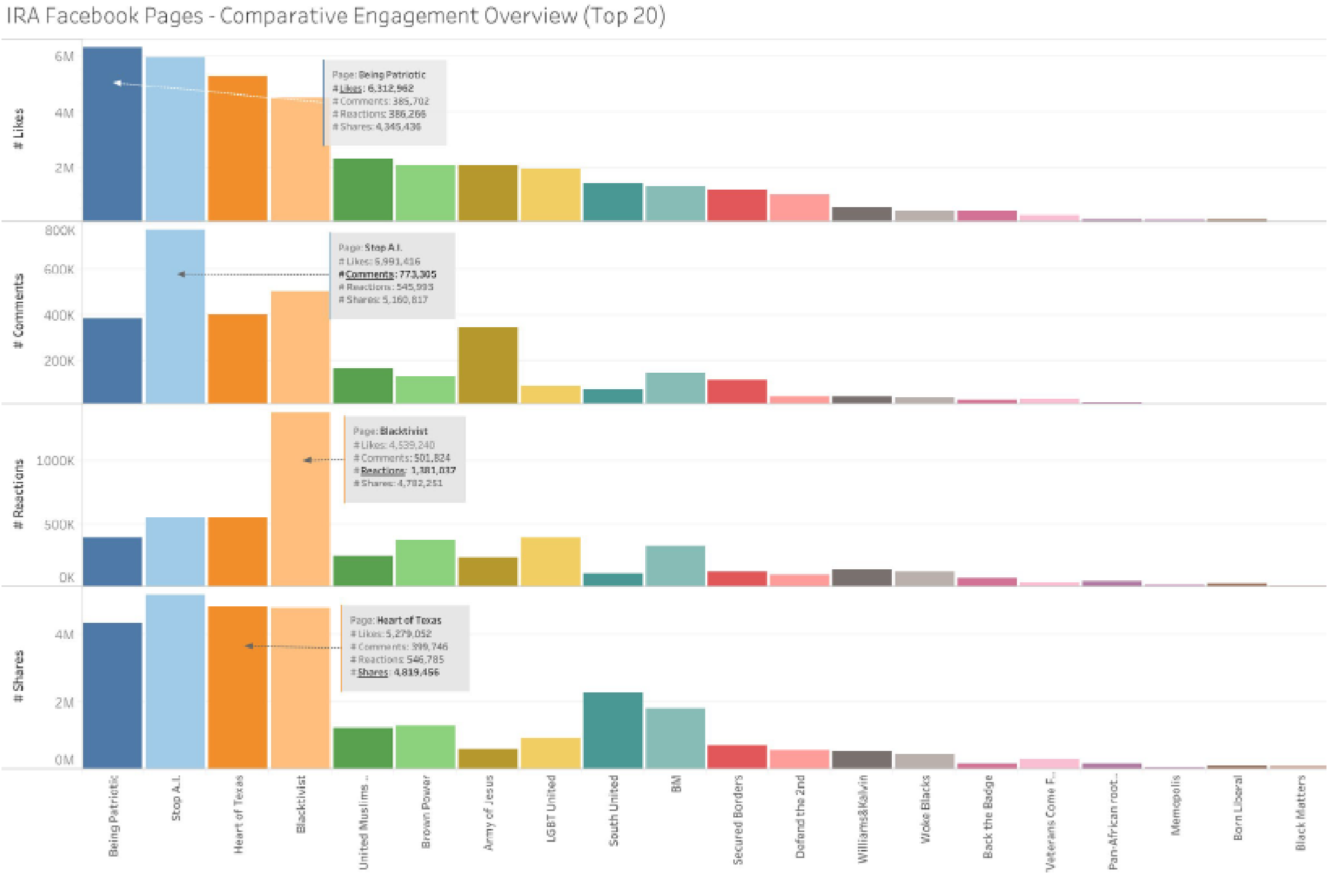


Figure 8.1: Different types of engagements across the top 20 pages on Facebook controlled by the IRA (Diresta et al., 2019, p. 22).

Paid advertisement for targeting Americans

One of IRA's tactics was to pay for advertisement targeting specific groups in the US. They paid for advertisements across different social media platforms were they wanted users to like pages, follow accounts, join events and visit websites. In total, the IRA paid for 3519 ads on Facebook and Instagram. The IRA wanted to create a community of people with similar ideologies. Of 3519 ads, 1852 was interest-based targeted. Their ads had different targeting specification set to reach people in different cities, gender, age, sex, political view and race (Diresta et al., 2019).

According to Facebook, the 3519 ads cost close to 100 000 dollars. In 2016, they increasingly started to support President Trump and oppose Sec. Clinton as a candidate explicitly. On the 18th of March, the IRA posted an ad on Facebook that had a picture of Sec. Clinton and the caption "If one the God lets this liar enter the White House as a president - that day would be a real national tragedy". On the 6th of April, the IRA posted an ad for their "Black Matters" account that encouraged people to take a picture and upload it with the hashtag "HillaryClintonForPrison2016" or "nohillary2016". Their ads on Sec. Clinton was almost always negative (Muller, 2019).

Their ads on President Trump was unlike their ads on Sec. Clinton, exclusively positive and showed support for his campaign. The first known IRA ad that explicitly endorsed President Trump was on the 19th of April. They had an ad on Instagram for their account "Tea party news", that asked young people to help them make a patriotic team of young Trump supporters by uploading pictures with the hashtag "KIDS4TRUMP". In the following months, the IRA posted their ads mostly through their other accounts that targeted Trump-supporters such as "Being Patriotic", "Stop All Invaders" and "Secured Borders" (Diresta et al., 2019).

The IRA put an extra effort into targeting people in the African-American community. One method was first to target a geographic location for a local event or rally. Then they targeted the people with content of police-brutality where African-Americans were involved. One domain controlled by the IRA, blackmattersus.com, spent in total 163 459,60 RUB on paid advertisement. Their advertisements generated 1 327 862 impressions and 87 837 clicks (Diresta et al., 2019).

Another way the IRA targeted specific groups was with age. They targetted people under the age of 15 with meme content, while people over the age of 17, who had an interest in weapons, were targeted with content related to the second amendment (Diresta et al., 2019).

When analyzing the numbers of clicks on the ads created by the IRA, they had a higher clickthrough rate than typical Facebook ads. In 2018 the average clickthrough rates on Facebook ads were 9 %. In 2015-2017 the IRA ads had an average clickthrough rate of larger than 9 %. Their ads were successful in targeting their wanted audience (Diresta et al., 2019).

Cross-platform branding

The term cross-platform branding means to build the brand using different platforms to promote the brand. In today's society, the use of social media in marketing and building the brand is beneficial. It entails spreading the brand's message across different platforms to reach a broader audience. Cross-platform branding was one of the tactics employed by the IRA. As stated earlier, the IRA created different websites, pages and accounts to build a community within the targeted group. They then used different accounts and pages to promote the same message (Diresta et al., 2019).

Their effort into cross-platform branding can be illustrated by "Black matters". Black matters was a website with linked accounts on various social media platforms. The domain had a Facebook group called "BM" where they also paid for Facebook ads and stickers. They had 31 paid advertisements on Google. They had an account on YouTube where they posted 95 videos. Their Instagram account Blackmattersus had 28 466 followers with 1 929 855 engagements. On Twitter, they had 5841 followers (Diresta et al., 2019).

The brand Black matters are one example of how the IRA used cross-platform branding to promote their agenda, seeking to influence a large number of people. Big influencers shared Black matters content on their social media accounts, giving the IRA controlled brand free advertisement and the possibility to expand even further. They were successful with some brands like Black matters, where people willingly shared their content without it being flagged (Diresta et al., 2019).

The IRA spread the "Black matters" group across the social media ecosystem, creating an inauthentic media property to spread their message and reinforce their brand. "Black matters" sought to create division online as well as in real life. They arranged real events and hired people to write and create content for their website blackmattersus.com. They engaged in real conversations and worked with Americans to decrease the risk of being detected as inauthentic behaviour. The one to one interaction with followers was a way for them to get in touch with immigration lawyers, content creators and everyone they saw as valuable. They even posted about a reality show, where they wanted possible contestant to send in a video describing the problems that African-Americans were facing (Diresta et al., 2019).

Memes

A meme is "an idea, behaviour, style, or usage that spreads from person to person within a culture" (Merriam-Webster, n.d.). Social media platforms are a popular forum to share memes. A meme can be innocent and funny, but also convey an ideological and cultural message. Every person that is active on social media is exposed to memes. People who are not interested in politics may be exposed to propaganda. In the case of the IRA's effort in swaying the 2016 US Presidential election, many Americans saw IRA's memes (Vasquez, 2019).

The creation and posting of memes were one of the tactics employed by the IRA. A meme is a simple but effective way of spreading a message to a large number of people. Memes are engaging, fun and can contain a powerful message. It is an effective way of spreading information or propaganda on social media. Memes are useful because they are easy to understand without having to read a large text. They often contain a large picture with minimal text. According to the United States Defence Department, a meme is a powerful tool to change values and behaviour. Memes are the propaganda of the digital age (Diresta et al., 2019).

The IRA used memes as a cultural signifier within each group. Memes are easy to reshare and to recontextualize to fit within the groups' agenda. The IRA created memes that stuck to the theme of the group. Memes were a large part of the IRA's visual content posted on social media. The IRA recycled the memes that got the most likes and shares within the groups and accounts. One example of this is a meme with Jesus that the Instagram account Army_of_jesus_ posted on the 2nd of March 2016. The post got over 87 000 likes. The same Instagram account posted the same meme on the 13th of June 2016, and this time the post got over 84 000 likes (Diresta et al., 2019).

The different pages and accounts shared other pages and accounts memes, and they also shared content posted on Tumblr on Instagram and visa versa. Sharing content from other social media platforms was also a form of cross-platform branding, where they tried to reach followers from other social media platforms. Different accounts also reused memes and branded them with their name, taking credit for creating the memes. It is unsure if the same employee at IRA operated the other accounts who rebranded the memes. A leak from an IRA hack reveals that they had a content managerial infrastructure containing folders of images for this purpose. The IRA used memes as a part of the many tactics they employed (Diresta et al., 2019).

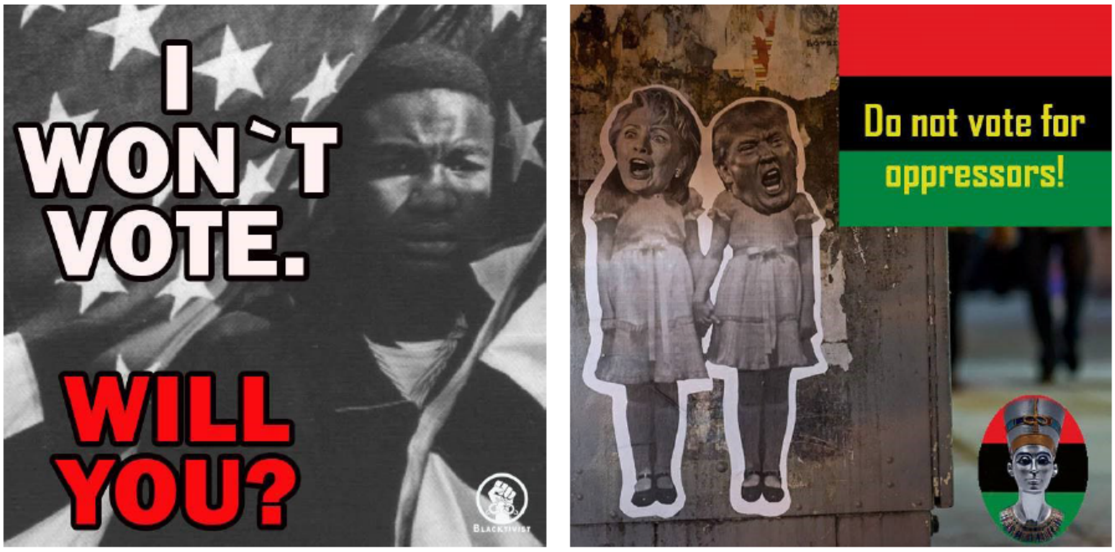


Figure 8.2: Two examples of memes posted in groups within the African-American community (Diresta et al., 2019, p. 82).

Amplify conspiratorial narratives

The IRA created conspiracy theories targeting the different groups they were manipulating. They conducted a form of psychological warfare through their social media platforms to create confusion and disorder, and to sow discord between cultural and political groups. The IRA wanted to create distrust in the system. The job was easy, considering the American peoples' trust in mass media was low in 2016. A study done by Gallup in 2016 showed that only 32 % of the American population had "a great deal" or "fair amount" of trust in the media. Only 14 % of people voting for the Republican Party had "a great deal" of trust in the media, and 51 % of people voting for the Democratic Party had "a great deal" of trust in the media (Swift, 2016). When the trust in media is low, it can open up for conspiracy theories to influence the public opinion (Farkas and Bastos, 2018).

The IRA targeted different groups with different conspiracy theories across their social media accounts. They used their Twitter accounts in particular about amplifying the conspiracy theories from their right-wing personas. Targeting people on the right-wing of the political spectre, they created conspiracies about vaccines, aliens and the famous murder of Seth Rich. The people on the left-wing were also targeted with the murder of Seth Rich and the message to trust Julian Assange. The IRA targeted the black community with historical conspiracies; such as that Mozart was black and that the Statue of Liberty originally was a black woman. The goal of creating conspiracy theories was to reinforce cultural and political identity (Diresta et al., 2019).

8.3.4 CIB as a tactic during the 2016 presidential election

How can we know that the Russian government was involved in the IRA's operation using CIB to influence the 2016 US election? Much of the content the IRA manufactured was propaganda, and they used their social media platforms to spread it. When looking at the ideology, purpose and context of the propaganda, it may lead back to the source. The Russian government had a lot to gain from this operation (Farkas and Bastos, 2018). There were many economic sanctions towards Russia at the time leading up to the 2016 US election. An example of this was the EU's sanctions after Russia's actions in Ukraine. The EU implemented economic sanctions on the 31st of July 2014. The sanctions targeted the Russian financial, energy and defence sectors. The US had also implemented trade sanctions with Russia. The Russian government would benefit from a Russian-friendly U.S president (Counsil of EU, 2016).

The tactics explained in the above sections are tactics Russian used as part of their coordinated inauthentic behaviour operation. The IRA started already in 2014, creating social media accounts and pages to manipulate US citizens and influence the political and cultural agenda. CIB does not automatically mean that the social media accounts and pages spread fake news; it means that they try to deceive people about who they are, where they come from and their agenda. The IRA created accounts and pages and targeted different groups of people. They convinced their followers that they were of the same ideology or culture, but in fact, they had another agenda.

The IRA's plan for the 2016 presidential election was to help President Trump's presidential campaign by creating support among US citizens and discrediting Sec. Clinton as a candidate. When it appeared that Sec. Clinton was going to win, the accounts linked to the IRA started to focus their content on undermining her future presidency (Muller, 2019).

8.3.5 Discussion

The official investigation into Russia's involvement into the 2016 US Presidential election conducted by Special Counsel Robert S. Muller concluded that there was not sufficient evidence to state that the Russian operation was successful in swaying the election. However, it was not because of Russia's lack of trying. The investigation shows that the Russian government ordered an operation to manipulate the 2016 US presidential election. A part of that operation was to manipulate public opinion through social media, using CIB. The research question of this paper is "what was IRA's most successful CIB-tactic when trying to sway the 2016 US Presidential election?" To answer the research question, the discussion uses the factors propagation and engagement as parameters for how successful each tactic was. Propagation is how far the content spread in the social media ecosystem, and engagement is how many people shared, liked or commented on the content. The reason that propagation is a factor for how successful they were is that it quantifies how many people saw their content. Total engagement is a factor because it reveals if, and to what extent, the content resonated with people.

The limitations of this case study are that it is difficult to prove causation between the IRA's tactics and how people voted. There could be other causes than IRA's covert operation that led to people voting differently (Wæhle et al., 2020, para. 7). Another limitation is that the results have only presented quantitative data of IRA's efforts from the largest social media platforms, and no qualitative data to gauge the impact IRA's CIB-operation had on users exposed to their content. There is no data to show if Americans acted differently because of exposure to the IRA's content. The significance of the results is that a thorough investigation lead by Robert s. Muller has looked into the Russian involvement in the 2016 US Presidential election. The Senate Intelligence Committee ordered also ordered a report to look into the IRA's involvement. The report written by Renee Diresta et.al and Robert S. Muller's investigation has greatly influenced this report. The results explain the amount of content produced by the IRA and how many Americans were exposed to that content. The dataset provided by the major social media platforms is evidence of a significant operation ran by the IRA.

8.3.6 Propagation and created engagement

To influence the election, the IRA did not have to manipulate people who already supported President Trump. They would have to influence the people who did not have an opinion about the election or was unsure about their vote. The IRA had to manipulate people who did not support President Trump to either not vote or vote for another candidate than Sec. Clinton. To be able to do this, they had to reach the people who were unsure about their vote with content that promoted President Trump's campaign and discredited Sec. Clinton.

In terms of propagation and engagement on Facebook and Instagram, the IRA's operation did well. IRA's accounts and pages on Facebook posted in total 61 483 posts and got in a total of 37 627 085 likes and 3 339 752 comments. That is an average of 612 likes and 54 comments per post. The number of followers combined was 3 334 202, and they shared IRA's posts 30 350 130 times. In total, the IRA's accounts on Facebook received 77 million engagements. When sharing, liking and commenting on posts on Facebook, it spreads to other accounts that are friends with the one doing the interaction. When engaging with content created by the IRA, it spread to more accounts than only the accounts that engaged with their post (Diresta et al., 2019).

IRA's accounts on Instagram had in total 3 391 116 followers. They had in total of 116 205 posts generating 183 246 348 likes and 4 017 731 comments. That is an average of 1568 likes and 34 comments per post. Their Instagram content created more engagement than their Facebook content. In total, the IRA's accounts received 187 million engagements on Instagram. The combined engagement on both Facebook and Instagram, which includes likes, comments, shares and reactions, was in total 263 769 228 engagements. Some accounts and pages had more followers and engagement than others did (Diresta et al., 2019).

The propagation and engagement on Twitter was also a success story for the IRA. IRA created over 3841 persona accounts and approximately 1.4 million people engaged with their tweets. The accounts created both original contents and retweeted content from other accounts. On their original content, they got 72 801 807 engagements (Diresta et al., 2019).

Facebook estimates that the content created on Facebook may have reached as many as 126 million people. In terms of propagation, 126 million people is a lot and far more than the 3.3 million people that followed the accounts and pages (Muller, 2019). The content created on Facebook had a significant reach in terms of people viewing their content. That massive reach increases the chances of reaching people who resonate with their message and possibly change their actions. When looking at how many retweets the IRA's accounts got of their original content, show that their message was resonating with people. The content about Clinton was almost exclusively negative, and it had the potential to influence people that were less likely to be critical about the information they receive.

The IRA's use of ads on social media platforms helped them spread their content. The IRA targeted people from different areas, age groups, sex and more. When looking at how much money the IRA spent on ads, it is clear that this was a pivotal tactic to spread their content. It was necessary to create communities to manifest a sense of tribalism within the group. Being a part of a community and feeling a sense of identity in the ideology can be a powerful way of influencing people.

Memes were one of the tactics the IRA used for all its worth. Memes are the new form of propaganda. A message of not to vote communicated through memes is a powerful way to make it resonate with the followers. IRA saw most engagement on their accounts on Instagram. That could be a result of Instagram being ideal for posting memes, and in turn, be the best battleground for their memetic warfare (Diresta et al., 2019).The propagation of memes created by the IRA was significant. The memes spread to different forums and pages on the Internet. Some of the memes created by the IRA are still relevant in the targeted communities. This speaks to the propagation of the memes. It shows how much the memes resonated with people from the targeted communities. If the memes manage to change people's beliefs or actions cannot be said (Diresta et al., 2019).

The page blackmattersus.com had a strong following across different social media platform. They rarely posted about Trump's campaign, and their focus was mostly about spreading negative stories about Clinton. The IRA's effort into cross-platform branding could be one reason for the large following that "blackmattersus" had. "Blackmattersus" were a large brand on several social media platforms. They reached many African-American voters and positioned themselves to influence the votes of the followers. The messages that the "Blackmatterus" conveyed to their followers was that Sec. Clinton was corrupt and that they should not vote (Diresta et al., 2019).

The IRA targeted the African-American community heavily. The purpose was to suppress voting or to vote for third-party candidates (Diresta et al., 2019). The IRA targeted the African-American community with advertisement discrediting Sec. Clinton as a candidate. The memes created by the accounts associated with the African-American community were showing distrust with the candidates and wanted the followers not to vote. According to Pew Research Center, the percentage of African American voters dropped to 59,6 % in the 2016 presidential election. That was a 7 % drop for the all-time high voting percentages from 2012 of 66,6 % (Krogstad and Lopez, 2017).

How large of an impact the IRA's operation against the African-American community had is difficult to read from the information. There can be several factors why the voter turnout of African-American people was low. The candidate in the 2012 presidential election, Barack Obama was a popular figure in the African-American community. From the election in 2004 to the election in 2008, voter turnout of African-American's increased with 4,9 % (*Dissecting the 2008 Electorate: Most Diverse in U.S. History*, 2009). From the election in 2008 to the election in 2012 the percentage of African-American voters increased again from 65,2 % to 66,6 %. The percentage of African-American voters had to that point been increasing for the last 20 years. In 2012, the percentage of African-American voters surpassed white American voters for the first time (Krogstad and Lopez, 2017).

The statistics show that Barack Obama was a popular president within the African-American community, and could be a factor to why there was a significant drop in voter turnout from the African-American community. Another factor could be that non of the candidates resonated with African-American voters, and it had little to do with the IRA's operation.

The IRA constructed conspiracy theories to create distrust in the system and to sow discord between different groups. The effects of conspiracy theories are that people who believe in them can make decisions based on them, which in turn can make them do things they initially did not want to do (Lucas et al., 2018). There is not much data on how the conspirational narratives created by The IRA propagated on social media, and not enough data about the created engagement on it.

To summarize the discussion, the IRA wanted to influence people in a few ways. They endorsed President Trump's campaign and discredited Sec. Clinton's. They also tried to suppress voters and make people vote for third-party candidates. The propagation of the content created by the IRA was significant. The content on Facebook reached 126 million people, who speak to how far their content spread through the social media ecosystem. The amount of engagement created on their content, millions of likes, shares and comments on Facebook, Instagram and Twitter, shows that people resonated with their message. The IRA used several tactics in their CIB-operation. They paid for advertisements to grow their groups and pages, an important tactic to spread their propaganda. They focused their efforts to cross-platform branding, establishing a brand within the communities. An example of their success in cross-platform branding was their brand "Blackmattersus". Spreading propaganda through memes is a powerful way of influencing people. The IRA used memes as a part of every tactic they employed. They reposted popular memes, rebranded them and posted them on different accounts, as it was their original content. The IRA also spread conspirational narratives as a tactic to create distrust in the system and to sow discord between cultural and political groups in the US.

The answer to the research question, what was IRA's most successful CIB-tactic when trying to sway the 2016 US Presidential election, is that there is not possible to conclude with one tactic being their best. The whole operation was impressing, and it might be due to a few key-tactics that was mentioned in this report.

8.4 Conclusion

IRA's covert operation to target the 2016 US Presidential election was extensive, well planned and well-executed. The amount of time and resources that went in to creating fake social media accounts and pages, creating original content, and engaging with people shows that the operation was essential for the Russian government. The extensive network of accounts and pages created to deceive the audience of who they were and their agenda shows how coordinated their inauthentic behaviour was. The IRA created different communities within different demographics. The propagation of their content was massive, possibly reaching upwards of 126 million people on Facebook. The engagements on their content were also significant, getting close to 264 million engagement on their content on Facebook and Instagram.

There was not enough data to give a definitive answer to the research question "what was IRA's most successful CIB-tactic when trying to sway the 2016 US Presidential election". The result shows that the IRA used different tactics, and they were successful with many of them. The IRA was successful with their approach to target many different political, cultural and ethnical groups. The significance of the results show how extensive IRA's operation was, and a testimony to that is the extensive investigation led by Robert S. Muller. The investigation shows that there was an apparent attempt to sway the election in favour of President Trump. Their tactics were many, and the attempts came from different angels. The IRA promoted President Trump's campaign and discredited Sec. Clinton's campaign. The IRA also tried to suppress voters in groups less likely to vote for Trump. There was not sufficient data to state that the CIB-operation succeeded in swaying the election.

8.5 Future work

Nations need to be aware of the possibility of foreign nations trying to manipulate democratic elections in the future. Foreign nations like Russia might have an interest in who is elected as head of state. Their covert operations show how far they are willing to go to succeed with manipulating democratic elections. Manipulation of democratic elections will undermine democracy, and it is something every democratic nation must prevent. The US Presidential election is essential for the geopolitical climate because they are a significant factor. Every state looks to the US Presidential election because they are either a vital ally or a great adversary. Manipulation of elections in other countries than the US can be a target for Russia, China and Iran, and it could already have happened.

For future work within the topic of foreign CIB-operations to manipulate democratic elections, it will be interesting to see case studies of other elections where election manipulation might have happened. A study like that can give insight into methods employed by other nations. If democracies can learn from previous attempts, they can build systems to guard for it. To manipulate the US Presidential election in 2020 was probably a target for some foreign nations. As Robert S, Muller concluded in his investigation of the 2016-election, Russia tried to manipulate it in favour of President Trump. Did Russia try to do the same in 2020, and what tactics did they employ then?

Bibliography

Bradshaw, S. and Howard (2019), 'The global disinformation order 2019 global inventory of organised social media manipulation'.
URL: *https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf*

Counsil of EU (2016), 'Russia: Eu prolongs economic sanctions by six months'.
URL: *https://www.consilium.europa.eu/en/press/press-releases/2016/07/01/russia-sanctions/*

Diresta, R., Shaffer, K., Ruppel, B., Sullivan, D. and Matney, R. (2019), 'Digitalcommons@university of nebraska -lincoln'.
URL: *https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs*

Dissecting the 2008 Electorate: Most Diverse in U.S. History (2009). Accessed 18.11.2020.
URL: *https://www.pewresearch.org/hispanic/2009/04/30/dissecting-the-2008-electorate-most-diverse-in-us-history/*

Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements | Permanent Select Committee on Intelligence (n.d.).
URL: *https://intelligence.house.gov/social-media-content*

Facebook Community Standards (2020).
URL: *https://www.facebook.com/communitystandards/inauthentic_behavior*

Farkas, J. and Bastos, M. (2018), 'Ira propaganda on twitter', *Proceedings of the 9th International Conference on Social Media and Society* .

Gleicher, N. (2018), 'Coordinated inauthentic behavior explained'. (Accessed: 01.09.2020).
URL: *https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/*

Internet Research Agency (2019), 'Internet research agency'.
URL: *https://en.wikipedia.org/wiki/Internet_Research_Agency*

Krogstad, J. M. and Lopez, M. H. (2017), 'Black voter turnout fell in 2016, even as a record number of americans cast ballots'.
URL: *https://www.pewresearch.org/fact-tank/2017/05/12/black-voter-turnout-fell-in-2016-even-as-a-record-number-of-americans-cast-ballots/*

Lucas, J., Galdieri, C. and Sisco, T. (2018), 'Conventional wisdom, parties, and broken barriers in the 2016 election'.

Merriam-Webster (n.d.), 'Coerce'. Accessed 13.09.2020.
URL: *https://www.merriam-webster.com/dictionary/coerce*

Muller, R. (2019), 'Mueller report highlights: Read the top moments from the 448-page report'. **URL:** *https://abcnews.go.com/Politics/justice-department-release-redacted-version-mueller-report/story?id=62201315*

Swift, A. (2016), 'Americans' trust in mass media sinks to new low'. **URL:** *https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx*

Thompson, N. and Lapowsky, I. (2019), 'How russian trolls used meme warfare to divide america'. **URL:** *https://www.wired.com/story/russia-ira-propaganda-senate-report/*

Vasquez, C. (2019), 'Case study analysis of shared visual arguments and propaganda techniques of russian propaganda posters and russian internet memes', *Master thesis* .

Wæhle, E., Dahlum, S. and Grønmo, S. (2020), 'case-studie'. **URL:** *https://snl.no/case-studie*

Chapter 9

The Severity of Coordinated Inauthentic Behaviour

By Michael T. Gebremariam

9.1 Excutive summary

Today there are many examples of big and small governments that have been victimised because of CIB. On the contrary there are also governments that have benefited from the phenomenon to suppress and attack groups and other foreign countries and vice versa. But how can governments, groups and individual got access to such data so that they can put others on misery and death? How does social media companies handling the situations so far, and what will be the role of governments and the society at large to handle or combat CIB? Those are some of the questions that will be addressed on this paper.

This paper is also going to see the wide spreading and severity of coordinated inauthentic behaviour by taking sample cases. It will not only include the effects of Coordinated inauthentic behaviour's on the Social media manipulations of elections during the 2019 European parliamentary elections, but also on other social media issues like the ongoing Corona virus (COVID-19) pandemic conspiracy theory and the Australian bushfire season on the late of 2019 until the early 2020.

Since the ability of social media platforms to handle CIB and the techniques being used to do so will be widely covered by the group members topics, here it will just be given general comments so that one can rate the status of CIB on the internet era.

9.2 Introduction

Science has been defied, climate change has been denied, religions has been attacked. Those are some of the consequences of coordinated inauthentic behaviour. Many countries are facing dilemmas where the very existence of their governments er at the hands of the social media platforms. Countries like Australia, and India have contacted the social media owners to come up with better solutions for their platforms so that the integrity of their nations would not be put on compromises (Kouvela, 2020) , (Vinson, 2010).

If a person wanted to attack a government or a group of a society to change political views or polarize opinions; he then can simply go to the dark nett to buy manipulation services with lots of inauthentic comments, clicks and followers. Those manipulations rely on inauthentic accounts that can interact with other online accounts to distort or influence the perception of the public audience. The accounts can be owned by unidentified people or bots that can respond or retweet contents automatically (Keller et al., 2020).

9.3 Coordinated inauthentic behaviour as a persuasion tool

Coordinated inauthentic behaviour now a days is being used as a persuasion tool on the social media platforms to affect beliefs or political views or divert the outcome of an electoral result ... etc. of a targeted audience. The social media platform companies not only provide social accounts that are used by individuals to socialise but also, they provide software-controlled accounts (bots) that are one of the main actors associated with manipulation of political campaigns. Those platforms are full of fake accounts, false narratives, low-credibility news sources, stat-sponsored operators and so on which acts negatively on the integrity of the social media intentions of socialising with fellow human beings (Luceri et al., 2020).

Online disinformation is causing a great deal of problems to the computer science, social science, and political science communities. The response to online disinformation demands automated fact checking mechanisms, psychological experiments concerning belief and persuasion, documentation of active disinformation campaigns that is taking place, and proposed responses to specific threats. It is the joint effort of such communities is expected to give results in combating coordinated inauthentic behaviour from being a persuasion tool (Stray, 2019).

9.4 The status of coordinated inauthentic behaviour

On an experiment conducted by the NATO Strategic Communication Centre of Excellence (Multinationally constituted and NATO-accredited international military organisation) to measure the ability of the social media platforms on combat coordinated inauthentic behaviour on their platforms. They concluded that Facebook, Instagram, Twitter, and You Tube are failing to tackle coordinated inauthentic behaviour. The experiment was conducted by applying manipulative contents like comments, views, and likes bought on the darknet from a range of European and Russian media manipulation service providers. The experiment was conducted on May-August 2019 and the report is published on the book “FALLING BEHIND: How social media companies are failing to combat inauthentic behaviour online” (of Excellenc, 2019).

9.4.1 Case 1: European parliamentary elections

The aim of this topic is to discuss and see if there were inauthentic behaviours acted by coordinated entities on the European Parliamentary Election that was held between 23 and 26 of May 2019 (Iosifidis and Nicoli, 2019), (Keller et al., 2020).

According to the Association of Internet Researchers (AOIR) 21st annual conference papers, the European parliamentary elections had been affected by the Twitter social media video sharing mechanism. The investigative report had further stated that it conducted a multi-lingual, cross-case analysis of twitter images posted by users from 6 different countries. The images were investigated in the context of the EU election and EU membership. From the findings of the investigations, it was concluded that visual media played a central role in the twitter political discourse before the elections, both as a channel for official campaigning and candidate communication and for new forms of political expression and user-generated political content (Keller et al., 2020).

Other sources also indicate the same results. For example, an article with the title "It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections" conducted an analysis of the phenomenon that there were coordinated inauthentic behavioural activities on the European parliamentary elections of the 2019, though the paper doubts its own approaches to the conclusion (Giglietto et al., 2020).

9.4.2 Case 2: Australian bushfire season

On the period between September 2019 and March 2020, Australia has faced the most horrific bush fire on its history ever. According to the journal media culture "Uncovering a climate catastrophe?" the incident had resulted on a 12.6 million hectares of land burn, and billions of animals and vegetations species extinction. Of course, this is without the human lives lost and displacement, the houses and infrastructures damaged, and the pollution it exerted on the environment (Mocatta and Hawley, 2020), (Pacific, 2020).

Although it is well documented that the global climate change resulted the incident, some opportunistic people found grounds to divert the public opinion from the subject of climate change to arsonists that caused the fires. They used coordinated inauthentic information to persuade some peoples opinions (Pacific, 2020).

The Association of Internet Researchers (AOIR) 21st annual conference papers noted that the hashtags like #ArsonEmergency played a pivoted role in ruling that the bushfire is caused by arsons rather than climate change. This hashtag and the likes are bot-controlled misinformation entities that had been deployed by some people that had the interest of keeping the established government regulations on their favour (Keller et al., 2020).

9.4.3 Case 3: Corona virus (COVID-19) pandemic conspiracy theory

Corona virus (COVID-19) pandemic outbreak on the late 2019 and is still spreading all over the world. The very human nature of socialising is put into question, thousands of people are dying every day, countries are facing economic challenges that some never anticipated (Gruzd and Mai, 2020), (Graham et al., 2020).

Yet some found means to mis and disinform fellow citizens and the public about the pandemic for political and other purposes. The Australia institute conducted a research on COVID-19 infodemic and released an article with the title "Like a virus: The coordinated spread of coronavirus disinformation". In the article it stated that "Analysis of over 25.5 million tweets over 10 days identifies 5,752 accounts that coordinated 6,559 times to spread mis- and disinformation regarding the coronavirus for either commercial or political purposes. Almost all politically motivated activity promoted right wing governments or parties. Coordinated spreading of the China bioweapon conspiracy theory is estimated to have made over 5 million impressions on Twitter users, spread by mainly pro-Trump, partisan conservative and/or QAnon accounts." Which clearly shows the use of coordinated inauthentic behavioural activities (Graham et al., 2020).

Another article "Enduring Information Vigilance: Government after COVID-19" wrote by the UNITED

STATES ARMY WAR COLLAGE PRESS showed that coordinated inauthentic information on the COVID-19 are been deployed on people by various state nations for political objectives. Furthermore, the article concluded that the COVID-19 infodemic highlighted a threat to democracy that the western world need to address (Jankowicz and Collis, 2020).

9.5 Handling coordinated inauthentic behaviour incidents

As Facebook's head of cybersecurity policy Mr. Nathaniel Gleicher in explaining the term coordinated inauthentic behaviour stated, Facebook take down pages and networks of criminals by the effort of manually laboured staff and automated technology. He further explained the process as if it is that "someone is looking for a needle on a haystack" to show the difficulty of the job. He did not explain or clarified as to how the exact process takes place. From the explanation we know that they take down if the perpetrator is from a different location than the place of origin or fail to self-identify. But the main issue of targeted audience disinformation or political inclination of either the perpetrator or the staff member that conducts the operation of taking down is not clear (Gleicher, 2018).

On an article wrote by Anita Gurumurthy Jai Vipra Apr 2019 "Misleading Takedowns: Facebook needs to be a lot more transparent when it comes to banning pages, groups" suggested that platforms like Facebook takes down pages on their own terms which some times can be very misleading specially on political parties disputes. This kind of shutting down one and leaving the other can give negative impact. The article argues that there should be much investigations on to the root of the causes, and further gave example of incidents happened in India (Vipra, 2019).

Another report written by Taibbi, Matt and Rolling Stone in December 2018 "Who Will Fix Facebook?" claimed that there were many authentic websites that has been taken down because they fell into certain categories. It also stated that the platform is going to work with the Atlantic Council on the matter of inauthentic behaviour which is something the writer of the article did not agree with because of the political inclination of the organization. The organization is a collection of as it is described by the article exspies, senior military leaders, former CIA heads, and neocons that are financially backed by weapons makers like Raytheon, energy titans like Exxon-Mobil and banks like JPMorgan Chase (Taibbi, 2018).

9.5.1 Facebook

Coordinated inauthentic behaviour is the most challenging task social media platform Facebook faces now. The platform releases monthly report on how many pages it removed. SocialMediaToday make track of Facebook's monthly report and presented it on a form of table. The table does not include June's report because Facebook fail to report on that month (Hutchinson, 2020).

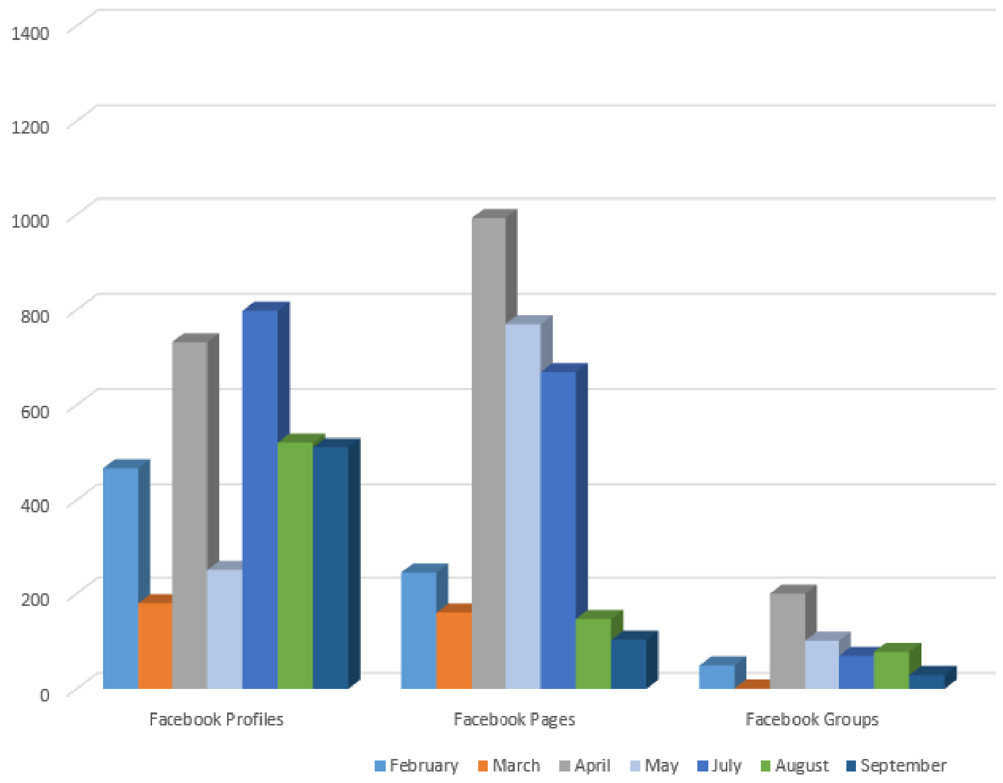


Figure 9.1: Table Facebook CIB Removals from SocialMediaToday

Facebook has stated that it takes out pages and networks from its platform by a means of manually working investigating staff members and automated technological processes. It combines those methods to facilitate the process of sorting out bad actors from its platform. It for example deletes and stop millions of attempts to create fake accounts every day with the help of automated technologies (Gleicher, 2018).

9.5.2 Twitter

The social media platform Twitter also faces a demanding challenge as the rest about coordinated inauthentic behaviour activities. According to SAGE journals article with the title “Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter”, twitter has a better control over its users information data internally than in its public API dataset. So, it made it very clear to its users by putting some rules that can act like guidelines for tweeting and/or retweeting. Some of the rules include:(Gruzd and Mai, 2020)

- artificially amplifying or suppressing information,
- interfering in elections,
- sharing synthetic/manipulated media which may cause harm, or
- promoting violence against, threatening, or harassing an individual or a group of people is not allowed.

This way there is a better tendency of control over coordinated inauthentic behaviour. Users have the chance to delete tweets or tweets disappear because the platform has flagged them as inappropriate or unverified (Gruzd and Mai, 2020).

9.6 Conclusion

The impact of coordinated inauthentic behaviour on the social media platforms as we have seen from this report has played big role in mis- and disinforming the public at large and local on issues like the European parliamentary elections, COVI-19 global pandemic, Australia bushfire season, and more. Those er few examples of the very critical issues so far. This comes to answering the question of severity of coordinated inauthentic behaviour as very bad. Experiments conducted by top NATO Strategic Communication Centre of Excellence show that the social Media companies er failing on handling the problem.

Purposefully acted Coordinated inauthentic behaviour has a great impact on our daily lives spiritually, politically, Health wise, and so forth. With these being said, we are left vulnerable for behavioural attacks from inauthentic actors, our democracy under threat, and yet no one is doing anything about it.

Bibliography

Giglietto, F., Righetti, N., Rossi, L. and Marino, G. (2020), 'It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 italian elections'. Accessed 22.11.2020.
URL: *https://www.tandfonline.com/doi/full/10.1080/1369118X.2020.1739732*

Gleicher, N. (2018), 'Coordinated inauthentic behavior explained'. (Accessed: 01.09.2020).
URL: *https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/*

Graham, T., Bruns, A., Zhu, G. and Campbell, R. (2020), 'Like a virus: The coordinated spread of coronavirus disinformation'. Accessed 22.11.2020.
URL: *https://eprints.qut.edu.au/202960/1/P904_Like_a_virus_COVID19_disinformation_Web_.pdf*

Gruzd, A. and Mai, P. (2020), 'Going viral: How a single tweet spawned a covid-19 conspiracy theory on twitter'. Accessed 22.11.2020.
URL: *https://journals.sagepub.com/doi/full/10.1177/2053951720938405*

Hutchinson, A. (2020), 'Facebook provides update on actions taken against coordinated inauthentic behaviour'. Accessed 22.11.2020.
URL: *https://www.socialmediatoday.com/news/facebook-provides-update-on-actions-taken-against-coordinated-inauthentic-b-1/586710/*

Iosifidis, P. and Nicoli, N. (2019), 'The battle to end fake news: A qualitative content analysis of facebook announcements on how it combats disinformation'. Accessed 22.11.2020.
URL: *https://journals.sagepub.com/doi/full/10.1177/1748048519880729*

Jankowicz, N. and Collis, H. (2020), 'Enduring information vigilance: Government after covid-19'. Accessed 22.11.2020.
URL: *https://press.armywarcollege.edu/parameters/vol50/iss3/4/*

Keller, T. R., Graham, T., Angus, D., Bruns, A., Marchal, N., Neudert, L.-M., Nijmeijer, R., Nielbo, K. L., Mortensen, M. D., Anja Bechmann, P. R., Stromer-Galley, J., Baptista, E. A. and de Oliveira, V. V. (2020), 'Coordinated inauthentic behaviour' and other online influence operations in social media spaces'. Accessed 22.11.2020.
URL: *https://journals.uic.edu/ojs/index.php/spir/article/view/11132/9763*

Kouvela, M. (2020), 'Bot detective: Explainable bot detection in twitter'. Accessed 22.11.2020.
URL: *http://ikee.lib.auth.gr/record/319595/files/GRI-2020-27499.pdf*

Luceri, L., Cardoso, F. and Giordano, S. (2020), 'Down the bot hole: actionable insights from a 1-year analysis of bots activity on twitter'. Accessed 22.11.2020.
URL: *https://arxiv.org/pdf/2010.15820.pd*

Mocatta, G. and Hawley, E. (2020), 'Uncovering a climate catastrophe? media coverage of australia's black summer bushfires and the revelatory extent of the climate blame frame'. Accessed 22.11.2020.
URL: *https://journal.media-culture.org.au/index.php/mcjournal/article/view/1666*

of Excellenc, N. S. C. C. (2019), 'Falling behind: How social media companies are failing to combat inauthentic behaviour online'. Accessed 22.11.2020.
URL: *https://images.politico.eu/wp-content/uploads/2019/12/Final-Falling-behind_StratCom_COE.pdf*

Pacific, G. A. (2020), 'Dirty power: Burnt country'. Accessed 22.11.2020.
URL: *https://www.youtube.com/watch?v=4FjmlNo-Ai8*

Stray, J. (2019), 'Institutional counter-disinformation strategies in a networked democracy'. Accessed 22.11.2020.
URL: *file:///C:/Users/Michael/Desktop/2019-2020/Master%20studiet/IMT%204115%20(2020)/Institutional%20counter-disinform%20strategies%20in%20a%20networked%20democracy.pdf*

Taibbi, M. R. S. (2018), 'Who will fix facebook?'. Accessed 22.11.2020.
URL: *https://search.proquest.com/docview/2211285243?rfr_id=info%3Axri%2Fsid%3Aprimo*

Vinson, K. E. (2010), 'The blurred boundaries of social networking in the legal field: Just "face" it'. Accessed 22.11.2020.
URL: *https://search.proquest.com/docview/860067897/41C2D9D34673468EPQ/4?accountid=12870*

Vipra, A. G. J. (2019), 'Misleading takedowns: Facebook needs to be a lot more transparent when it comes to banning pages, groups'. Accessed 22.11.2020.
URL: *file:///C:/Users/Michael/Downloads/MisleadingTakedowns_Facebookneedstobealotmoretransparentwhenitcomestobanningpagesgroups.pdf*