What It Takes to Regulate Content on Social Media and Are There Effective Ways to Regulate Political Advertisements? – The Case of Facebook

Abstract

The aim of this paper is to understand the difficulties in regulating social media, specifically Facebook, through an exploration of the components that comprise social media. The differences between social media and traditional mass media is presented by comparing four of their elements. The nature of the algorithm, that plays a significant role on social media platforms, how it creates a filter bubble for each user, who is then exposed to filtered content and exempt from another. One of the major issues the filter bubble presents for the public are the consequences of the exposure to personalised news and political advertisements that appear on each one of the Facebook's users' feeds. Further, the regulation of political advertisements on social media are explored in more detail. After the 2016 US presidential election scandal, involving fake news, disinformation, data stolen from Facebook's users and then used to micro-target them with tailor-made advertisements and the influence of foreign nationals on the outcome of the election, the truth in politics and regulation of social media platforms have been one of the most frequent topics of public concern. It is not the case that there are not any regulations for social media or the advertisements they display. There are multiple social media regulation strategies, however, all of them have a number of flaws. There is the content moderation strategy that every social media platform has itself, there are legislative regulations such as Section 230 of the Communications Decency Act or the General Data Protection Regulation, as well as non-legislative regulations, such as fact-checking, that try to reduce the disinformation and promote user privacy on the social media.

Content

1. Introduction

Social media has opened many different opportunities to the world. People can connect with one another effortlessly and the quick access to information that social media offers has never been so easy before. However, with every positive comes a negative.

The unique structure of social media has posed many difficulties in regulating the content that the social media platforms host. Since social media are privately owned companies, they are not regulated the same way as the state companies (Jørgensen and Zuleta, 2020). From the advertising point of view, social media is not mass media, hence the advertising laws and regulations for mass media do not apply to social media (van Dijck and Poell, 2013). Another point to note is their global reach, therefore the regulations of one country, do not apply for the other. They feature news but their CEOs argue that they are not media companies but technology companies (Fuchs, 2019). Not to mention the algorithm that more and more social media platforms are incorporating. I address concrete examples concerning social media by using Facebook as the social media network. Facebook's algorithm collects information, also known as data, about its users (Sneed, 2020). Then the algorithm manages the user's feed in accordance to the data it has collected about them. This system places the social media users into filter bubbles that consist of content, that includes political advertisements as well, that corresponds with the user's views and opinions. All this happens without the users knowing (Napoli, 2014). As a free platform, Facebook sells the data collected about its users and sells them to advertisers that use it to micro-target the users on Facebook with their advertisements (Sneed, 2020).

All of these factors have presented a difficult role in regulating the content on social media. However, the amount of scandals that social media has been at the centre of are calling for urgent solutions. Even though not the first but certainly the one that had the biggest global reach, the 2016 US presidential election scandal showed what social media are capable of. Not only did a company called Cambridge Analytica misuse Facebook's users' data to micro-target them with political advertisements, the technology giants confirmed that foreign nationals used their platforms to spread fake news and disinformation campaigns in order to influence the outcome of the election (Rathi, 2017).

What are the content regulations in place and are they effective? Apart from the social media platforms self-regulating their content, there are legislative and non-legislative regulations in place. Even though Facebook has multiple features on their platform to encourage its users to report fake news and otherwise inappropriate content, their content is regulated half by artificial intelligence and half by human content moderators (Andersen and Søe, 2020). Due to the public pressure for transparency and more serious take on responsibility, Facebook announced a formation of Facebook Oversight Board that is meant to make decisions about Facebook's content according to global norms. Isn't this merely another way Facebook is avoiding the blame? Do global norms exist? The legislative regulations such as the Honest Ads Act, Section 230 of the Communications Decency Act or the European Union General Data Protection Regulation (GDPR) have been in place to bring

more seriousness to the public concern about the use of their data online. However, their effectiveness has got few loopholes. From the non-legislative side, the fact-checking might be the answer to the battle with fake news. How do they assess the truth if the matter isn't simply just a yes or no answer and what do the users think about the content corrected by fact-checkers? Is, perhaps, the public awareness and education about people's data in the digital world the next solution?

2. Social Media Logic

Media is constantly evolving. At first, it was the traditional mass media that was the most influential. However, in the past two decades, it has been the new media that has been significantly shaping the present state of media (Dutta-Bergman, 2004). With the second stage of the Internet evolution, the Web 2.0, social media is on the rise and people see this as a positive move towards democracy (Yerlikaya, 2020). Before delving into the social media logic, it is worthwhile to look at it in comparison to the traditional mass media. On the one hand, mass media and social media may seem to be fairly similar. Stefanone et al. (2010) compare social media to reality TV by saying that the behaviour of the users on social media reflects the model of a reality TV programme. Some scholars say that social media and traditional mass media complement each other, since each of them offers a different type of enjoyment to its audience (Dutta-Bergman, 2004).

According to Fuchs (2019), social media consists of blogs, user-generated content sites, and social networking sites, like Facebook. I will be addressing concrete examples concerning social media by using Facebook as the social media network. Even though some scholars argue that social media and mass media are rather similar, this statement does not apply when it comes to regulating both of them. It is important to note that Facebook, Twitter, Instagram and other social media platforms, are not media companies, which is the case of the traditional mass media. They are technology companies, in general, however, some academics argue that they are also advertising agencies (Fuchs, 2019) or even data collection companies (Sneed, 2020). Due to this, the general rules and regulations for media regarding their content, advertising and publishing do not apply. According to Kaplan and Haenlein (2010), "Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" (Kaplan and Haenlein, 2010, p.61). As stated, one of the reasons why it is difficult to regulate anything that is posted and advertised on social media is the fact that all the regulations that existed before the rise of social media were designed for traditional mass media and the internet was exempt from these regulations. As van Dijck and Poell (2013) state, there is a social media logic and a mass media logic and in order to understand the difference between these two systems, and hence understand the difficulty in regulating these types of media, we need to understand the logic of both of them. It is expected that they will be different, as they come from a different political and technological descent. To get a better understanding, we will have a further look at the four main elements, as van Dijck

and Poell (2013) explain, that work together to create their so-called logic. These are programmability, popularity, connectivity and datafication (van Dijck and Poell, 2013).

According to van Dijck and Poell (2013), mass media has been programmed as a one-way flow, whose primary goal is to catch the user's attention, whereas social media programmability refers to a two-way traffic between users and programmers. This means that social media, as a user generated content, allows its users to publish content freely. On the other hand, the social media platforms adjust their algorithm to influence the data flow based on the information (data) they store about its users. This data is any activity the users perform on the social media platforms (e.g. liking, commenting, sharing) (van Dijck and Poell, 2013).

The second element of van Dijck and Poell's social media logic is popularity (2013). Mass media's popularity lies in their power to choose whom they give voice to and whom they do not. On the other hand, social media claims to be more democratic than this and gives voice to everyone equally (van Dijck and Poell, 2013). However, this is reasonably questionable since the algorithm that plays a significant part on social media platforms influences what content (e.g. topics, people, articles, posts) will be displayed at the top of each user's social media feed which, theoretically speaking, makes the popularity choice and decides what is more important for each user, similar to the system in mass media. Napoli (2014) identifies the popularity aspect of social media as being fundamental for the platform's development. He further points out that it is the main drive for the platform's algorithm (Napoli, 2014). It is important to note that this also affects the advertisements that are an essential part of the social media platforms, as that is how the platforms make revenue. When targeting users with their advertisement, the advertisers take advantage of the algorithm and either target them with tailored advertisements (Fuchs, 2019) or filter through the most popular users and pay them for a promotional job (van Dijck and Poell, 2013). Dutta-Bergman (2004) explains that social media does not necessarily have to compete with mass media, but merely adds a new dimension. It depends on the user's goal to choose which medium he or she turns to. Van Dijck and Poell (2013) explain that in the popularity aspect, social media only enhances mass media by adding a new dimension. Dogaru-Tulică supports the claim that social media is interlinked with mass media by arguing that politicians use social media to promote themselves and then, as a result, more people watch them on TV (2019). Based on these examples, we can see a clear correlation between mass media and social media. Social media announces someone as popular, mass media takes this information in and promotes this person on their media channels (van Dijck and Poell, 2013).

The third element is connectivity. Van Dijck and Poell (2013) further argue that the difference in mass media and social media logic is that mass media aims to connect people with content, whereas social media connects people with each other or with user generated content. Moreover, Welbers and Opgenhaffen (2018) use the phrase "going viral" (Welbers and Opgenhafen, 2018) and state that it is the shareability of social media that is its advantage. To understand this from an advertising perspective, mass media connects its advertisers to consumers. Social media, on the other hand, with its person-to-person connectivity with the addition of the algorithm strategy, allows advertisers to target a specific

audience (van Dijck and Poell, 2013). Social media platforms pride themselves on advocating individualism. However, van Dijck and Poell (2013) argue that this is not exactly as it seems. A social media user, either consciously or unconsciously, customises their online space by the actions he or she makes on the social media platform, due to the platform's algorithms. This also encourages an "automated personalisation", where the algorithm, based on the user's activity history, almost assumes the user's personality, which causes the platform's algorithm to customise the social media feed layout for the user (van Dijck and Poell, 2013). Van Dijck and Poell (2013) point out the users' opinions about the customisation of the social media feed based on their previous engagement with a social media platform and say that "some users appreciate the service offered by platforms to connect them to likeminded people, preferred items, or individualized taste; others loathe networked customization as a signal of intruded privacy or commercial exploitation of user information" (van Dijck and Poell, 2013, p.9).

Lastly, there is the datafication element. Datafication is the platform's ability to collect and analyse data about its users and then customise the feed to its users accordingly (van Dijck and Poell, 2013). Even though mass media has always aimed to produce content that was tailored to their target audience, social media has brought this element to the next level, where the information communicated on their platforms is much more customised for each individual user. Van Dijck and Poell (2013) explain that "what makes datafication a crucial characteristic for social media logic is its ability to add a *real-time data dimension* to mass media's notion of liveness" (van Dijck and Poell, 2013, p. 10). The data collected traditionally by mass media through polls and surveys are almost being replaced by metadata, which is a collection of data that holds information about the user's interaction with the platform. These are, for instance, the user's likes, comments and even the relationship status updates (van Dijck and Poell, 2013). Fuchs (2014) points out that the more time the user spends on social media the more data the platform collects about them.

These four elements that van Dijck and Poell (2013) present to us give us more clarity in understanding the differences between mass media and social media, which is important when looking at social media critically and finding the reasons why the already established rules for mass media do not work for social media, and perhaps why social media causes one scandal after another. It is important to note that from the initial plan that the CEO of Facebook, Mark Zuckerberg had, which was to create a platform that would connect people, has become a "multifaceted communication space, which not only allows people to interact by means of posting and commenting on statuses, but also features extensive advertising, as well as feed of trending topics […], all of which is also influenced by the personalisation algorithm" (Seargeant and Tagg, 2019, p.43).

3. The Facebook Algorithm

One feature that is dominating social media platforms is their algorithm. Algorithms can be understood as pieces of code that social media platforms use that "directly structure user

behaviours, impact preference information, and impact content production decisions" (Napoli, 2014, p.343). Algorithms are an important integration of social media and we see them being incorporated by more and more social media platforms, i.e. Facebook Newsfeed, Instagram Feed, Twitter's Timeline, etc. (DeVito et al., 2017). However, these sets of code have caused a lot of debate, whether it is in regard to the user privacy, manipulation, politics, advertising and even democracy. The Facebook algorithm creates a so-called "filter bubble" for the user and exposes them to data (e.g. posts, ads, news) tailored to each one of them personally (Sindermann et al., 2020).

The more these algorithms interfere with everyday decisions the more the question about their transparency gains importance (Shin and Park, 2019). Why these algorithms are especially dangerous is the fact that they are not transparent, and hence they are limiting the user's understanding of it (Skin and Park, 2019). This fact is another one of the reasons why it is difficult to regulate the content on social media platforms. Moreover, the technology companies refuse to make their algorithms transparent. On the contrary, the technology companies have the right to do so. Not only is the algorithm the platform's property but, as Napoli points out, the algorithm is being constantly updated to keep up with the current demands (2014), which is an additional reason that makes the regulating of social media content more difficult (van Dijck and Poell, 2013). Equally relevant to this issue is the argument of Gorwa et al. (2020) which states that "minimum standards for decision transparency are essential to allow both ordinary users and critical experts to understand the patterns of governance within which they are embedded" (p.11). Likewise, a study of the Facebook Newsfeed found that the more transparent the platform's algorithm is the more it increases the users' trust in that platform (Kizilcec, 2016).

The algorithm needs something that fuels its decisions. That will be the data it collects and stores about the social media users. What exactly is this data? On social media, everything is considered data. Essentially, the users themselves are data comprised of smaller data. A page is data, a like is data, user's relationship status is data, the user's demographic is data, the GPS location of the user's smartphone is data, user's behaviour is data, and so on (Fuchs, 2019). Lupton (2018) points out that these types of data, when gathered together, comprise Big Data. Therefore, Facebook's algorithm tracks its users' behaviour to collect data, based on which the algorithm customises each individual users' Newsfeed (Napoli, 2014). The more the social media platforms grow the more they have become some sorts of data enterprises. This is very useful for advertisers that are now able to tailor their ads to target specific audiences, using the data the technology companies have collected (van Dijck and Poell, 2013). To better understand how exactly the advertisers benefit from the algorithm and the data collection, we need to understand how the filter bubble works.

3.1. What Is a Filter Bubble?

Due to the algorithmic personalisation of social media platform feeds, so-called filter bubbles and echo chambers have been created for their users. Seargeant and Tagg identify

an algorithm that personalises the platform's experience for its users, a filter bubble (2019). This definition has two sides. One, where they argue that Facebook users can freely choose the actions they make on the platform; from liking what they want, commenting on what they want, sharing what they want, and even potentially clicking on the advertisements they want (Seargeant and Tagg, 2019). The definition has a second side, which raises a question about the amount of freedom the users actually have. According to Sindermann et al. (2020) the filter bubble is a pre-selected content for particular social media users. Therefore, considering the users are only presented with certain content in the first place and excluded from others, they are not completely free to choose what they want to see on their social media feed (Seargeant and Tagg, 2019). Facebook's algorithm is programmed to only show its user content it assumes they will be interested in, and hence will most likely interact with. This is essentially the business model of social media – to create the most engagement, so the users keep using their platforms (Seargeant and Tagg, 2019). Here, Sindermann et al. (2020) point out a number of issues, amongst which is the inability of the user to see the filter bubble, the inability of the user to choose what will be in their filter bubble and the inability of the user to choose whether they want to be in the filter bubble at all.

On the contrary, as already mentioned, the filter bubble that the algorithm creates for the user is based on their engagement history. This is a never-ending cycle, since the way the users interact on Facebook depends on their previous experience with the platform (Seargeant and Tagg, 2019). In addition to the user's engagement history, the algorithm also takes into account which content is more popular in general and pushes this content to the top of the user's feed to give it more importance and drive more engagement (van Dijck and Poell, 2013). According to Tredinnick and Laybats (2019) "the users have been transformed into the product" (Tredinnick and Laybats, 2019, p.93) and it is their minds that can be hacked with targeted advertisements particularly, and not the social media platforms.

Garrett (2017), based on his three-month observation study about Facebook users, supports Seargeant and Tagg's (2019) argument and acknowledges the fact that the platform's algorithm is looking for an engagement from its user. However, he is looking at it from a different point of view and argues that the algorithm's intention is not to purposely exempt the users from the content they do not get to see but merely wants to present them with the content they will most likely engage with instead. The Facebook users that were observed in Garrett's (2017) study, which was specifically aimed at exploring the effects of being in the filter bubble, said that they could surprisingly see more content that was promoting political views opposite to their own. Garrett (2017) explains, that if the problem merely lied in the algorithm's narrow selection of content it presented to the user, then there would be a simple solution of diversifying the algorithm. The filter bubble might be one part of the problem but another one is the well thought out strategies of the political disinformation campaigns (Garrett, 2017).

Being in the filter bubble not only enhances the already established beliefs and opinions of that particular user, but also causes a misleading anticipation that everyone 'thinks the same as me' (Sindermann et al., 2020). As explained, the reason for that is the fact

that the user is exposed to a small percentage of information and does not know that other information, perhaps contradicting his or her beliefs, exists (Garrett, 2017). On the contrary, Sindermann et al. (2020) argue that the algorithmic filtering of information is necessary as there is simply an excessive amount of information available to us.

3.2. The Impact of Facebook's Algorithm on News and Political Advertisements

Apart from many other things already mentioned, it has been widely discussed that the filter bubble distorts the user's political views and limits the exposure to the news in general (Sindermann et al., 2020). Due to this fact, there has been a fear about the effects of the filter bubble on democracy and claims that it influences voting (Sindermann et al., 2020). Anderson (2018) points out research which shows that the filter bubble causes the social media user to only be exposed to his or her own political views. Due to the nature of the filter bubble, the user's political views are even more amplified.

In the United States, almost 50% of the population turn to Facebook in search of news (Seargeant and Tagg, 2019). This makes Facebook one of the people's most preferred sources of news. As with any other content on social media, the news circulating on Facebook is difficult to regulate. On top of that, Facebook's users are only exposed to the information that appears in the filter bubble the algorithm has created for them based on the data it has collected about them. According to Sindermann et al., those users that consume news only via the social media networks are more likely to be consumed by the filter bubble than other individuals who seek news from traditional mass media as well (2020).

From the point of view of news and political values, being in the filter bubble means that the user only gets to see the content that supports their own civic and political beliefs and opinions. The same applies to political ads. Provided that the user has ever liked, shared or searched for anything on Facebook, as already discussed, the algorithm will remember it and store it as data (Seargeant and Tagg, 2019). Seargeant and Tagg (2019) argue that this essentially compromises the democratic side of the platform, which is supposed to be promoting the sharing of values. He further points out the benefits of being exposed to opinions contradicting the ones we have. People who are exposed to different opinions broaden their minds and become more tolerable, whereas people only presented with the views they agree with are less sensible to contrasting opinions (Seargeant and Tagg, 2019). Schäfer (2020) supports this statement with findings from her experiment. She states that the user's will to talk about and participate in a political debate increases, even if indirectly, when he or she is exposed to diversity of news.

The social media datafication element (van Dijck and Poell, 2013), the ability to customise advertisements and the social media algorithm have opened many new attractive possibilities for advertisers (Idan and Feigenbaum, 2019). Fuchs (2014) defines targeted advertising as a system, where advertisers pay companies like Facebook for the access to their data and then, based on this data, send specific advertisements to specific users to maximize

their profit. What does this mean for political advertising? The answer is the 2016 US presidential election, which I will be discussing later.

There are a number of reasons why a good political campaign strategy is effective at influencing people's opinions. One of these is the targeting of specific people that will see a particular advertisement, which is available on Facebook makes the advertisement more effective (Garrett, 2017). Even though a lot of people hold strong political opinions, there are many people that are not strongly opinionated but merely hold some weak theories. These people are easily influenced by a campaign that is presented to them and their already weak opinion is changed almost effortlessly by this advertisement (Garrett, 2017). Disinformation campaigns are especially successful because the tactics used in the disinformation campaigns are targeting emotions. To be even more specific, they use anger and mistrust to manipulate the public opinion (Garrett, 2017).

There has been a lot of debate about the impact of the algorithm and the filter bubble on political advertising and news. According to Andersen and Søe (2020), it is not the platform's programmer's problem to be dealing with issues in politics. Moreover, speaking of news, there are countless types of fake news; "fabrications, fakeness, falsity, lies, deception, misinformation, disinformation, propaganda, conspiracy theory, satire" (Andersen and Søe, 2020, p.6). For that reason, it is inefficient to deal with all of them in the same way. Furthermore, if a politician's intention is to propagate their persona, they do not care if the statement they share is true or false. In fact, they do not even need to know whether it is true or false. All that is important for them is the propagation of their persona (Andersen and Søe, 2020). Which brings up a question: is it feasible to be fact-checking political advertisements if politicians intentionally lie?

Andersen and Søe's (2020) explanation of fake news comes from the fact that fake news has always been and always will be around us and that we must learn to live with it. More specifically, they are referring to rumours, gossip and fiction. Tredinnick and Laybats (2019) support this claim and add that fake news has likewise been present in the traditional mass media. The different is the fact that social media helps to amplify the fake news much more. In reference to the discussions about fake news, where it is argued that thinking and implying that the world only consists or should consist of nothing but the truth, Andersen and Søe (2020) further point out that science fiction is not trying to be truthful but that does not mean that we should get rid of its existence. Andersen and Søe (2020) explain that these fictional stories are supposed to show us a different way of thinking, different realities, important values and norms, based on which we are supposed to judge or compare our reality with. They further argue that the research shows that fiction is everywhere around us.

Schäfer (2020) examines the types of news that are featured on social media and the effect they have on the user. To identify news shared on social media, Schäfer (2020) uses the term "snack news". She argues that political issues presented in these "snack news" are only damaging to the audience. The reason for this is the fact that "snack news" does not improve the factual knowledge of its readers. In addition, it has been proved that it distorts the perception of the users that are exposed to them. As already mentioned, a significant

percentage of people use social media to get their news. Even if they do not learn the actual facts from Facebook news, they claim to feel better informed by the fact that they are consuming information about politics in some form, which Schäfer (2020) explains as a general familiarity with the topic. Studies show, that users consuming news merely through social media are in reality less knowledgeable about politics than those who turn to other sources to read about political matters and do not use social media platforms for these purposes (Schäfer, 2020). Even though this applies to all social media news consumers, it is especially concerning in the case of those people who maintain more extreme opinions (Schäfer, 2020). Some researchers argue that what can be causing this effect is the user's friends resharing a post on a particular topic or multiple pages posting about the same issue. This is similarly amplified by the algorithm that is designed to promote posts that are the most engaged with by the users. Schäfer (2020) emphasizes that "citizens have to possess political knowledge to completely ensure proper representation of their interests" (Schäfer, 2020, p.3).

The detrimental effects of social media on politics have also been confirmed in the past two US presidential elections. Schäfer's (2020) experiment proves the fact that the more the users are exposed to "snack news" about one topic, the more it reinforces their perceived knowledge, regardless of the fact that some of the posts regarding that particular topic may not even be discussing concerns about the actual matter. Based on her experiment, Schäfer (2020) further argues that "it can be concluded that SNS [social network sites], but also snack news in general, has the potential to foster an illusion of knowing if users are frequently exposed to redundant information" (Schäfer, 2020, p.9).

4. The 2016 US Presidential Election

"is social media a tool of emancipation or is it a threat?" (Yerlikaya, 2020, p.178)

Political campaigns, fake news, disinformation, propaganda and all, have caused social media platforms quite a few scandals worldwide. It is important to note that scandals involving disinformation did not start at the rise of social media. Disinformation has been used to manipulate public opinion in the past (Garrett, 2017). From the hoax about life on the moon in 1835, through the oil industry's campaign to induce doubt about climate change in 1960s, the claims that the US actively participated in the 9/11 terrorist attack, to the more recent ones concerning social media and politics (Alcott and Gentzkow, 2017). Even though the 2016 US presidential election scandal involving social media was the one that drew the whole world's attention, it was in 2010 when social media first appeared to have an influence on politics, after "the Arab revolutions that began in Tunisia and spread to Egypt and Syria" (Yerlikaya, 2020, p.179).

Yerlikaya (2020) argues that certain actors specifically use social media to influence elections. It is not only the 2016 US presidential election that is the evidence. Elections influenced by foreign nationals and disinformation campaigns through social media happened

in France, Germany and Turkey and during the Brexit referendum (Yerlikaya, 2020). The evidence shows that Donald Trump used an enormous number of false, misleading and unsupported statements that circulated social media, especially Facebook, leading up to the 2016 US presidential election (Garrett, 2017). Another alarming fact about the 2016 election scandal was the involvement of the Internet Research Agency that was in close contact with Vladimir Putin which, likewise, used disinformation in the political advertisements on social media to manipulate people and hence, support Trump's victory as a president (Garrett, 2017). What's more, from as early as the year 2014, there has been evidence leading to the above mentioned Russian based company, that was creating fake American profiles to spread advertisements promoting Donald Trump and demoting Hilary Clinton (Beyersdorf, 2019). Facebook, Twitter and Google officially confirmed the fact that foreign nationals used their platforms to influence the elections, however, it was long after the election ended. These technology giants have also said that this was the start of the cyberwarfare (Beyersdorf, 2019). To get a better understanding of how serious the situation was, it is said that almost 130 million Facebook users saw the propaganda campaigns. Although Mark Zuckerberg stated at the beginning that he is of the opinion that Facebook's role in the 2016 election was only to help and that it was accomplished by giving voice to more people. On the contrary, it took Facebook eleven months to recognise the Russian interference with the election that took place on their social media platform and failed to cooperate with the US Congress (Beyersdorf, 2019).

It was not only the Russian interference that was the biggest scandal in the 2016 US presidential election. Another striking debacle involved a company called Cambridge Analytica, based in the United Kingdom, that likewise influenced the outcome of the 2016 election "by showing different advertisements on the same issue, to different people" (Rathi, 2017, p.1). The whistle-blower and a former Cambridge Analytica employee, Christopher Wylie, stated that the company helped Trump to get elected with the use of their microtargeted advertising (Rathi, 2017). Cambridge Analytica's microtargeted advertising is a perfect example of how van Dijck and Poell's (2013) elements of social media can be used, or misused, for these purposes. Facebook's datafication element allowed Cambridge Analytica to obtain data about its users. They did so by creating an app through which the Facebook users took a psychology test (Peters, 2020). As a result of how Facebook's privacy policy was designed, not only did the app allow them to access the data of the users who took the psychology test but also to the data of their Facebook friends unless they manually changed their privacy preferences in the settings (Rathi, 2017). Based on the data collected, Cambridge Analytica constructed personalities and, thus, sent different advertisements to different users (Rathi, 2017).

Even though some people say that the microtargeted advertisements were not used in the Trump campaign, it raises a concern of social media abilities and potential threats (Rathi, 2017). Although, that does not change the fact that the company then sold the data to other advertisers, such as the Leave.EU one, which then targeted users with their own campaign advertisements (Peters, 2020).

5. Social Media Content Regulation – The Case of Facebook

"[…] far-right and populist actors […] use social media to influence and manipulate the masses. The fact that social media is free from control and exempt from legal and economic constraints makes it more attractive for populist and far-right figures" (Yerlikaya, 2020, p.179).

Since the 2016 US presidential election debacle, there have been debates about the need for a stricter and more efficient social media regulation (Tandoc et al. 2020). Governments have been putting legislative measures in place (e.g. GDPR, The Communications Decency act Section 230, The Honest Ads Act), non-legislative organisations introduced more fact-checking and social media platforms themselves started coming up with new features to battle fake news (Tandoc et al. 2020). Perhaps the question is "what are social media users themselves doing to stop the spread of fake news?" (Tandoc et al. 2020, p.382).

5.1. Facebook's Content Moderation

5.1.1. The Platform Features

The regulation policy Facebook uses to moderate its content is called the Community Standards (Klonick, 2020). The Community Standards are the representation of Facebook's values, reflecting the culture of people who have initially created them, who were Americans. It goes without saying that, as a result, the Community Standards strongly depict the First Amendment (Klonick, 2020). Hence, as expected, Facebook's primary value is the freedom of expression, which comes first when the content moderator is deciding whether or not to remove specific content. After the freedom of expression, Facebook's secondary values come into place, which are "safety, privacy, authenticity and dignity" (Klonick, 2020, p. 2463).

The Community Standards had been private the entire time (even though Facebook's content regulation has been private since 2004, it was only in 2008 that some form of regulation policy started to be created at all) and were only made public in 2018 due to the immense pressure for transparency from many sides (Klonick, 2020). Not only did Facebook encounter a lot of criticism for not making their content moderation strategy public earlier, they were also criticised for what the strategy was. It was inexcusable that only a small number of people sitting in the office in Silicon Valley were responsible for decisions made about the content on a global level (Klonick, 2020). As Klonick (2020) further argues: "the rules for Facebook had always come from an American and Western tradition, which was part of the problem" (Klonick, 2020, p.2455).

However, it is not only the Community Standards that are flawed. Other aspects of Facebook's content moderation have pitfalls too. At the moment, Facebook's content is regulated half by artificial intelligence and half by human content moderators. Further, there

are features on the social media platform itself for the users that encourage them to participate in the platform's regulation. For instance, there is the flagging option that Facebook's users can use to report fake news or otherwise inappropriate content (Andersen and Søe, 2020). It is important to point out that most of the content that has been flagged by Facebook's users merely reflects their personal opinion or a clash of interests with another user (Andersen and Søe, 2020). Therefore, the flagged content is not very relevant for consideration for its removal, and hence, the flagging is not helpful. Eventually, Facebook decided not to progress further with their flagging function, as they have been informed by academics about the unintentional outcomes it can cause (Andersen and Søe, 2020). If a user sees that a particular post has been flagged, he or she will then think there must be something about the post that is true, since it has sparked so much attention, which thus enforces the user's belief in what the post promotes, which was the opposite of Facebook's intention for flagging (Andersen and Søe, 2020). As one of the solutions for Facebook's flagging strategy Garrett (2017) suggests that once the users flag a post, they should then be encouraged to inspect the corrections they have received and share it with the user who posted it originally. In addition, the information should be available and ready to be shared with the wider audience of the user that created the post in the first place.

As a response to Facebook's flagging option and the people fighting against fake news, Andersen and Søe (2020) argue that fake news is some sort of a "communicative action" that cannot be simply erased by fact-checking. They further claim that this way of tackling fake news does not delete the problem but it rather "raises serious social, cultural and political questions about the values, motives and ideologies inscribed in the algorithms designed to tag or flag fake news" (Andersen et al., 2020, p.2).

Another update that Facebook has introduced in order to fight fake news is an updated Newsfeed. Facebook claims that their updated Newsfeed algorithm is meant to make the content more authentic, which has been requested by its users (Andersen and Søe, 2020). In practice, it is up to the user to decide whether or not they believe what they see. Likewise, it is up to them whether they decide to take advantage of Facebook's features and report it as spam, tag it or flag it. What this essentially means is that if Facebook users do not recognise fake news, the algorithm understands that they want to keep seeing content like this, resulting in the new algorithm change having no effect (Andersen and Søe, 2020). The algorithm still continues working based on the engagement of the users and the popularity of certain posts. Considering the fact that the nature of fake news is to spread quickly and the nature of Facebook's algorithm is to put the most engaged with stories at the top of the users' feed, it can be argued that Facebook's recent adjustment to their algorithm, in reality, supports the spread of fake news (Andersen and Søe, 2020).

Furthermore, what Facebook is putting in place to battle fake news is a new button that would appear next to posts on the Newsfeed. The users will be able to click and see the story in context. This feature would likewise include relevant links that would explain the post in more detail (Andersen and Søe, 2020). The downside of this feature is the fact that the links

discussed above will be Facebook's and Wikipedia articles. Therefore, the disinterest of the links is quite debatable (Andersen and Søe, 2020).

Moreover, Facebook claims to have reduced the news that appears in their Newsfeed to only contain 4% of their overall content (Schäfer, 2020). The fact that Facebook is working on improving this problem is supported by findings in Andersen and Søe's (2020) article, in instances such as fake news terminology clarification or the button next to their posts on the Newsfeed. However, Schäfer (2020), based on her findings, argues that even though reading a full article provides the user with more facts and, hence, reinforces their knowledge, it does not enhance their feeling of being knowledgeable.

Certainly, there are areas, where Facebook is successful in moderating its content. In almost 100% of cases, Facebook has detected spam and fake accounts before the users did so. On the contrary, content such as hate speech was detected and removed by Facebook only about 16% of the time, before the user flagged it (Klonick, 2020).

It is without a doubt that Facebook is trying to fight against fake news in multiple ways. Andersen and Søe (2020) applauded Facebook for its take on fake news when Facebook de facto decided to abandon the term fake news altogether and exchanged it with more specific terminology. The new terms Facebook uses are *false news*, which stands for the news articles that claim to be factual but are intentionally spreading invalid information, and *disinformation*, which can be any content that is inaccurate, false or misleading (Andersen and Søe, 2020).

5.1.2. Artificial Intelligence

Half of Facebook's content is regulated by artificial intelligence. Zuckerberg states that artificial intelligence is the future of Facebook's content moderation (Gorwa et al., 2020). Although artificial intelligence has been involved in Facebook's content moderation for some time, instances such as the case of the livestream incident in Christchurch show it is not 100% reliable (Gorwa et al., 2020).

One of the artificial intelligence systems in place is the algorithmic moderation. This is a range of systems that identify content by its context and decide whether to take it down or not (Gorwa et al., 2020). That is accomplished by identifying, matching, classifying or predicting (Gorwa et al., 2020). Gorwa et al. (2020) emphasize that algorithmic content moderation is an essential asset for any platform to control their content, serve their users and advertisers, which makes them a feasible business.

On the contrary, advocates for academic human rights argue that if a platform's content moderation is fully automatic, it is dangerous (Gorwa et al., 2020). Gorwa et al (2020) explain this further by stating that sometimes it is difficult for automated systems to distinguish between complex contexts such as hate speech, where they tend to make a lot of mistakes. This can result in over blocking, for instance, which can be an issue in regard to politics. Another reason why this is particularly problematic is the fact that certain issues vary in context depending on which social group they are shared in (Gorwa et al., 2020).

The same applies for sharing and blocking terrorist content. Facebook claims that almost 100% of terrorist content is taken down by their algorithmic moderation. It is important to note that Facebook only reports the terrorist content that was taken down regarding Al-Qaeda and ISIS (Gorwa et al., 2020). What about other types of terrorism?

With all this in mind, how confident can we be that artificial intelligence will fairly analyse political advertisements, and based on what criteria? Would having a human in place of the artificial intelligence make any difference? Gorwa et al. (2020) are of that opinion, that merely having technological solutions in place for issues such as misinformation without a human control over it, should be thought about twice. Especially considering all the vulnerabilities the artificial intelligence has had so far (Gorwa et al., 2020).

5.1.3. Facebook Oversight Board

As a result of being pressured to be more transparent and responsible for moderating its content from different sides, Facebook announced in 2018 that it will create an Oversight Board, which will be something like a Supreme Court for Facebook (Klonick, 2020).

Once a user's content is removed, the user has to go through Facebook's procedures to request to get the content published again. When the user exhausts all the possible ways within Facebook, he or she can appeal about their removed content to the Oversight Board (Klonick, 2020). When it comes to removing content from the platform, the Board members' role is to control whether Facebook is complying to their policies and values. The Board members are supposed to represent the most diverse members of the society as possible (Klonick, 2020).

There are a number of challenges that Facebook is facing. Because Facebook was the one who established it and invested money in it, one of the issues is the questions about the Board's independence. Another one is the idea Zuckerberg has about the Board. According to Zuckerberg's idea, the Oversight Board will "make the final judgement call on what should be acceptable speech in a community that *reflects the social norms and values of people all around the world"* (Klonick, 2020, p.2474), which Klonick (2020) argues is far from being realistic, because there is no such thing as universal norms and values. Another key fact to remember is that Facebook's values already incline towards Western ones (Klonick, 2020).

There are various opinions about whether Facebook's Oversight Board will be efficient or not. The pessimistic one is that the Board will only lead to more content to be deleted and that Facebook will now have somebody to blame, when arguing about whose fault the removed content is (Klonick, 2020). Then there is the realistic one, which is that it simply will not work. Klonick (2020) points out that looking at the number of appeals that come through to Facebook about the removed content, it could possibly mean that the Oversight Board will have to review over 150 000 cases a day. One should note here that Zuckerberg stated that the Board will have somewhere between 11 to 40 members (Klonick, 2020). Nevertheless, there is also the optimistic opinion, which is about the efficiency of the Oversight Board. The optimists state that the Board will give Facebook users more transparency and fairness when

it comes to resolving issues about deleted content, since in the past this used to be about who you knew (Klonick, 2020).

5.1.4. How Does Facebook Regulate Its Advertisements?

Sneed (2020) sees the lack of clarity regarding whether Facebook is a media company or a technology company as the major problem in establishing effective advertising regulations. On the contrary, Klonick (2020), points out that Facebook is facing a very difficult and perhaps impossible task; to keep their business strategy private whilst meeting the demands for being transparent, and thus, maintaining the engagement from its users.

The problem is not that there are no regulations for political advertisements that Facebook adheres to. As already mentioned, one of the reasons why an effective regulation for political advertising on social media platforms has been difficult to establish is the fact that the already existing rules and regulations for advertising for mass media do not apply to social media (van Dijck and Poell, 2013). The US federal law for advertising states that "any broadcast, cable, or satellite communication, newspaper, magazine, outdoor advertising facility, mass mailing, or telephone bank to the general public, or any other form of general public political advertising" (Beyersdorf, 2019, p.1073) has to include a disclaimer in their advertisements. Generally speaking, it does not include the advertisements on the internet. Although, there are some advertisements on the internet the law does mention, these are the ones placed on other people's websites for a fee (Beyersdorf, 2019).

The technology companies argue that the law also lists exceptions for the disclaimers on "small items", such as pens, pins, buttons, etc. and any other advertisement that a placement of a disclaimer would be inconvenient (Beyersdorf, 2019). In 2011, Facebook was trying to find out whether any of their small advertisements fall into one of the exceptions categories or they are supposed to contain the disclaimer. Since half of the Federal Election Commission members said that Facebook's advertisements fall into the exceptions category and the other half said they do not, Facebook did not get a clear response. Thus, Facebook proceeded as if their advertisements belonged to the exceptions category (Beyersdorf, 2019).

Moreover, in 2006, the Federal Election Commission stated that Internet advertising does not pose any threat of corruption (Beyersdorf, 2019). Many activists from the technology industry have been fighting against a lot of the proposed regulations for social media platforms, which has led to social media platforms becoming highly underregulated. Clearly, there has been the fear of over-censorship when it comes to regulating political advertisements, which could be witnessed during the first week of Facebook running its updated content regulation policy. Political advertisements had been mixed up with political news which caused even harmless things, such as the British Royal Family, to be scored as being not compliant with this new policy (Sneed, 2020).

Some of the political advertisements were distributes by foreign nationals (Beyersdorf, 2019). One would say that the issue with foreign nationals influencing elections should be taken care of by the federal law that forbids foreign nationals to spend money on influencing

elections (Beyersdorf, 2019). Indeed, there is a law for that. However, this law does not forbid anyone to spend money on political advertising on the internet (Beyersdorf, 2019). Sneed (2020) argues that it is the US Congress that should require a label with each political advertisement on the internet. What has been happening is that the agencies responsible for political advertising send one type of an advertisement to one group of social media users and a different advertisement to another group of users, which merely causes confusion about the motive of the particular political party (Sneed, 2020).

Facebook announced that by the 2018 mid-elections a new system regarding political advertisements will be available that will require the advertiser to link the advertisement with a Facebook page (Beyersdorf, 2019). Facebook also claimed that it will keep an archive of all political advertisements relating to the election. This will include the information about the advertisements, money spent on them and what the target audience for the advertisement was. In addition, Facebook would require advertisers to disclose their location and identity and whether their advertisement is directly related to elections (Beyersdorf, 2019).

Nevertheless, Facebook does require a disclaimer from the advertisers before their advertisement is approved. However, if the advertisement includes a "paid for by" disclaimer, Facebook will approve it, regardless of what the individual puts in this field (Beyersdorf, 2019). Even though Facebook is trying to be more transparent, loopholes like this are still present. Therefore, it is necessary for the federal officials to get involved. According to Sneed (2020), this could finally allow for a consistent political advertising regulation on all social media platforms.

5.2. The Legislative Regulation of Social Media Platforms

One cannot deny that Facebook has been putting new measures in place to self-regulate their content as well as the political advertisements. However, the scholars argue that without an official law, the self-regulation of the social media platforms will merely be haphazard (Beyersdorf, 2019). There have been attempts to regulate political advertising and the obscene content on social media platforms by the government officials, such as The Honest Ads Act, the Section 230 in the Communications Decency Act or the European Union General Data Protection Regulation (GDPR) to the more recent formation of the research agency Social Science One, which examines the impact of social media on democracy (Klonick, 2020). The problem with these regulations has either been that they have not been fully developed and, therefore, have never been passed, or they have actually served the opposite purpose, such as the case of the Section 230 and the ongoing debate about its much-needed update, or they do not apply to every country like the EU GDPR. What Sneed (2020) stresses the most is that the US Congress should create clear boundaries and characterisation for political advertisements that would likewise not constrain the first amendment.

5.2.1. The Honest Ads Act

It has been suggested that in order to prevent foreign nationals from interfering with elections, the Honest Ads Act should be passed. The Honest Ads Act is a proposed bill that was introduced in 2017 concerning the transparency of online political advertising and the ways it can be regulated (Beyersdorf, 2019). Since the Federal Election Commission (FEC), which takes care of the disclosure of money spent on political campaigns and other disclosure requirements, does not involve advertisements placed in the digital media, the Honest Ads Act was meant to fill this gap (Sneed, 2020). Based on this Act, the identification of purchasers of online political advertising has to be revealed (Beyersdorf, 2019).

In 2014, the FEC's Vice Chair Ann Ravel pointed out the urgency of the FEC to establish regulations for online political advertising and in fact predicted the Russian involvement in the 2016 US presidential election. The more new kinds of online advertisement forms evolved, the more the FEC commissionaires could not agree on its regulations. One of them stated that "increased transparency in internet political advertising was censorship" and "could threaten the continued development of the internet's virtual free marketplace of political ideas and democratic debate" (Beyersdorf, 2019, p.1080). In 2017, due to the FEC's malfunction, Ann Ravel resigned (Beyersdorf, 2019).

Finally, it was decided that the most significant regulation of social media will be done by the platforms themselves by self-regulation (Beyersdorf, 2019). However, the three technology giants (Facebook, Google and Twitter), despite their self-regulating policies regarding political advertisements on their platforms, reached out to FEC for help to likewise contribute to the regulation. They pointed out an issue with placing disclaimers on the advertisements and argued that if the advertiser wanted to place their advertisement without a disclaimer or omit any of the policies put in place by the platforms, they would simply choose to use a different platform instead. Therefore, the idea regarding the disclaimers on the political advertisements would be ineffective (Beyersdorf, 2019).

In 2018, the FEC drafted a regulation applying to political advertisements placed on social media platforms by broadening the original statement "communications placed for a fee on another person's website" with "communications placed for a fee on another person's 'internet-enabled device or application'" (Beyersdorf, 2019, p.1087-1088). However, disclaimers are a much more complicated concept and this statement change was the only aspect that the commissionaires could agree on. Decisions like what kind of disclaimer would be placed on each type of the advertisement and what should be an exception, for instance, prevailed undecided and, therefore, the social media platforms remained self-regulated (Beyersdorf, 2019).

The Honest Ads Act was in reality introduced in 2017 by three Congress senators. The Act was meant to extend the regulation statement concerning online political advertising to apply to all paid political advertising online. It is divided into four parts, which consist of: the regulation of the capacity of online political advertisements, the guidance about disclaimers that need to be present in the political advertisements, the retention of all political

advertisements and information about them by the platform they have been placed on, and, lastly, the prevention of the involvement of foreign nationals from interfering with elections (Beyersdorf, 2019).

When establishing the restrictions, freedom of speech would not be violated, since the new rules employ a disclosure of fundamental information to voters, protect against corruption, and help to promote the campaign laws, which were the government's primary concerns (Beyersdorf, 2019). Disclaimers are especially important and legible in assessing the credibility of the advertisement and the advertiser. This aspect is critical, since a lot of the advertisements on the internet target and manipulate specific groups of people (Beyersdorf, 2019).

However, the Honest Ads Act has never been passed (Beyersdorf, 2019). Even though the technology companies have adopted the suggested policies of the Honest Ads Act, in order to arrive at an efficient regulation of all political advertisements in the online space, the urgency of the Act to be passed officially has been increasing (Beyersdorf, 2019). As already mentioned, Facebook approves any advertisement that includes a "paid for by" disclaimer, which confirms the fact that the platform's self-regulation about political advertising is considerably flawed (Beyersdorf, 2019).

As a matter of fact, if the Honest Ads Act was passed, apart from solving the flawed self-regulating policies of the technology companies, a legal action could be taken against the advertisers for violating any of the regulations. (Beyersdorf, 2019).

The Honest Ads Act is not the only way of stopping foreign nationals from influencing elections. On the contrary, the Honest Ads Act would not solve all the issues that surround foreign nationals using social media to influence elections. There are elements of social media like bots, trolls and fake accounts that are able to produce viral content, which likewise play a significant role in this matter and may even have a much greater impact than the online political advertisements (Beyersdorf, 2019). It is important to note that the Act does not cover content that is not an advertisement. Therefore, the Russian bots used to influence the 2016 election would not fall under the regulations of the Act. As Beyersdorf (2019) explains, "bots can form social media networks known as 'botnets', which are frequently hundreds of 'automated accounts built to follow and re-message one another'" (Beyersdorf, 2019, p.1095). For instance, if the bots are programmed to release specific fake news at the key times during the elections, it can create a false sense of consensus and authenticity. It is this pseudo group behaviour the bots create that intensify the effects on elections (Beyersdorf, 2019).

Another point to note is the loopholes the Honest Ads Act has. A foreign national can come to the US, buy a US SIM card and still purchase political advertisements on Facebook, stating that their identity is from the US (Beyersdorf, 2019). Moreover, Facebook confessed that developing a process that would be able to distinguish political advertisements from commercial advertisements might in reality be impossible (Beyersdorf, 2019). Nevertheless, the Act merely focuses on the political candidates and the mentions of elections in the advertisements. If the advertisements did not refer to any of these points but instead

expressed statements concerning social issues, these would not be regulated under the Honest Ads Act either (Beyersdorf, 2019).

5.2.2. Communications Decency Act Section 230

In 1996, one of the laws that the US Congress passed in order to regulate the profane and illegal content spreading on the internet was the Communications Decency Act (Ehrlich, 2002). With the digital world and technology always evolving, the Section 230 of this Act was passed to protect internet service providers (ISPs) that host user generated content from being held liable for this content or be treated as publishers, and thus, could not be sued for it (Hsia, 2016). The ISPs are defined as any online service that hosts third party content or allows access to multiple users (Hsia, 2016). Section 230 meant a complete immunity for the ISPs, resulting in them not wanting to make changes to any of the user generated content that may have been offensive, illegal or in any other sense defamatory. The reason for that being the fact that if the ISP made any changes to the user generated defamatory content, the content could be considered the content of the ISP (Hsia, 2016). As a result, Section 230 would not apply to them. Due to this, Section 230 states that an effort to remove defamatory content is encouraged and the ISPs will not become liable at any rate. On the contrary, the ISPs are encouraged to remove defamatory content, as a matter of fact, they are not legally required to (Hsia, 2016).

Hsia (2016) examined Twitter's removal requests for defamatory content. Even though the study shows that the US requests for the content removal was the fourth highest, Twitter did not remove any of the content requested for removal whereas other countries, such as Germany and the Netherlands, had almost a 100% removal rate. The US First Amendment, the freedom of speech, plays a very significant role in this case and does not allow for content to be removed. It can be argued "that the laws in Germany and the Netherlands are more effective" (Hsia, 2016, p. 428). The reason for that could be the GDPR regulation that the European Union has put in place in recent years (Sneed, 2020).

Meeran (2018) makes various suggestions how Section 230 could be amended to accommodate the current issues. Firstly, Meeran (2018) suggests that when a suit is filed on a content published on one of the ISPs, what the court should first consider is whether the ISP is in this case acting as the publisher or the speaker. Secondly, the court should examine whether the ISP is also acting as the information content provider or not. According to Meeran (2018), this way it would be much clearer to determine whether Section 230 applies. The problem faced here is that the court is inconsistent in determining the scope when the ISP is the information content provider partly, completely or not at all (Meeran, 2018).

Initially, the reason for the Communications Decency Act was to remove pornography from the Internet, and the purpose for Section 230 was to encourage the growth of technology start-up companies (Hsia, 2016). Fast forward more than twenty years and we live in a different world, Section 230 does not merely protect the ISPs but gives them absolute

immunity. The then start-ups, are now technology giants, and the law from 1996 needs to be amended accordingly.

5.2.3. General Data Protection Regulation (GDPR)

The European Union took steps to protect its nations' citizen's data and privacy on the Internet and in 2018 the EU General Data Protection Regulation (GDPR) came into effect (Bennett, 2018). It was the first big legislative regulation concerning the collection of people's data in the world (Bennet, 2018). The GDPR applies to anyone within the European Union. However, if an organisation outside of the EU collects or in any way monitors data about the EU residents, it still has to adhere to the GDPR (Bennet, 2018).

The GDPR lays out fundamental actions an establishment is required to take to lawfully collect or monitor someone's data if they fall into any of the previously mentioned categories. The main one being the user's consent, whose data is being processed (Bennet, 2018). Bennet (2018) points out that the consent has to be made consciously by the user's free-will and it cannot simply be a pre-ticked box. Scholars suggest that the user's approval to process their data could be incorporated from the EU GDPR in other countries as well (Sneed, 2020).

5.3. Non-legislative Regulation of Social Media

5.3.1. Fact-Checking

After the 2016 US presidential election the truth in politics has become an important matter (Hameleers and van der Meer, 2019). One of the non-legislative content moderation measures that has been available to tackle this issue are fact-checking companies. However, the 2016 election raised a question about the actual effectiveness of fact-checking (Sanders, 2018). Fact-checking directly focuses on checking political statements on the internet. Fact-checking is supposed to serve as a tool controlling politicians from sharing false claims (Lim, 2018). Likewise, it is meant to prevent false and misleading content from appearing on users' social media feeds, and hence protect them from being influenced by exposure to false claims (Lim, 2018). As Hameleers and van der Meer state, fact-checking is a new journalistic medium (2019). However, to distinguish fact-checkers from news reporters, fact-checkers' intention is not to come up with new information, but to merely correct the already existing one (Lim, 2018). There are different fact-checkers for different kinds of statements, but the two major ones that check statements from politicians that I will be examining in more detail are the Fact Checker and the Politifact (Politifact being the fact-checker particularly for Facebook) (Lim, 2018).

Blocking political content that is false and misleading and keeping the politicians accountable to the truth may sound like a wise and effective idea that the fact-checkers could provide. However, Lim (2018) argues that the fact-checkers more often than not disagree with their evaluations. Moreover, they do not check the same content. Hameleers and van der

Meer (2019) agree that the effectiveness of fact-checking is questionable, which they base on the evidence of a number of studies. They further point out the fact that the internet users can choose to opt out of seeing the fact-checking corrections. In addition, Politifact's and Fact Checker's scales for evaluating misleading content are different. Moreover, most of the politicians ignore fact-checkers and claim that they are biased (Lim, 2018). Trump, for instance, is a perfect example of a political candidate that does not respond to fact-checkers when they reach out to him for a clarification or a correction of his claims (Lim, 2018). It is not only this behaviour that is challenging for the fact-checkers but also the fact that politicians tend to be vague most of the time, which makes it difficult to fact-check their statements. Ultimately, politicians use this tactic to be purposely misleading (Lim, 2018). Lim (2018) further states that fact-checking these kinds of statements is especially valuable for the users. It is important to add that the vague statements are also the ones that are the most difficult to be fact-checked accurately (Lim, 2018). Moreover, Hameleers and van der Meer (2019) argue that the effectiveness of the fact-checkers likely depends on the user's political views, since the users tend to believe statements that correspond with their points of view but not the ones that are of an opposite opinion. According to Sanders (2018) "[p]eople know what they believe and go in search of facts to support those beliefs, and the internet delivers" (Sanders, 2018, p.19).

Sanders (2018) further argues that fact-checking is not meant to influence people's voting choice but is merely allowing people to see the truth before they decide what they believe. It is without a doubt that fact-checking has been a great addition to the lives of internet users. According to Lim (2018) multiple fact-checkers checking the same claim of a politician and arriving at the same conclusion has the biggest value. Fact-checkers do sometimes cross check each other's fact-checks, which they themselves consider important. However, the accuracy here is slightly problematic, especially in cross checking the facts from Politifact's with the Fact Checker. The reason for that is the way they evaluate the result. One fact-checker uses a six-point scale and the other a five- point scale (Lim, 2018). In an attempt to find some consensus among these two fact-checkers, Lim (2018) states that the final evaluation of the statement that is being fact-checked often depends on the fact-checker's subjective opinion. This is not in the case of true or false statements, but it rather relates to the statements which accuracy is evaluated on a scale of five or six. Lim (2018) further explains that "in many cases, while both fact-checkers agreed that the statement was inaccurate, they disagreed on how inaccurate it was, and this resulted in discrepancies in the ratings" (Lim, 2018, p.6). Nonetheless, Lim (2018) also points out another fact that plays a role in the result of the fact-checked statement, which is the varying resources the fact-checkers use to assess a statement.

Even though these inaccuracies happen, according to Lim's (2018) research, it is very rare that fact-checkers from Politifact and Fact Check cross check each other's facts. Lim (2018) emphasizes that if they did cross check each other's facts, it would be one of the biggest assets of fact-checking. Based on her research, she admits that even though it is

reasonably challenging to objectively assess a political candidate's truthfulness, fact-checkers have a potential and provide a big value to the public (Lim, 2018).

Andersen and Søe (2020) look at the fact-checkers' role from the advertisement point of view and explain that Facebook stated that they are not permitting the advertisements that link to posts or articles that have been indicated by fact-checkers as false. What's more, if advertisers attempt to run advertisements that link to information that has been fact-checked and indicated as false number of times, they will be banned from advertising on Facebook altogether (Andersen and Søe, 2020).

5.3.2. Public Campaigns

Apart from merely policing the social media platforms, Sneed (2020) suggests getting social media users educated about the usage of their data. Not only does one third of Facebook users not know how a free platform generates money, but nearly two thirds are not aware of the fact that Facebook keeps a list of their likes and personality traits (Sneed, 2020). Tredinnick and Laybats (2019) emphasize that even though an increased social media platform regulation does not cause a harm, what is critical in the era of ever-increasing disinformation on the internet is the education of the public about these matters.

Campaigns such as the ones that draw attention to the effects of smoking are effective in educating the public about this matter (Sneed, 2020). There could be campaigns spreading the awareness of online data sharing, which could even be in the form of a short TV commercial. Whereas the commercials about smoking are meant to discourage people from smoking, the purpose of public campaigns about the user's online data would be to raise awareness and educate (Sneed, 2020). In a world where social media has such a strong influence on people, the understanding of how someone's data is used online is very important. According to Sneed (2020) this matter should simply be incorporated into the curriculum in school, as a part of a computer class, for instance.

Although the federal agencies are making a big effort in educating the public about cybersecurity, social media needs to get more attention as well (Sneed, 2020). Another good example of a campaign that is meant to raise awareness is the one for cybersecurity: STOP. THINK. CONNECT by the Department of Homeland Security in the US, which is "encouraging users to think about what information gets posted to online accounts, who may be able to access it, and how those who access it might use it" (Sneed, 2020, p.126).

Perhaps the next solution to the social media content regulation is educating the public about the data shared online through public campaigns. Likewise, since people are the content creators responsible for what is shared on social media, they should be the ones being responsible in what they share. Public campaigns could be a good way to start.

6. Discussion

This paper sought to understand the social media content moderation and the limitations it is facing. A significant part was of the paper was dedicated to the political advertising regulations, with Facebook being the representative for the social media platforms. Social media is often thought of as an extension of traditional mass media (Dutta-Bergman, 2004). Even though they seem to be fairly similar, there are significant differences between them. Unlike traditional mass media social media hosts user generated content and the long-established regulation for mass media does not apply to social media (Kaplan and Haenlein, 2010). It goes without saying that all of the social media elements that van Dijck and Poell (2013) presented to us are some of the significant factors, due to which social media has difficulties in moderating its content.

Facebook claims to be advocating freedom of expression and its users can post content freely on their platform (Klonick, 2020). However, there are certain conditions under which this free posting on the social media platform is possible. One of the important features of majority of social media is the algorithm. Facebook algorithm places its users in a filter bubble and based on the data the algorithm has collected about them and assumed what they want to see, arranges the content on the user's feed (e.g. users' posts, news, advertisements) (Napoli, 2014). This includes showing the user particular content as well as not showing the user other content. As a result, when it comes to news and advertisements, the user's social media feed is surrounded by content that corresponds with his or her views. Not only the fact that the user lives in an online echo chamber merely surrounded by his or her values, views and opinion, but also the fact that the user turns into a product that is sold to advertisers (Tredinnick and Laybats, 2019).

Facebook operates as a free platform in exchange for its users' data (Sneed, 2002). To make a revenue, Facebook sells this data to advertisers that micro target Facebook's users with customised advertisements (van Dijck and Poell, 2013). Social media platforms, with its lack in regulation, are especially attractive to advertisers (Idan and Feigenbaum, 2019). As an example of the user data exploitation was the 2016 US presidential election that showed what social media platforms are capable of. Not only did the advertisers try to influence the users voting choices by micro-targeting them with their political campaign advertisements but the technology giants (Facebook, Google, Twitter) admitted that foreign nationals, one of them being Russia, used their platforms to spread fake news and disinformation in order to influence the outcome of the election (Beyersdorf, 2019).

There are multiple measures in place to regulate social media platforms. However, these measures have significant loopholes, due to which scandals on a global level are possible to escalate. Facebook moderates in content by human content regulators, artificial intelligence and features on the platform that encourage the users to flag in any form inappropriate content (Gorwa et al., 2020). The challenges the Facebook's content regulation is facing are; the artificial intelligence being ineffective in terms of context, a group of people sitting in the office in Silicon Valley making decisions on a global level and Facebook's platform flagging

options mostly being abused by the users' subjective opinions, thus, not helpful. Due to the demands of the public for Facebook's transparency and more responsibility, Facebook announced that it will create an Oversight Board that will be responsible for making decisions that reflect the world's norms and values (Klonick, 2020). Even though this step may seem like Facebook is taking all the accusations seriously, scholars argue that; first, there is no such thing as norms and values on a global scale, and second, that this is just a way of Facebook avoiding to be blamed (Klonick, 2020).

Apart from social media self-regulating their platforms, there are legislative and non-legislative regulations in place. Among the legislative ones, the European Union GDPR being referred to as the most effective one (Bennet, 2018). Scholars argue that the rest of the world should incorporate some of its procedures. Fact-checking, as the non-legislative regulation, is getting mixed opinions whether it is that effective or not (Lim, 2018).

Another solution to social media content regulation is the public awareness Sneed (2020). Where the legislative and non-legislative measures fail, perhaps it is time for the social media users to get educated and take the responsibility into their own hands in how they approach the time they spend on social media.

7. Conclusion

This paper analysed number of aspects of social media in regard to its regulation. The aspects discussed pointed to the differences between social media and mass media, the algorithm that collects data about the social media users and the environment it creates for them, the so-called filter bubble. The second part of this paper focused on the measures in place that regulate the social media content, the disinformation and fake news, and the influence on democracy via foreign nationals misusing social media platforms and micro-targeted political advertisement.

References

Allcott, H. and Gentzkow, M., 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives,* [e-journal] 31(2), pp.211-36. DOI: 10.1257/jep.31.2.211.

Anderson, K., 2018. How to Win an Election: An Analysis of the Power of Social Media in Politics. *Public Interest Law Reporter,* [e-journal] 24(1), pp.96-100. Available at: <https://lawecommons.luc.edu/cgi/viewcontent.cgi?article=1565&context=pilr> [Accessed 7 September 2020].

Andersen, J. and Søe, S. O., 2020. Communicative Actions We Live By: The Problem with Fact-Checking, Tagging or Flagging Fake News – the case of Facebook. *European Journal of Communication,* [e-journal] 35(2), pp.126-139. DOI: 10.1177/0267323119894489.

Bennet, C. J., 2018. The European General Data Protection Regulation: An instrument for the globalization of privacy standards?. *Information Polity,* [e-journal] 23(2), pp.1-8. DOI: 10.3233/IP-180002.

Beyersdorf, B., 2019. Regulating the "Most Accessible Marketplace of Ideas in History": Disclosure Requirements in Online Political Advertisements After the 2016 Election. *California Law Review*, [e-journal] 107(3), pp. 1061-1100. DOI: 10.15779/Z38C53F20R.

Couldry, N., 2015. The myth of "us": digital networks, political change and the production of collectivity. *Information, Communication & Society,* [e-journal] 18(6), pp.608-626. DOI: 10.1080/1369118X.2014.979216.

DeVito, M. A., Gergle, D. and Birnholtz. J., 2017. "Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems,* [e-journal], pp.3163-3174. DOI: 10.1145/3025453.3025659.

Dogaru-Tulică, A., 2019. The Personalisation of Political Communication in Social Media Era. Case Study: Gabriela Firea, The Mayor of Bucharest. *The Youth of Today - The Generation of the Global Development,* [e-journal] 14(1), pp. 369-376. Available at: <http://www.proceedings.univ-danubius.ro/index.php/eirp/article/view/1940> [Accessed 7 September 2020].

Dutta-Bergman, M. J., 2004. Complementarity in Consumption of News Types Across Traditional Media and New Media. *Journal of Broadcasting & Electronic Media,* [e-journal] 48(1), pp.41-60. DOI: 10.1207/s15506878jobem4801_3.

Ehrlich, P., 2002. Communications Decency Act Section § 230. *Berkeley Technology Law Journal,* [e-journal] 17(1), pp.401-419. DOI: 10.2307/24120113.

Fuchs, C., 2014.Critique of the Political Economy of Information Capitalism and Social Media. In: Fuchs, Christian and Sandoval, M. (ed.) *Critique, social media and the information society* [online] New York: Routledge, pp. 51-65. Available at: <http://fuchs.uti.at/wp-content/FuchsSocM.pdf> [Accessed 24 July 2020].

Fuchs, C., 2019. Social Media, Big Data, and Critical Marketing. In: Tadajewski, M., Higgins, M., Denegri-Knott, J. and Varman, R. ed(s). *The Routledge Companion to Critical Marketing*. [online] London: Routledge, pp. 467-481. Available at: <https://doi.org/10.4324/9781315630526-29> [Accessed 24 July 2020].

Garrett, R. K., 2017. The "Echo Chamber" Distraction: Disinformation Campaigns are the Problem, Not Audience Fragmentation. *Journal of Applied Research in Memory and Cognition,* [e-journal] 6(4), pp.370-376. DOI: 10.1016/j.jarmac.2017.09.011.

Gorwa, R., Binns, R. and Katzenbach, Ch., 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*. DOI: 10.1177/2053951719897945.

Hameleers, M. and van der Meer, T. G. L. A., 2020. Misinformation and Polarization in a High-Choice Media Environment: How Effective Are Political Fact-Checkers?. *Communication Research*, [e-journal] 47(2), pp. 227–250. DOI: 10.1177/0093650218819671.

Himanshu, G., 2019. *Is Facebook really concerned about privacy?* [online] Available at: <https://www.cjr.org/tow_center_reports/facebook-merges-encrypted-messages.php> [Accessed 22 May 2020].

Hoofnagle, C. J., van der Sloot, B. and Zuiderveen Borgesius, F., 2019. The European Union general data protection regulation: what it is and what it means. *Information & Communications Technology Law*, [e-journal] 28(1), pp.65-98, DOI: 10.1080/13600834.2019.1573501.

Hsia, J., 2016, Twitter Trouble: The Communications Decency Act in Inaction. *Columbia Business Law Review,* [e-journal] *2017*(1), pp.399-452. DOI: 10.7916/cblr.v2017i1.1718.

Idan, L. and Feigenbaum, J., 2019. Show me your friends, and I will tell you whom you vote for: Predicting voting behaviour in social networks. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,* [e-journal], pp. 816-824. DOI: 10.1145/3341161.3343676.

Jørgensen, R. F. and Zuleta, L., 2020. Private Governance of Freedom of Expression on Social Media Platforms. *Nordicom Review,* [e-journal] 41(1), pp. 51-67. DOI: 10.2478/nor-2020-0003.

Kaplan, A. M. and Haenlein, M., 2010. Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons,* [e-journal] 53(1), pp.59-68. DOI: 10.1016/j.bushor.2009.09.003.

Kizilcec, R. F., 2016. How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems,* [e-journal], pp.2390-2395. DOI: 10.1145/2858036.2858402.

Klonick, K., 2020. The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *The Yale Law Journal*, [e-journal] 129(8), pp.2418-2499. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3639234> [Accessed 23 August 2020].

Lim, Ch., 2018. Checking how fact-checkers check. *Research & Politics.* [e-journal], DOI: 10.1177/2053168018786848.

Lupton, D., 2018. How do data come to matter? Living and becoming with personal data. *Big Data & Society,* [e-journal] 5(2). DOI: 10.1177/2053951718786314.

Meeran, Y. A., 2018. As justice so requires: Making the Case for Limited Reading of Sec. 230 of the Communications Decency Act. *George Washington Law Review,* [e-journal] 86(1), pp.257-286. Available at: <http://heinonline.org/HOL/Page?handle=hein.journals/gwlr86&div=9> [Accessed 12 August 2020].

Napoli, P. M., 2014. Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption. *Communication Theory,* [e-journal] 24(3), pp. 340-360. DOI: 10.1111/comt.12039.

Peters, M. A., 2020. Platform ontologies, the AI crisis and the ability to hack humans 'An algorithm knows me better than I know myself.' *Educational Philosophy and Theory,* [e-journal] 52(6), pp.593-601. DOI: 10.1080/00131857.2019.1618227.

Peterson, K., 2018. GDPR: What (and Why) You Need to Know about EU Data Protection Law. *Utah Bar Journal,* [online] 31(4), pp.12-17. Available at:

<https://www.kmclaw.com/media/article/247_July_Aug_2018_Peterson_Data_Protection.pdf> [Accessed 7 September 2020].

Rathi, R., 2017. *Effects of Cambridge Analytica's Facebook ads on the 2016 US Presidential Election* [online] Available at: <https://towardsdatascience.com/effect-of-cambridge-analyticas-facebook-ads-on-the-2016-us-presidential-election-dacb5462155d> [Accessed 03 September 2020].

Sanders, K., 2018. Fighting Fake News. *The Quill,* [e-journal] 106(1), pp. 16-21. Available at: <https://www-proquest-com.ezproxy.lib.bbk.ac.uk/docview/2051739559/fulltextPDF/E76E2F3132AA4B17PQ/1?accountid=8629> [Accessted 7 September 2020].

Schäfer, S., 2020. Illusion of knowledge through Facebook news? Effects of snack news in news feed on perceived knowledge, attitude strength, and willingness for discussion. *Computers in Human Behaviour*, [e-journal] 103(2), pp.1-12. DOI: 10.1016/j.chb.2019.08.031.

Seargeant, P. and Tagg, C., 2019. Social media and the future of open debate: A user-oriented approach to Facebook's filter bubble conundrum. *Discourse, Context & Media,* [e-journal] 27(3), pp.41-48. DOI: 10.1016/j.dcm.2018.03.005.

Shin, D. and Park, Y. J., 2019. Role of Fairness, Accountability, and Transparency in Algorithmic Affordance. *Computers in Human Behaviour*, [e-journal] 98, pp. 277-284. DOI: 10.1016/j.chb.2019.04.019.

Sindermann, C., Elhai, J. D., Moshagen, M. and Montag, C., 2020. Age, gender, personality, ideological attitudes and individual differences in a person's news spectrum: how many and who might be prone to "filter bubbles" and "echo chambers" online?. *Heliyon,* [e-journal] 6(1). DOI: 10.1016/j.heliyon.2020.e03214.

Smith, A., 2020. What Is Section 230? How Trumps Fight with Twitter Could Fundamentally Change the Way Internet Works. *Independent.* [online] Available at: <https://www.independent.co.uk/life-style/gadgets-and-tech/features/section-230-what-is-trump-twitter-executive-order-social-media-internet-a9538681.html> [Accessed 14 August 2020].

Sneed, M., 2020. The Key to Regulating Facebook and Data Collection Companies is Transparency, *Albany Law Journal of Science & Technology,* [e-journal] 30(1), pp.109-136. Available at: <http://heinonline.org/HOL/Page?handle=hein.journals/albnyst30&div=8> [Accessed 12 August 2020].

Stefanone, M. A., Lackaff, D. and Rosen, D., 2010. The Relationship Between Traditional Mass Media and Social Media: Reality television as model for social network site behavior. *Journal of Broadcasting & Electronic Media*, [e-journal] 54(3*),* pp.508-525. DOI: 10.1080/08838151.2010.498851.

Tandoc, E. C., Lim, D. and Ling, R., 2020. Diffusion of disinformation: How social media users respond to fake news and why, *Journalism*, [e-journal] 21(3), pp. 381–398. DOI: 10.1177/1464884919868325.

Tredinnick, L. and Laybats, C., 2019. Reality Filters: Disinformation and Fake News. *Business Information Review,* [e-journal] 36(3), pp.92-94. DOI: 10.1177/0266382119874267.

Van Dijck, J., 2020. Governing digital societies: Private platforms, public values. *Computer Law & Security Review,* [e-journal] 36(4). DOI: 10.1016/j.clsr.2019.105377.

Van Dijck, J. and Poell, T., 2013. Understanding Social Media Logic. *Media and Communication,* [e-journal] 1(1), pp.2-14. DOI: 10.17645/mac.v1i1.70.

Welbers, K. and Opgenhaffen, M., 2018. Presenting News on Social Media. *Media Login in the Communication Style of Newspapers on Facebook,* [e-journal], pp.45-62. DOI: 10.1080/21670811.2018.1493939.

Yerlikaya, T., 2020. Social Media and Fake News in the Post-Truth Era. *Insight Turkey,* [e-journal], 22(2), pp.177-196. DOI: 10.25253/99.2020222.11