

# Democracy, Social Media and Viral Lies

An comparative framing analysis of three influential actors in  
disinformation policy in the Dutch digital landscape



Gido van Os  
Student ID: 1065293  
MSc Thesis Human Geography  
Conflicts, Identities & Territories

7 July 2021  
Word count: 21.287  
Supervisor: dr. O.T. Kramsch  
Second reader: dr. N. Stel

Cover image: Album cover of “Revolutions per Minute” by Rise Against (2003).

## **Abstract**

Since 2016, the issue of disinformation on social media has become an urgent topic. Scholars, policy makers and experts argue that this phenomenon can form a substantial threat to contemporary democracies, undermining trust in established institutions and electoral processes. Yet in current research on disinformation, the connection between theory and practice is often absent. Therefore, in order to establish a first step in investigating what kind of anti-disinformation policies would be fair, effective and broadly supported, this study takes a closer look at three important actors that influence the debate on a practical level. These three actors in the Dutch digital landscape are governmental institutions, social media platforms and relevant NGO's. Subsequently, this comparative qualitative research uses framing theory to operationalize and analyze how these actors perceive the problem of disinformation on social media, how they perceive its solutions, and how these solutions are motivated and legitimized. This study finds two major similarities in used framings – concerning the nature of the problem and the focus on media literacy – and four major differences – concerning the cause of the problem and perspectives on transparency, content moderation and freedom of expression.

## Acknowledgements

There is no way to get around it: the past year has been rather “unusual” to say the least. Like many things this year, this Master’s thesis is almost entirely produced at home. Hence, conducting this research and writing the thesis report felt like a complete solo-endeavor at many times, but the presence of loving friends and family reminded me that we are all in this together. I couldn’t have done it without them. It’s cliché but it’s true.

I want to thank my supervisor, Olivier, for giving me a lot of space and trust during the last six months. This certainly gave me extra confidence to explore my own path during my thesis journey. You understood that such a project – especially during times of a worldwide pandemic – does not merely pose an intellectual challenge but also a mental one. Your assurance and calmness helped me to go through this period almost stress-free.

Thank you Eline, Marte, Bart and Charlie for reading some of my early drafts. Your comments and critiques inspired me to explore new perspectives and helped me to steer this project in the right direction. Thank you Peter, for inviting me into the tiny two-seated office inside your student dorm. On that note, I also want to thank all my other house mates for being such great friends. With countless coffee breaks, shared lunches and sports intermezzos in our small but cozy flat we always kept each other motivated while working and studying from home.

Last but not least, I want to thank those who participated in this research. Thank you for making time in your busy schedules for me. I enjoyed our conversations and you provided me with great insights.

The cover of this thesis features the album cover of one my favorite albums of all time. Its image fitted the topic of this thesis particularly well. Now, as this thesis also couldn’t have been written without the presence of great music in my ears – as a passionate music fan I do not like to sit in silence – I want to end this section with a chorus written by Rise Against, a band that has always been at my side since I was 14 years old. Its lyrical sentiment certainly resonates with the societal problem that this thesis reflects upon...

“Can we be saved, has the damage all been done? – Is it too late to reverse what we've become? – A lesson to learn at a crucial point in time – What's mine was always yours, and yours is mine.”<sup>1</sup>

---

<sup>1</sup> “Chamber the Cartridge” by Rise Against (2006).

# Table of contents

<b>1. Introduction: the Infocalypse</b> .....	<b>1</b>
1.1 Social media and (geo)political discourse.....	2
1.2 From diagnosis to treatment.....	2
1.3 Academic relevance .....	3
1.4 Synthesis .....	4
<b>2. Literature review</b> .....	<b>6</b>
2.1 Definitions: “misinformation” versus “disinformation” .....	6
2.2 Psychology: the cognitive stickiness of misinformation.....	6
2.3 Psychology meets policy: the effectiveness of counter-measures .....	7
2.3.1 Algorithms & content moderation.....	7
2.3.2 Fact-checking & labelling.....	8
2.3.4 Synthesis .....	8
2.4 Framing theory .....	9
2.5 Research questions.....	10
<b>3. Methodology</b> .....	<b>12</b>
3.1 Data sampling.....	12
3.2 Data collection.....	14
3.3 Data analysis .....	15
3.4 Ethical considerations .....	15
3.5 Methodological limitations .....	15
<b>4. Governmental institutions</b> .....	<b>17</b>
4.1 National level: Ministry of Domestic Affairs (BZK) .....	17
4.2 International level: European Commission (EC) .....	20
<b>5. Social media platforms</b> .....	<b>24</b>
5.1 Facebook.....	24
5.2 Twitter .....	28
<b>6. Internet Freedom Organizations</b> .....	<b>31</b>
<b>7. Discussion</b> .....	<b>35</b>
<b>8. Conclusion</b> .....	<b>38</b>
<b>References</b> .....	<b>39</b>
<b>Appendix I – Data overview</b> .....	<b>45</b>

“A lie gets halfway around the world before the truth has a chance to get its pants on.”

– Winston Churchill

# 1. Introduction: the Infocalypse

In the fragmented media landscape of today, debates on the honesty and truthfulness of political information has increasingly become more prominent, both in the public sphere and in the academic world (Hameleers et al., 2020; Weeks & Gil de Zúñiga, 2021). As many scholars have argued, the health of a democracy depends on many variables, and a well-informed and educated populace is a crucial one (e.g. Kuklinski et al., 2000; Lewandowsky et al., 2012). Ideally, citizens should have access to accurate information and hold well-informed opinions – e.g. on which political actors are involved in their democracy, how their political aspirations actually relate to the challenges facing the state, and what the rules of the political game are – which guide them in their participation in civil society (van Aelst et al., 2017). However, the recent epidemic spread on political disinformation threatens to sabotage these very foundations on which contemporary democracies are built (Brown, 2018). Back in 2012, Lewandowsky et al. warned in an influential paper that “[i]f a majority believes in something that is factually incorrect, the misinformation may form the basis for political and societal decisions that run counter to a society’s best interest (...) and can have serious consequences” (p. 107). Moreover, they argued that the internet would have a crucial role in the spread of political disinformation, by changing the media landscape in a unprecedented way. In retrospect, they hit the nail on the head.

Although many experts acknowledge that the surge of political disinformation on the internet is a deeply concerning development, most contemporary digital ecosystems are still deeply vulnerable to the spread of disinformation. Tristan Harris, co-founder of the *Centre for Humane Technology* and main protagonist in the documentary *The Social Dilemma*, often brings up a powerful analogy to illustrate the massive differences in national security doctrines in the physical world versus the digital world. He points out that in the physical world, if a rogue nation sends a warplane into the United States airspace, or a lone terrorist attempts to enter the country, they are likely to be stopped by billions of dollars’ worth of infrastructure protecting the borders – ranging from passport controls to border walls and air defence systems. This network of infrastructure is carefully set up to make sure that the country is safeguarded from any perceivable threat. Yet in the digital world the borders are left almost completely open. To make matters worse, if a bad actor tries to drop an information bomb into the digital ecosystem of the United States, it is met by a Facebook algorithm that asks: “OK, what zip code do you want to target?” (Harris, 2021).

The weaponization of information has been around for centuries, but the Digital Age and the rise of social media platforms has drastically increased the speed and reach in which false information can circulate, thereby increasing its impact (P. N. Howard, 2020; Lazer et al., 2017; Zeitzoff, 2017). Disinformation expert Nina Schick dubs this “the Infocalypse”. In recent years, growing numbers of digital bombs are used by various actors – ranging from foreign nation states to lone domestic actors – to influence political discourse (Schick, 2020). These bombs are not equipped with an explosive charge of ammonium nitrate, but with false information shared via social media platforms and shady web domains, designed to be provocative, polarizing and misleading – which can be equally destructive to contemporary democracies (Howell, 2013; P. N. Howard, 2020; Swire-Thompson et al., 2020). There is no shortage of recent examples. Notable ones are the Russian disinformation campaigns via social media, targeting the 2016 US presidential elections (Allcott & Gentzkow, 2017), the Brexit referendum (Bastos & Mercea, 2019; The Economist, 2017) and the war in Eastern Ukraine (Chen, 2015; Mölder & Sazonov, 2019). Disinformation has influenced elections in other major democracies, such as in France (Ferrara, 2017), India (Jain, 2019) and South Korea (Keller et al., 2020). Last year, research confirmed that the Netherlands has also been targeted by spreading geopolitical disinformation using the Dutch language. Indonesian fake accounts on social media deliberately spread falsehoods about the separatist movement in West-Papua in order to manipulate the public attitude towards the movement (NOS, 2020). According to a study by I&O research (2017), 82% of Dutch citizens sees disinformation as a threat to democracy and rule of law.

## 1.1 Social media and (geo)political discourse

In many cases, social media platforms takes the centre stage in the spread of disinformation. Once it has entered the digital ecosystem, it proceeds to live a life on its own and spreads organically through platforms such as Facebook and Twitter. Citizens themselves may unknowingly participate in their own disenfranchisement by using social media to consume, distribute or even generate false information, thereby granting it more legitimacy and authority (Mejias & Vokuev, 2017). Hence it has been argued more and more in the last decade that “[s]ocial media increasingly constitutes an important arena not just for political communication, but also for political conflict” (Barberá & Zeitzoff, 2018). The next section presents a brief historical perspective on the intersection of social media and (geo)political discourse.

During the mid to late 1990’s – when the World Wide Web was still in its infancy, making her first steps towards global dominance – scholars of geopolitics already questioned how rapid ICT developments would change the nature of political relationships and discourse. Hence Manuel Castells (1996, p. 476) noted that “politics has to be framed in the language of electronically based media, [which has] profound consequences on the characteristics, organizations, and goals of political processes, political actors, and political institutions”. In other words, he saw back in the 1990’s that the introduction and diffusion of new technologies were going to fundamentally change the roles of states and citizens in society. New ICT developments could bridge gaps between traditional spaces and reduce spatial differences, thereby having the potential to challenge the status quo (Brunn, 1998). The growing prominence of digital space has been “changing the role of states, those residing within state boundaries, and regional and global governmental organizations” (ibid, p. 108). These changing circumstances thus require new forms of governance, where its rules need to traverse traditional legislative borders and need to be constantly updated. In short, physical spaces and digital spaces are intertwining in numerous dimensions – politically, socially, economically – and are increasingly harder to separate (Herrera, 2008).

At the brink of the 2010’s, social media was becoming a vital part of the digital ecosystem. Initially this development was cheered at – by scholars, the public and obviously the technology companies themselves – mainly because of its promise to democratize conversation online, thereby realizing real-world political changes (Ferrara, 2017; Schiffrin, 2017). For instance, social media received a lot of public and academic praise for its role during the Arab Spring (e.g. Howard et al., 2011) and other protest movements around the world. Hence it was argued at the time that any regulation of social media would not be desirable, given the democratizing potential these new “tools” brought to the world (Shirky, 2011). However, this positive image of social media has drastically shifted since then. Although some researchers did warn about the potential for abuse of social media for political disinformation during that early rosy period (such as Hwang et al., 2012 and Morozov, 2011), the general narrative definitely turned net-negative after disturbing stories of excessive online political disinformation – such as those discussed in the introduction – drew increasing international attention during the mid- to late-2010’s (Hameleers et al., 2020).

## 1.2 From diagnosis to treatment

So, during the last five or six years, many people started to realise that these developments on social media might actually pose a substantial threat to democracies around the world. This has now repeatedly been confirmed by researchers, journalists and experts. Thus one could argue that – when it comes to disinformation on social media – society has now moved beyond the *diagnosis* phase and entered into the *treatment* phase. Today, experts and policy makers are mainly thinking about the ways in which this problem can actually be solved. What steps can be taken to minimize the damaging effects of disinformation?

The answer to this question is debated by three major actors in society (Ganesh & Bright, 2020). First, large social media platforms such as Twitter, Facebook and Google now acknowledge this problem – due to both internal reflection and public pressure – so they have been experimenting with various measures to

minimize the spread of disinformation on their platforms. Examples of such measures are the suspension of specific accounts and communities (Roose, Isaac & Frenkel, 2021), the addition of warning labels next to posts that may include false information (Roth & Pickles, 2020) and the tweaking of algorithms to demote posts that contain disputed claims (Paul, 2020). Second, governmental institutions have also noticed the significance of this problem. Hence the European Commission (EC) has designated expert groups to generate feasible policy recommendations to minimize the effects of disinformation in Europe's digital ecosystem (European Commission, 2018, 2020a). In addition, the European Digital Media Observatory was established in June 2020, which aims at "creating and supporting the work of an independent multidisciplinary community, capable of contributing to a deeper understanding of the disinformation phenomenon and to increase societal resilience to it" (European Commission, 2020b). On the national level, the Dutch Ministry of Domestic Affairs (BZK) is formally aware since 2017 that disinformation on social media poses a threat to Dutch democracy and the integrity of its elections in particular (Kamerstuk 26.643, nr. 508, 2017). The subject has come up in several parliamentary questions and plenary debates from 2017 to 2021, and this led to some specific measures against disinformation taken by BZK (Kamerstuk 30.821, nr. 91, 2019). Third, there is a group of smaller actors – representing a part of civil society – that are wary of any set of rules that may jeopardize free speech on the Internet. Social media platforms regularly cause controversy and pushback in their anti-disinformation efforts, for example when certain accounts or pages are suspended. These actions can be perceived as deliberate "silencing" of parts of society, or even as censorship. This side of the argument can be found in various parts of the online realm – e.g. in the YouTube comment sections or in forum threads – but it is formally voiced by non-governmental organizations such as Bits of Freedom, Article19 and the Electronic Frontier Foundation.

These three actors frame the problem of disinformation on social media in their own way, and subsequently play a role in how the problem is perceived and in deciding what would be the best way forward. It is therefore important to reflect on how these different actors – i.e. social media platforms, government institutions and civil society organizations – perceive the problem and its potential solutions, and investigate who *they* think should have the prime responsibility to act. Yet when it comes to establishing policies against disinformation, there is a delicate balance between protecting freedom of expression and establishing effective regulation, between intentional harm and accidental ignorance, and between fact and fiction. These treacherous grey areas are still to be explored if we are to make progress on this problem. Hereby we could eventually establish evidence-based disinformation policies that are both effective and uncontroversial.

### **1.3 Academic relevance**

Due to the novelty of social media, research on the intersection of political conflict and social media has only been present for a decade. During this period, most research in the field of conflict studies focused on online (de)radicalization via social media (Bilazarian, 2020; Conway, 2017). Many of these studies tended to emphasize Islamic extremism – especially since the rise of the Twitter-savvy ISIS (Awan, 2017; Benigni et al., 2017) – but recently there has been more interest in the way that social media draws individuals into political extremism (Hodge & Hallgrimsdottir, 2020; Nouri & Lorenzo-Dus, 2019). Polluting the digital ecosystem with outrage-fuelling disinformation seems to be the very mechanism by which extremist ideas spread most efficiently (Ganesh & Bright, 2020). The intentional spread of disinformation through social media pages has repeatedly and recently been linked to populism, extremist groups and various political parties from the radical right (Baele et al., 2020; Bennett & Livingston, 2018). Additionally, certain inherent features of social media – such as its business model of targeted advertisements – are known to contribute to polarization and social unrest (Stewart et al., 2018; K. Thorson & Wells, 2016). Hence there is a link between contemporary disinformation on social media and political conflict, which needs more exploration and understanding. Although research on disinformation has exploded in the last few years (Weeks & Gil de Zúñiga, 2021), there are still many reasons why more research into this topic is needed. The next section discusses four of them.

First, it is argued that “[t]he effects of new communication technology have not received its due respect from scholars of political violence” (Zeitsoff, 2017, p. 1972). In this Digital Age, the research community needs to look toward the future and pay more attention to the changing media- and communication landscape, also when coming from a perspective of conflict studies. Maura Conway – a professor of International Security at Dublin University – rightly observes that there has been a surge of research on the intersections of (social) media and political conflict recently, but frankly “without reference to the key scholars and existing studies of media and communication research generally and Internet studies particularly” (2017, p. 88). Therefore, this research attempts to offer a multi-disciplinary perspective on this topic – using knowledge of political geography and conflict studies, but combining this with insights coming from the fields of computer science, communication science and political psychology. Notable journals in these fields are *Policy and Internet*, *Information Communication and Society* and *Political Communication*. Its relevant insights will be discussed in the next chapter.

Second, the vast majority of research on this topic has merely focused on the political context of the United States (Jerit & Zhao, 2020; Tandoc et al., 2018). The political, social and legal context in the US is significantly different than the Dutch or European context, so one has to be careful with extrapolating American results across the Atlantic. This is why more research dedicated to specifically the Netherlands is needed. Although recently there has been some valuable research done in the Dutch context – e.g. by Michael Hameleers at the University of Amsterdam – this is mainly conducted from the perspective of communication psychology. Most of this research tends to stop after conclusions such as *disinformation is damaging to society, because people are very susceptible to it on social media*. Yet this research project aims to go a step further and also review the solution-thinking on the topic of disinformation. It reflects upon proposals by relevant organizations in the Dutch digital landscape that attempt to reduce the impact of disinformation on society.

Third, the global digital ecosystem is developing in such a rapid pace that some scholarly work from a couple of years ago can already seem outdated. The attitudes of scholars, governments, the tech sector and the general public are constantly changing when it comes to this subject. These dynamic attitudes can be spotted in scientific discourse, but also in recent news events, such as Twitter’s decision to suspend Donald Trump from their platform (Conger & Isaac, 2021) or the measures that the Dutch government took to counter disinformation in the 2021 elections (BZK, 2021). Five years ago, these actions may have been unimaginable. Therefore, conducting research on this topic in 2021 will most likely result in radically different conclusions compared to, for example, 2011 or 2016.

Fourth, various reports of and opinion pieces state that a strong and healthy relationship between theory and practice – with experts and scholars on the one hand, policy makers and tech employees on the other – is crucial in finding an effective solution to the problem of disinformation on social media (European Commission, 2018). This research project therefore aims to connect the attitudes of different actors together that are directly concerned with the problem and its solutions on a policy level.

## 1.4 Synthesis

This research aims to fill the four research gaps discussed in the previous section by investigating the current policy landscape in the Netherlands on the topic of disinformation. It analyses how the three relevant actors perceive the problem and its solutions. This is done by critically reflecting on how these actors actually diagnose the problem, how they think about possible solutions, and how they motivate and legitimize these perspectives. By focussing on the actors who have a direct impact on disinformation policies and debates, this research stays close to practice.

The next chapter reviews important literature – written by scholars from different academic backgrounds and expertise areas – about disinformation on social media. It also introduces the main research question and its sub-questions, which flow from “framing theory” as introduced in the chapter. Chapter three describes the methodology of this research. Chapter four, five and six are the ‘results’ chapters – it includes the analysis of the three different actors respectively. Chapter seven discusses and interprets these results. Finally, this thesis reports ends with concluding remarks in chapter eight.

## 2. Literature review

This chapter introduces important strains of research in the domain of disinformation. The first two sections elaborate on *what* disinformation is, *why* people are falling for it and *who* is most susceptible to it. This type of research emerged in the 1990's and gained more and more interest in the 2000's. It was mostly conducted from the perspectives of psychology, communication science and (behavioural) political science later on. This body of literature has grown cumulatively, building upon previous research and reaching a broad consensus among scholars. Then, since half a decade or so, more research is conducted to answer the question of how to effectively counter the spread disinformation, which will be discussed in the third section. According to the literature, this question seems certainly more challenging to answer than the former. There is a lot of disagreement among scholars and policy makers on the question of content moderation and the prevention of the dissemination of disinformation (Jerit & Zhao, 2020). Additionally, these counter-measures raises ethical questions which are also discussed in this section. The fourth section discusses framing theory by Benford & Snow (2000), which forms the foundation of the operationalization of this research. Lastly, the fifth sections discusses the main research question and its subquestions.

### 2.1 Definitions: “misinformation” versus “disinformation”

Before discussing the current literature, it is important to clear up the distinction between some key terms that are relevant to this research. In popular culture and media, the terms “misinformation”, “disinformation” and “fake news” seem to be used interchangeably. This research will attempt to keep a distance from the loaded term “fake news”, because it has become a buzzword during the last few years (Tandoc et al., 2018, p. 2). Although it does refer to false information on social media in many cases, the term has also repeatedly been used to discredit critical reporting from traditional news sources, thereby further “muddying” our shared understanding of the term (ibid). Hence the term is avoided in this research. It will only refer to the terms misinformation and disinformation, following most papers and policy reports (European Commission, 2018). Misinformation can be any type of shared information that is factually false, while disinformation refers to false information that is deliberately created and shared when *known to be false* (Wardle, 2018). Both are false information, but the key difference ultimately lies in the intention of the actor who spreads the information.

Yet from the receiver's point of view, misinformation and disinformation can be understood as the same thing: false information. Now, most of the psychological research focuses on the effects of false information on the receiver, regardless of the intention of the actor who spreads it. Hence one can be *misinformed* by *disinformation*, and be ignorant of this fact. This relation between the two terms is why – when researching disinformation – misinformation is still a relevant concept. In an influential paper, Kuklinski et al. (2000) introduced the concept of misinformation in the political setting, and defined it as follows: “[p]eople are misinformed when they confidently hold wrong beliefs – [it] acts as an obstacle to educating the public with correct facts” (p. 790). Due to this psychological phenomenon, the impact of disinformation on political discourse can be significant. Research on the effects of false information from the receiver's point of view has mainly been conducted from the perspective of political psychology. This large body of scholarly work led to some overarching conclusions, as explained in the next section.

### 2.2 Psychology: the cognitive stickiness of misinformation

To get straight to the point: it is not easy to convince people that they are wrong about something. We generally believe to be right, and we are very good in reasoning why this would be true (Kunda, 1990; Lewandowsky & Oberauer, 2016; Taber & Lodge, 2006). This is known as “motivated reasoning”. Time and time again, this phenomenon is empirically confirmed by psychological research in various contexts (ibid). This great trust in one's own correctness is present when people hold factually correct beliefs – such as the conviction that the earth is round – but also when they are misinformed – such as the belief that the

earth is a flat disc. There are cognitive variables within each person that render misinformation “sticky” in the mind (Lewandowsky et al., 2012) and leads to motivated reasoning (Bode & Vraga, 2015). These cognitive processes make it particularly hard to correct for misinformed beliefs. Furthermore, even after misinformation is successfully corrected, some parts of the brain can still cling onto the previously-held beliefs (Swire-Thompson et al., 2020). This is what political scientist Emily Thorson calls “believe echoes” (2016). All of these cognitive processes also translate to online environments, where misinformation can be ubiquitous. Therefore, as this psychological research shows, diminishing the impact mis- and disinformation on social media will always be an inherently challenging process due to the nature of the human mind.

Although practically every human is susceptible to false information to some degree, there is a gradient in the stickiness misinformed beliefs. For example, research has been done on the political asymmetry in the cognitive stickiness of misinformation. Ecker & Ang (2019) found in an experimental study that – when presented with factual information that disproved previous misinformed assumptions – self-ascribed “conservative” participants were significantly less willing to change their worldview than self-ascribed “progressive” participants would. Other studies have repeatedly reached similar conclusions (e.g. Jost, 2017; Lewandowsky & Oberauer, 2016). Simply put, social media users cannot be seen as one homogenous group – some groups may be more vulnerable to the impact of false information than others.

## **2.3 Psychology meets policy: the effectiveness of counter-measures**

Because misinformation can be sticky in the mind of the receiver, and there has been a surge in political mis- and disinformation on social media recently, it has become more and more important to investigate what would be the most effective way to counter it. Back in 2010, Nyhan & Reifler remarked that “only a handful of studies in political science have analysed the effect of attempts to correct factual ignorance or misperceptions” (p. 305). Since then, the interest in this particular sub-topic – both academic and public – has been growing exponentially (Weeks & Gil de Zúñiga, 2021). In the next section, two widely used categories of counter-measures on the level of content will be discussed.

### **2.3.1 Algorithms & content moderation**

There has been a lot of attention to the way in which algorithms determine what people do and do not see on their social media feeds. Without venturing into the technical details of these algorithms, in essence these algorithms are argued to be the foundation of heavily personalised news feeds and the controversial phenomenon known as “filter bubbles” or “echo chambers” (Allcott & Gentzkow, 2017). Some prominent scholars such as Shoshana Zuboff (2019) argue that this algorithmic content curation is essential to how social media platforms generate online engagement – and thereby monetize their service – but that it simultaneously enhances the very existence of filter bubbles and emotion-triggering misinformation on the platform. Consequently, scholars noticed quite early that tweaking this algorithm – in order to diversify the information space of social media users, e.g. by introducing more serendipity – may contribute to the diminishment of extreme partisanship and polarization on social media (Bode & Vraga, 2015; Reviglio, 2019). After all, a diverse information space is effective to reduce users’ need for motivated reasoning in the first place, and this pluralism is therefore considered to be “essential for the functioning of a democratic society” (Möller et al., 2018, p. 959). Indeed, in the following years, social media companies started to experiment with tweaking their algorithms as a counter-measure, in order to burst bubbles and minimize the impact of misinformation (Paul, 2020; Roose et al., 2021). However, the actual effectiveness of these types of private interventions have not yet been empirically evaluated by independent researchers, because the inner workings of algorithms are generally not yet shared by social media platforms. Besides, some studies suggest that algorithms in themselves do *not* explain the existence of filter bubbles per se (Haim et al., 2018; Möller et al., 2018). In conclusion, the scholarly debate remains ambivalent on the topic of algorithmic intervention.

### 2.3.2 Fact-checking & labelling

Given the existence of misinformation on personalised social media feeds, there have been efforts by journalists and social media platforms to fact-check disputed stories and subsequently label those posts as false<sup>2</sup>. Fact-checking and subsequent labelling has become one of the most popular strategies address the prevalence of misinformation – its usage by traditional media outlets and social media platforms has been growing significantly in the previous two decades (Walter et al., 2020). Researchers in turn have been conducting experiments and meta-analyses to investigate which kind of these interventions may be the most effective. Yet the empirical evidence on this issue points towards two different directions, which will be discussed below.

On the one hand, researchers find that fact-checking and labelling of false information is not an effective or desirable counter-measure. Walter & Murphy (2018) conducted a comprehensive meta-analysis in which they studied the effectiveness of correcting misinformation across various contexts. By comparing the results of a large number of empirically verified studies of the past decade, they concluded that (1) political misinformation is the hardest to combat, followed by misinformation in the contexts of (2) marketing and (3) health (ibid). Furthermore, they compared the effectiveness of various corrective attempts. It seems that simple fact-checking and/or general warnings are not sufficient in most situations – “a successful correction would also include a coherent explanation for how and why the false rumour started” (ibid, p. 436). Another meta-analysis on “debunking” misinformation – conducted by Chan et al. (2017) – reached a similar conclusion: correcting misinformation with comprehensive arguments and providing new details prove to be far more effective than simply labelling the misinformation as wrong. To go even further, some research suggests that merely fact-checking and labelling misinformation as wrong could even cause a “backfire effect” (Nyhan & Reifler, 2010). If ideologically vulnerable groups see their posts labelled as false by platforms or fact-checkers, they may dig themselves even further into the trenches of their believe system. Furthermore, some fact-checking organizations and social media moderators are increasingly accused of being partisan or biased, thereby undermining their independence and legitimacy (J. Kelly, 2021).

Yet on the other hand, some research suggests that mere fact-checks and/or warning labels do in fact work in certain circumstances. For example, Paul Mena (2020) studied the effectiveness of simple warning labels that Facebook places under news stories that may contain false information. He found that this process of flagging reduces users’ intentions of sharing those posts with their network. These types of positive conclusions are confirmed by scholars such as Clayton et al. (2020), who found that “a negative effect of general warnings on belief in misinformation” (p. 1090), and Fridkin et al. (2015), who concluded that “fact-checks influence people’s assessments of the accuracy, usefulness, and tone of negative political ads” (p. 127).

### 2.3.4 Synthesis

All in all, research concerning the effectiveness of counter-measures on social media remains divided and has not yielded overarching conclusions. Hence “the process of correcting misinformation is complex and remains incompletely understood” (Chan et al., 2017, p. 1531). Some researchers seem favour certain measures, some other researchers can deny its effectiveness. In other cases, the effectiveness of counter-measures cannot be evaluated properly at all. Hence there seems to be no one-size-fits-all solution to the problem of disinformation on social media. Furthermore – and this is crucial – although there is this large body of excellent research, pointing towards what might be effective measures, there is one elephant in the room: in practice, social media platforms essentially pull the strings themselves when it comes to countering disinformation on their platforms (Gillespie, 2018). As private companies, they are free to intervene and

---

<sup>2</sup> Probably the most famous example: Donald Trump’s posts about the “stolen election” of 2020 were labelled by Twitter and Facebook as “containing misleading information”. As an example in the Dutch context, Twitter added similar warnings next to Thierry Baudet’s speculating tweets about the danger of COVID-19 vaccines.

experiment with different counter-measures. Yet they are also free to sit back and do nothing. Some of the (non-) actions by these companies have already caused public outcry and controversy. At this point, government regulations and/or public pressure can make a difference and push social media platforms in a certain direction.

According to Philip Howard (2020), professor at the Oxford Internet Institute, contemporary disinformation is “both a social and a technical phenomenon” (p. 21). It exists by virtue of the actors who spread the disinformation intentionally – the social aspect – and by the existence of social media platforms who facilitate and distribute it – the technical aspect. Thus it is however an important question to ask: who ought to have the prime responsibility to tackle this problem? As scholars have noticed, this question raises ethical concerns (Bjola, 2018). Yes, contemporary disinformation spreads rapidly on social media digital infrastructure, but do these private companies have the inherent responsibility to act on it? If they do, can they go too far in their actions? Or should the government be more proactive in establishing regulations and/or laws? Should these laws focus on the social or the technical aspect? Can specialized NGO’s make a meaningful difference? And what role does the individual citizen play in this story? Corneliu Bjola (2018, p. 310), a scholar who specializes in strategic communication and countering disinformation, writes about these dilemmas:

“having a governmental institution in charge of combating disinformation can be beneficial in terms of access to human and financial resources, technical expertise, and coordination mechanisms, but sometimes less so in terms of public credibility, as governmental actions in the area of strategic communication are generally viewed with suspicion, both by domestic and (especially) international audiences. Non-governmental organizations may suffer less from this problem so long as their political independence is well established, but this advantage may come at the expense of reduced effectiveness, as resources for an NGO can be scarce.”

As Bjola notes, governmental institutions that intervene in the information space of liberal democracies can indeed look suspicious, because freedom of expression is one of the keystones of these societies (Somaini & Pollicino, 2020). However, as mentioned before, there is not always a clear distinction between real or fake news, between authentic or inauthentic online speech. Hence even if anti-disinformation efforts by governments – or by powerful social media companies – are entirely coordinated in good faith, there is still a concern about its impact on online freedom of expression. Therefore, according to Mortera-Martinez (2019, p. 6), “the line between suppressing freedom of speech and shutting down disinformation campaigns is thin”.

Given the uncertainties about which counter-measures against disinformation could be fair and effective, and which actors should be responsible for initiating these measures, it is a relevant question to ask: what kind of answers are given by those who are working on this problem in practice? In order to investigate how the involved actors perceive the problem of disinformation and its solutions, this research uses framing theory as operationalization, which will be discussed in the next section.

## 2.4 Framing theory

In order to analyse how different actors perceive the problem and its possible solutions, this research uses framing theory as an analytical lens. The landmark paper of Benford & Snow (2000) on Collective Action Frames still holds much legitimacy in this regard. In this heavily-cited work, Robert Benford and David Snow review a body of earlier research devoted to framing. They adopt a definition by the influential sociologist Erving Goffman: “frames denoted ‘schemata of interpretation’ that enable individuals to locate, perceive, identify, and label occurrences within their life space and the world at large” (Goffman, 1974 in Benford & Snow, 2000, p. 614). Benford & Snow (2000) identify three core framing tasks that appeared in the literature of the preceding thirty years: “scholars have devoted considerable empirical attention to identifying and analysing the various types of *diagnostic*, *prognostic* and *action mobilization* framings” (p. 615).

In short, *diagnostic frames* are used to identify a problem, *prognostic frames* are used to describe a solution to this problem, and *motivational frames* are used to justify a certain approach towards this problem (Benford & Snow, 2000). Although governmental institutions or private companies are not perceived as Collective Action groups in the traditional sense, this analytical tool is arguably still very useful to identify important distinctions in how the three respective actors approach the problem of disinformation, and what is the difference between their approaches. All in all, framing theory provides important handles that will be used throughout this research to operationalize the data collection, interpret the collected data and draw deliberate conclusions.

## 2.5 Research questions

The previous sections outlined the problem of disinformation on social media, its consequences to contemporary democracies and the need for a way forward to tackle this problem. The digital ecosystem is vulnerable to the spread of disinformation, people on social media are susceptible to it, and the attitudes towards potential solutions to this problem vary substantially between different experts, scholars, policy makers, civilians, and social media platforms. Therefore, to investigate the framing of this problem by relevant actors in the disinformation debate, this study uses the following research question:

**How do the three relevant actors in the Dutch digital landscape – governmental institutions, social media platforms, and Internet freedom organizations – frame the problem of disinformation on social media and its potential solutions from 2017 to 2021?**

By taking this research question as a guideline, this study aims to provide a comprehensive overview on where these actors stand on possible solutions to the problem of disinformation on social media – e.g. on legal regulation, on content moderation, on a way forward in general. Therefore, the operationalization of this research question can be divided into three parts, which is based on the framework by Benford & Snow (2000) as explained in the previous section. First, the diagnostic frames are investigated, i.e. how these different actors describe and approach the problem of disinformation. Second, the prognostic and motivational frames are investigated, i.e. what their attitude is towards potential solutions and how they legitimize those solutions. Third, this process sets up a vantage point from where the answer to the last sub-question can be identified. It allows for an comparison of the different viewpoints, in order to identify a general consensus between all actors – which may render policy implications with a broad support base – and identify aspects upon which there is still much friction and disagreement.

This three-way process can also be formulated into the following sub-questions below. The first two questions guide the analysis, and the written results – chapters 4, 5 and 6 – are structured accordingly by using sub-headings. The third question guides the discussion in chapter 7.

### 1. How do these three relevant actors perceive the problem of disinformation on social media?

- Diagnostic framing:
  - i. What problems are identified?
  - ii. What are the important aspects of this problem?
  - iii. How is blame or causality attributed?

### 2. What is their approach towards potential solutions?

- Prognostic framing:
  - i. What potential solutions to the problem are identified?
  - ii. What concrete strategies are proposed to carry out these potential solutions?

- Motivational framing:
  - i. What justification(s) for these potential solutions are offered?

**3. What are similarities and differences between to the used framings?**

It should be clear by now that this research takes an empirical rather than theoretical approach. By keeping close to the attitudes of policy makers in the field and using the concept of framing (Benford & Snow, 2000) as analytical tool, this research contributes towards a practical understanding of what could be effective solutions to this problem, which enjoy a broad support base among different actors in society.

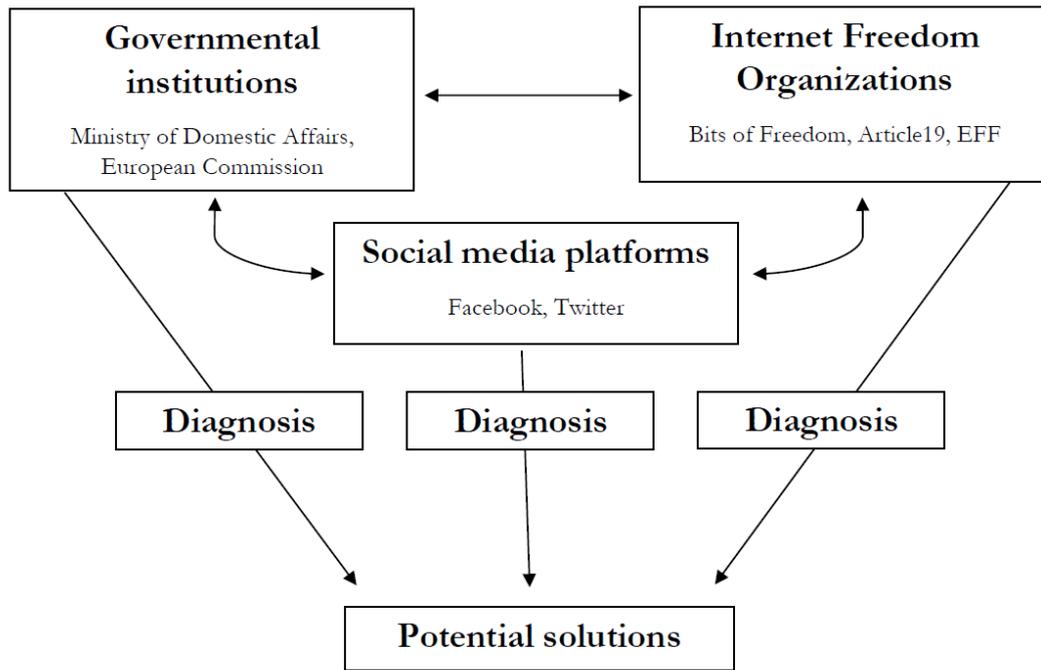


Figure 1 – Abstract representation of the policy landscape on disinformation in the Dutch digital landscape

### 3. Methodology

This study incorporates comparative qualitative research with an inductive approach. Methodologically, this implies a (1) thorough document analysis as primary data, combined with (2) semi-structured expert interviews that allow for more depth. This has two main advantages. First, this method offers great insight into the subjective attitudes of the different actors and how they compare to each other. Second, this type of research could be done from home, analysing documents that are publicly available on the Internet, and conducting interviews through video calls. In other words: COVID-19 restrictions – which had enormous complications on society at the time of writing – did not form a limitation here. The next section discusses the three different actors that function as units of analysis. Hereafter, the method of data collection and analysis are discussed in more detail, and some ethical principles that guided this research process are discussed. The last section elaborates on some important limitations of this research project.

#### 3.1 Data sampling

As this research takes a qualitative approach, random sampling and statistical inferences do not apply here. Instead, the sample is “intentionally selected according to the needs of the study – commonly referred to as ‘purposive sampling’ – [because the particular sample] can teach us a lot about the issues that are of importance to the research” (Boeije, 2010, p. 35). Now, from the literature review and research questions flow three units of analysis: the key actors in the debate around disinformation on social media. These are (1) governmental institutions; (2) social media platforms; and what this research calls (3) Internet Freedom Organizations (IFO’s). The next section discusses why these three actors are selected and which documents published by these actors are collected as primary data sample. A complete list of all the documents and interviews included in the data sample can be found in appendix I.

##### (1) Governmental institutions

When a problem in society needs to be solved, it is not uncommon to turn towards governmental institutions. Those institutions then can create policies and legal regulations that attempt to solve it. This has also been the case when it comes to the problem of disinformation. As mentioned before, the Dutch government has been formally concerned with this problem since early-2017. Since then, several parliamentary documents and questions (“kamerstukken” and “kamerbrieven”) have been published, which propose certain policies in order to mitigate the damage of disinformation on social media. The Ministry of Domestic Affairs (BZK) is particularly involved in this dossier. Furthermore, Dutch security services, the AIVD and NCTV have reported on this issue as well. Therefore, the primary data for this research includes all the relevant parliamentary documents – which are all publicly available on [rijksoverheid.nl](http://rijksoverheid.nl) – in combination with the relevant sections of AIVD and NCTV reports (n=14). Furthermore, an expert interview was conducted with an expert at BZK who works on policy on disinformation.

Additionally, on the European level, the European Commission (EC) has released various reports about the issue of disinformation on social media. In these documents, the Commission proposes policy recommendations and guidelines for the European Union and its individual member states. It is important for the analysis to take this international level into account, because the problem is almost inherently borderless. Brussels has a large impact on how one thinks about solutions to this problem on a national scale. Hence the primary sample of documents consists of relevant documents on the topic of disinformation published by the EC (n=5). These documents give an comprehensive insight, because the reports are rather extensive and detailed. Additionally, an expert interview was conducted with Bart Groothuis, member of the European Parliament for the liberal Renew Europe party, and member of the Committee on Foreign Affairs (AFET) and the Special Committee on Foreign Interference and Disinformation (INGE).

## (2) Social media platforms

One cannot talk about solutions to the problem of disinformation on social media without involving the social media platforms themselves. Following Boeije (2010), methodological choices should be guided by a principle of “maximization” of the setting. This means that “a location should be determined where the topic of study manifests itself most strongly” (p. 34). In this case, social media platforms simply cannot be ignored, because the problem manifests itself in exactly those digital spaces. Social media platforms come in different shapes and sizes, and therefore “their ability to take measures against disinformation differs from platform to platform” (Instituut voor Informatierecht, 2019, p. 24). Hence the question is which social media platforms to include in the unit of analysis. However, it is well-known that the three big Silicon Valley enterprises – Facebook, Google and Twitter – are the most important in shaping online discourse (Gillespie, 2018). However, due to limited time and resources, this research merely includes an analysis of Facebook – the biggest social media company of them all – and Twitter – a smaller, but still highly influential social media platform, due to its focus on global news events and political discussions. These two influential platforms are also chosen because of their prominent role in disinformation campaigns during the last five years, as mentioned in the introduction.

The primary data sample of this actor consists of the publications of Facebook and Twitter on the topic of disinformation. Both companies run extensive ‘about’ websites that function as a public outlet.<sup>3</sup> On these web pages, the companies elaborate extensively on the things that are happening on their social media platforms, such as internal policy decisions and product changes. To put it simply in Facebook’s terms, these web pages emphasize “what we do” and “why we do it”. Hence the data sample for this research consists of all the relevant publications on the topic of disinformation by both Facebook (n=15) and Twitter (n=16). Additionally, an interview was conducted with a public policy expert who works at Facebook in the Benelux region. Unfortunately it was not possible to conduct an interview with someone who works at Twitter. I did have a back-and-forth e-mail correspondence with a senior employee of Twitter’s Public Policy department in Brussel, who at some point agreed to an interview. Yet, strangely enough, the employee has not responded to any of my e-mails since then. Hence the Twitter data sample merely consists of the primary data: the relevant documents from its website.

## (3) Internet Freedom Organizations (IFO’s)

This third actor consists of various NGO’s which all voice similar concerns about issues in the digital realm. To include this group into the analysis is to include a critical opinion that exists in civil society, because – as these organizations are keen to point out – some potential solutions to the problem of online disinformation could pose genuine risks to online freedom of expression. By warning everyone about the treacherous path of Internet regulation, they function as a watchdog for free communication and privacy rights in the digital ecosystem. If government institutions propose legislation that concerns disinformation, or social media platforms call out for more content moderation, these organizations raise concerns, respond with critical articles and try to influence policy making by lobbying. In most cases, their employees have a background in ICT or information law, so they bring relevant expertise to the table. For categorization purposes this research labels these organizations as “Internet Freedom Organizations” (IFO’s) – to indicate this specific type of NGO – because their common denominator is that they all emphasise the need for a free and open flow of information on the Internet. The following IFO’s are included in this research:

---

<sup>3</sup> These can be found at <https://about.facebook.com> and <https://blog.twitter.com>

- **Bits of Freedom** is a movement that stands for Internet freedom in the Netherlands, and it is probably the most well-known IFO in the country. They focus on two essential rights – freedom of expression and privacy – and make sure that these principles guide better policies in governments, institutions and companies (Bits of Freedom, n.d.). This organization publishes insightful articles and opinion pieces, has expertise in technology and enjoys an engaging support base. Furthermore, they have direct contact with BZK and social media companies. In the last few years, they have published a handful of documents on the topic of disinformation. These documents are used in this research (n=5). Additionally, an expert interview was conducted with Rejo Zenger. He works as senior policy advisor at Bits of Freedom, which means that he focuses on influencing policies of governments and companies by speaking to all involved stakeholders. Besides his lobbying work, Rejo Zenger co-authored the publications that are included in the document analysis.
- **Article19** is committed to two fundamental freedoms: freedom to speak and freedom to know (Article19, 2020). This organization publishes critical pieces on the grip that governmental institutions and Big Tech have on the free flow of information in the digital sphere. Article19 does not merely focus on the Western world; a significant part of their work emphasises the need for Internet freedom in the Global South as well. During the last five years, Article19 has repeatedly voiced their concerns about legislation around disinformation that could put online freedom of expression in jeopardy. These concerns are inscribed in various publications on the topic of disinformation, which are included in this research (n=10).
- **The Electronic Frontier Foundation (EFF)** “is the leading non-profit organization defending civil liberties in the digital world. Founded in 1990, (...) EFF's mission is to ensure that technology supports freedom, justice, and innovation for all people of the world” (EFF, 2021). This organization was established 30 years ago in San Francisco, not far away from the cradle of the World Wide Web. Since then, the organization has grown into the one of the most influential NGO’s on the topics of digital rights and internet activism. They use expertise from technologists, activists and attorneys from all over the world. The EFF published various documents on the topic of disinformation since 2017, which are included in this research (n=10).

This adds to a total of n=25 primary data points for this actor. Bits of Freedom agreed to do an interview – which led to a very insightful conversation – but unfortunately both Article19 and EFF declined the interview requests by.

### 3.2 Data collection

Relevant documents and reports of the three actors were gathered as primary data. These documents allow for an insight in the position that the three different actors take on the problem and its solutions. This method is particularly suitable to qualitative studies, because it provides rich descriptions of a particular phenomenon, organization or program (Bowen, 2009). The documents that are included in the sample all meet two simple requirements: (1) all documents involve the problem of disinformation or a response to it, almost exclusively in the context of Netherlands or Europe, and (2) all documents are published between March 2017 and May 2021, which corresponds with the timeframe of this study<sup>4</sup>. As data saturation was reached, any further questions or clarifications would shape the foundation of the secondary data collection method.

---

<sup>4</sup> This timeframe is chosen because the issue of disinformation on social media has been prominent on the public agenda since 2017. When searching for the term ‘disinformation’ on various relevant websites – news outlets, governmental institutions, social media platforms, NGO’s – search results from before 2017 are rare.

This secondary data consists of interviews with experts from the three respective actors. The interviews were semi-structured, because the research question, literature review and document analysis provided talking points for an interview guide, which was created beforehand. However, if the respondent brought up interesting subjects, there was the option to spend a more time on those points. The guiding questions in the interview guides were changed slightly per interview to properly fit each respondent. A total of four interviews were conducted, which were 30 to 45 minutes in length. Three of them were done through video conference software, one was a regular phone call.

Data collection took place from February 2021 to June 2021, a period of approximately five months. During this time there was a back-and-forth interaction between data, reminiscent of the influential Grounded Theory approach by Glaser & Strauss (1967). Data collection and data analysis took place around the same time. The next section will discuss the analysis part in further detail.

### **3.3 Data analysis**

After the relevant documents were collected and thoroughly read, the next step involved a comprehensive analysis of that data. As Ragin & Amoroso (2019) explain, the analysis is a crucial step in a research process because it includes “breaking phenomena into their constituent parts and viewing them in relation to the whole they form” (p. 51). In order to do this meticulously, the documents were imported to Atlas.ti, a data analysis program. Further analysis took place in this program, in a methodological process called “axial coding” (Boeije, 2010, p. 114). Some suitable codes were already devised from the research question and literature review, particularly to operationalize framing theory. Subsequently these codes could be applied to the relevant fragments of the documents and interview transcriptions. However, if there was information in a document for which no suitable code had yet been created, new codes were added – this connects to the inductive approach of the study. Finally, after all documents were read and coded, general framings and narratives became evident. These findings became the foundation on which chapter 4, 5 and 6 of this thesis report are written.

### **3.4 Ethical considerations**

Although the experts that were interviewed for this research do not fall into particularly vulnerable groups, it is nonetheless important to respect the respondent’s rights as a researcher. Therefore, this research is guided by two important ethical principles as described by Boeije (2010): (1) informed consent and (2) confidentiality and anonymity.

First, informed consent “is the obligation to outline fully the nature of the data collection and the purpose for which the data will be used to the people or community being studied” (Boeije, 2010, p. 45). Thus, before the interview as well as within e-mail exchanges, the respondents were informed about the nature of the study and how the respondent’s perspective would fit into the research. Additionally they were asked if they allowed for the recording of the interview. All of them approved, so the transcriptions were made with their consent. Second, confidentiality and anonymity refers to the “agreement as to how the data are to be handled in the research in order to ensure privacy” (Boeije, 2010, p. 46). Hence this builds upon the “informed consent” foundation. All respondents were asked if their names could be mentioned in the thesis report, or if they would like to stay anonymous. In response, two of the respondents stated that they would rather stay anonymous.

### **3.5 Methodological limitations**

“Researchers need to acknowledge the limitations of their study design and instrument” (Price & Murnan, 2004, p. 67). This section discusses the inevitable limitations that this research process encountered, which can have implications for the research findings. The main limitations concern (1) the scope of this study and (2) the internal validity due to the interpretative nature of this study.

The first important limitation concerns the scope of this study. Due to limited time – this research was conducted in a timespan of five months – difficult methodological choices had to be made. Because the disinformation is such a broad societal issue at the moment, many stakeholders are involved in some way. Including everything and everyone would not be feasible with the available resources. Therefore, this research includes the perspectives of Facebook and Twitter, but it leaves out Google, owner of YouTube, a major social media platform that has also stepped up their efforts against disinformation during the last couple of years (Alexander, 2020). Additionally, other smaller platforms with nonetheless a sizeable influence in particular target audiences were left out, such as TikTok and Reddit (M. Kelly, 2021). Similarly, this research chooses to include the perspective of relevant NGO's: Article19, EFF and Bits of Freedom. Yet other organizations such as Freedom House, Internet Freedom Foundation, The Centre for Law & Democracy and the Pirate Party are left out. Although these organizations could provide their unique perspectives, their impact on policy around disinformation (in the Netherlands and Europe) is relatively slim in comparison to the three IFO's that are included in this research.

While it can be argued that this limited scope of this study could be a shortfall of this study – as certain actors are left out, notably Google – the opposite can also be true: the scope of the study may be too broad. This research incorporates many different viewpoints from highly variant actors, which means that it cannot delve incredibly deep into any one of them. For example by incorporating the EC in this research, the “European perspective” on disinformation policy in the Netherlands is included, but this perspective may be oversimplified because the influence of Brussels is not merely the Commission. The chosen approach aims to give a birds-eye view of the policy landscape on disinformation in the Netherlands, so some nuances may be lost in this bigger picture. Thus further research may be needed which properly investigates one of these actors in greater detail – e.g. by conducting a comprehensive discourse analysis with more data points per actor – in order to analyze its contribution in policy around disinformation on social media.

Another limitation, though more subtle, has to do with the interpretative nature of this research, which affects its internal validity. This refers back to the work of Anthony Giddens from the early 1980's. His “double hermeneutic” theory explains that, in the social sciences, researchers and objects of study have a two-way relationship. This contrasts with the “single hermeneutic” in natural sciences. The current research project falls in the former category. Thus, as a researcher, I “acquire knowledge by making an (academic) interpretation of how actors understand their social world” (Demmers, 2016, p. 16). These interpretations will inevitably be distorted by personal biases. I am aware that my interest in the topic of disinformation was initially incited by influential public figures such as Tristan Harris, Nina Schick and Shoshana Zuboff, who are all fierce critics of social media platforms and Big Tech in general. However, this research attempts to avoid conformation biases by taking in all perspectives from the different actors in a similar way: open yet critical, objective yet self-aware.

## 4. Governmental institutions

### 4.1 National level: Ministry of Domestic Affairs (BZK)

The Dutch government's official involvement in the debate around disinformation on social media started somewhere in the autumn of 2017. Earlier that year, stories about Russian interference in the US presidential elections started to pick up a lot of attention in the international press. Evidence suggested that Russia could have played a significant role in the outcome of the election, by hacking the Democratic National Committee and combining it with “a disinformation campaign waged predominantly on Facebook and Twitter, in an effort to use automated bots to spread fake news and pro-Trump agitprop” (Sheth & Bertrand, 2017). Hence Dutch politicians were concerned that this might occur in the Netherlands as well. Some parliamentary questions – directed at BZK – were asked in November 2017 by members of D66. Among other things, the minister was asked if she was aware of the Russian disinformation campaigns, and if BZK was taking any precautions for the forthcoming local elections. In her response in December 2017, the minister made clear how the government understands the problem of disinformation and what possible solutions would look like. Over the subsequent four years, more parliamentary documents would follow, which describe the approach that the Dutch government takes with regards to disinformation on social media. The next section discusses the analysis of these documents (n=14) – combined with relevant insights from an interview with an anonymous BZK employee who works on disinformation policy – by using framing theory as an analytical tool (Benford & Snow, 2000).

#### *Diagnostic framing*

##### Perception of the problem

The analysis reveals how the Dutch government perceives the problem of disinformation on social media, and how this slightly changed over time. In particular during the early phase – in 2017 and 2018 – there was much emphasis on the nature of these novel threats that democracy could face. The problem is officially diagnosed by BZK as “covert foreign interference through disinformation” (“heimelijke buitenlandse beïnvloeding via desinformatie”). References are made to relevant reports from both the General Intelligence and Security Service (AIVD) and the National Coordinator for Security and Counterterrorism (NCTV). These governmental security organizations – operating under the ministries of BZK and J&V respectively – have also mentioned the problem multiple times by then. These organizations tend to frame the problem security terms, such as “hybrid threats” or “information warfare”. Hereby they rightly note an important aspect of this problem: it forms a substantial threat to national security, although it merely manifests itself in digital space. In particular, these organizations note that disinformation flourishes on social media platforms. This digital dimension is what makes the problem novel and dynamic, and it thus requires more forward-looking solutions. Furthermore, there is a recognition that this is both a *permanent* and *multi-stakeholder* problem. Hence it does not merely require considerable attention during elections periods, but at all times, from all involved stakeholders:

“It’s not a static topic. It’s always developing, so you always need to think of new solutions to tackle it.”<sup>5</sup>

##### Important consequences

According to the government, the central potential harms of this problem are (a) unwanted influence on public opinion, (b) interference in election results, (c) loss of trust in governmental institutions and practices and (d) increase of social polarization in society. But, above all, the problem is harmful to democratic society as a whole. This broad point is frequently emphasized throughout all documents.

---

<sup>5</sup> Interview BZK, 21 April 2021, Utrecht.

## Blame and causality

First, the majority of the blame for this problem is attributed to foreign state actors, especially in 2017 and 2018, inspired by the warnings of the AIVD and NCTV. Russia in particular is mentioned frequently, as the country was known by then to be meddling in various other democracies through disinformation campaigns. Thus, for the government, “foreign state actors” are initially seen as an important cause of this problem, thereby emphasizing that this is essentially a threat from the outside coming in. However, in more recent documents, this blame is being attributed to not only foreign states, but “malicious actors” in general. This conceptual expansion leaves room for foreign individuals or even domestic actors to also be part of the problem:

“I think that in the beginning – when we started with this – we approached the problem mainly from a state-perspective: the foreign state threats. But surely the COVID crisis has shown that this is also a problem within the internal information domain of the Netherlands.”<sup>6</sup>

Second, the aspect of “new technologies” has a role in the causality chain too, although it is notably less emphasized in comparison to the first aspect. The government recognizes that (sophisticated) technologies in contemporary society – notably the existence of social media platforms, advanced algorithms, artificial intelligence, automated bots – increases the likelihood that democracy could be affected by the problem of disinformation. Due to this technology, the problem manifests itself in anonymity and is relatively risk-averse. It operates at high speeds with low costs. Hence the government recognizes that the problem of disinformation is inherently tied to digital space, social media platforms, and advanced technologies.

## ***Prognostic framing***

### Proposed solutions

Since this problem first appeared on the agenda, the government has mainly been emphasizing the need for *enhancing awareness* in citizens and public officials. In other words, individuals need to realize that online information is not always trustworthy, and they should therefore make their decisions – e.g. who to vote for during elections, or what local legislation to impose – by reviewing various sources and rationalize responsibly. The emphasis on increasing the public’s knowledge and awareness – which aims to prevent that disinformation negatively influences people – has remained the focal point for government policy throughout the past five years. Following from this, the government also underlines the importance of conserving the plural, open and *high-quality media landscape* in the Netherlands, which diminishes the chances for much disinformation to take root in society.

In addition, there is much emphasis on the need for *cooperation* with other actors. This is rooted in the recognition that this is a permanent and multi-stakeholder problem. First, the European Union is mentioned most frequently in this context, especially after the publication of the EU’s Code of Practice on Disinformation and other relevant reports and proposals by the EC – which will be discussed later. Hereby the Dutch government recognizes that the solution may best be found at the European level, because “disinformation does not stop at the borders”. European solutions could be more effective due to a bundling of resources, knowledge and experience. Second, the government sees potential in cooperation with technology companies and – to a lesser extent – researchers and academics. This is mainly framed in terms of “conversations” with technology companies, which aim to find common ground between the different interests. The government also wishes to accommodate more independent research on this topic.

---

<sup>6</sup> Interview BZK, 21 April 2021, Utrecht.

Lastly, the government would like to see that technology companies – such as Facebook, Google and Twitter – recognize their responsibility on countering disinformation on their own online platforms. These companies should also *enhance transparency* in their decision-making around policy guidelines and content moderation on their platforms. However, the government is wary of proposing national legislation to force these companies into a particular direction, because this would have implications for freedom of speech and free flow of information – which will be further discussed in the analysis of the motivational framing.

### Concrete strategies

As of 2021, the government’s approach towards disinformation now consists of three main strategies: (1) prevention, (2) strengthening of information position, and (3) – if absolutely necessary – a reaction. First, the prevention strategy focusses on enhancing awareness around the topic of disinformation, which includes major government programs for media literacy trainings and awareness campaigns (“bewustwordingscampagnes”), funding for investigative journalism, funding for high-quality media initiatives, and new transparency laws for Dutch political parties with regards to online advertising. These actions aim to create a resilient information society (“weerbare informatiesamenleving”), which diminishes the chance that disinformation will effectively influence the minds of Dutch citizens.

Second, the government needs to know exactly how the problem manifests and involves – it wants to strengthen its information position. This is done through various programs created by the EC, involving cooperation with other EU member states. Additionally, there is cooperation and information exchange with various security services, civil society groups and journalistic institutions in the Netherlands. The information position is also enhanced by stimulating further independent research into the threats that new technologies can pose to society, because a greater understanding of the problem will generate more sophisticated solutions.

Third, the government repeatedly emphasizes that, generally, it will not be concerned with reacting to disinformation – understood as fact-checking or debunking specific online content. This is seen as primarily a task for the journalistic and scientific communities, in cooperation with social media platforms. However, one caveat remains: if there is concrete evidence that national security or public health is under *direct* threat, the government might be obliged to step in and correct the false information, while explicitly not removing it. For example, this could be seen in the government’s response to disinformation around COVID-19 and its vaccines:

“And we can see that disinformation – if there is danger for public health, national security, civil unrest – that it might be a rationale for us saying: okay, this message is published, it is clearly false, don’t fall for it, the truth is actually...”<sup>7</sup>

### ***Motivational framing***

The government’s main focus on enhancing awareness among the public is framed from the fundamental starting point of free speech. After all, in a liberal society such as the Dutch, it is “one’s own responsibility” to move through a treacherous digital terrain and recognize which information is true and which is false. It is not the government’s task to determine this – the last thing that BZK wants to become is an Orwellian Ministry of Truth. Therefore, the government repeatedly reiterates the utter importance of an open society with freedom of speech and a free flow of information. This debate between right and wrong is emphasized as very important. The government does not want to prohibit false information, but it might – in the light of direct harmful information, as a last resort – provide correction and context to it:

---

<sup>7</sup> Interview BZK, 21 April 2021, Utrecht.

“Disinformation as a phenomenon is not illegal or unlawful, it is undesirable. And we are not going to prohibit it. (...) Especially because everyone has the right to express their opinion – [removing certain information by law] would really oppose freedom of speech.”<sup>8</sup>

From this standpoint, imposing a regulatory framework that explicitly forces social media companies to adopt certain content moderation strategies – as recently introduced in other major democracies such as India and Indonesia – also does not rhyme with the Dutch government’s liberal principles. Still, the government supports the initiatives of the EC that impose (self-)regulatory frameworks for social media platforms. Additionally, it encourages technology companies to step up their game when it comes to recognizing and reacting to disinformation on their own platforms, because this is not perceived as a governments’ task. Thus, the common thread throughout all documents is that the Dutch government clearly recognizes this problem in society and has ideas on how this could be tackled. Solving this problem certainly has high priority, yet the government recognizes that it does not have the primary responsibility to face it on the level of content, but rather by enhancing the public’s and individual’s defense systems. Furthermore, the Dutch government recognizes that other actors – notably the social media platforms and the European Union – have an equally large, if not larger, responsibility to come up with effective solutions.

## **4.2 International level: European Commission (EC)**

A sufficient analysis on this topic is not complete without the incorporation of the international perspective. The importance of international cooperation is frequently acknowledged by national governments such as the Dutch, as explained in the previous section. The problem of disinformation is widely recognized as a cross-border issue, and the EC has been working on concrete solutions and strategies during the past few years – setting up special committees in parliament, social media observatories and fact-checking organizations, among other things. Additionally, the EC introduced a self-regulatory mechanism in 2018 that attempts to set norms for social media platforms – The Code of Practice on Disinformation – and in 2020 it proposed an actual legislative framework – the Digital Services Act – which will set clear regulations for social media platforms when it has been implemented.

In light of these developments, the European Union recognizes itself as the global trailblazer when it comes to effective yet liberal internet regulation and privacy policies – given the reserved character of the United States when it comes to internet regulation, and the fragmentation of other transnational political organizations in Asia or Africa. Hence, in the words of Bart Groothuis, “there is only one option: Europe. The norms we set will be the standard that [the rest of the world] will adopt”. Indeed – as became clear in the analysis of the Dutch government’s approach towards disinformation – in the Netherlands there is a lot of focus on European cooperation and coordination. Therefore, the next section discusses the approach that the EC takes on this topic.

### ***Diagnostic Framing***

#### Perception of the problem

In early 2018, the EC assigned an High Level Expert Group (HLEG) – a team of smart minds that includes internet researchers, experienced journalists, employees from social media companies, cybersecurity experts, etc. – to clearly diagnose the problem that Europe is facing when it comes to disinformation. Consequently, this diagnosis was used to propose various solutions to it. Starting with a clear definition, disinformation is spelled out as “verifiably false or misleading information”, which is created and disseminated for economic or political gain, with the intention to deceive the public. Additionally, it “may cause public harm” in the form of threats to democratic processes and public goods. Hence the problem is

---

<sup>8</sup> Interview BZK, 21 April 2021, Utrecht.

<sup>9</sup> Interview Bart Groothuis, 20 April 2021, Utrecht.

perceived as “a major challenge for Europe”. The perceived dominant aspects of the problem – which are frequently emphasized throughout the analyzed documents – are its sheer complexity and sensitivity, due to the fact that it is multi-faceted, multi-stakeholder and ever-evolving. The problem is also strongly tied to the development of new digital media. Social media platforms and online messaging services play an important role in the distribution of disinformation, and its role can therefore not be underestimated.

### Important consequences

The EC is clear and comprehensive when it comes to their understanding of the consequences that disinformation could have on European society. They spell out clearly that the problem can reduce trust in governmental institutions and media, support extremist beliefs, sow distrust and tension in society, endanger election integrity and influence policy and debates particularly on sensitive topics – such as migration, climate change and public health. Specifically, there are deliberate attempts to undermine the legitimacy and trustworthiness of the European Union, fueling the already present anti-EU sentiment across the continent. For example, this happens when false stories about the legislative abilities of the European Parliament are repeatedly disseminated, or when the “European Deep State” gets the blame for the fall of the Austrian government. In the end, just as the Dutch government, the EC perceives disinformation as a threat to democratic societies at large.

### Blame and causality

Similar to the Dutch government, the attribution of causality is roughly twofold here. The EC recognizes that this problem exists by the virtue of various (1) actors – either individuals or state-sponsored – and (2) infrastructures. So firstly, a part of the blame gets attributed to foreign state actors that have an interest in undermining the political stability within Europe. In this narrative, disinformation is used as a geopolitical tool by bad actors, particularly in the form of coordinated disinformation campaigns. Unmistakably, Russia is the main antagonist in this context, but other actors such as China and – to a lesser extent – Iran are also emphasized. Yet domestic individuals may also play a role in the dissemination of disinformation, for example in the case of COVID-19 and vaccinations. The second part of causality relates to the (digital) infrastructure on which the disinformation spreads. Hence, to a certain extent it refers to the way how social media platforms operate, or how the field of misleading information is expanding due to the development of advanced artificial intelligence and synthetic media.

When Bart Groothuis was asked who or what should get the majority of the blame, he acknowledged this two-way split in attribution of causality. Yet he added his own twist on the narrative, pushing for much more focus and strictness on the *first* aspect, which could lead to direct sanctions or penalties directed at the actors who coordinate disinformation campaigns for a strategic goal:

“It’s both, but the problem is mainly that [the European Union] mostly looks at this problem as a technological problem that you can solve by improving algorithms, transparency, media literacy, education, and all that stuff. Sure, that’s all true, and maybe there should be more of this. But what I’m trying to say here – and I think that I’m one of the few, or maybe the only one, in Brussels that actively promotes this narrative – is that we shouldn’t merely look at those conditions, but also ask ourselves: ‘who exactly are these actors that exploit these favorable conditions? Who are these people?’ (...) Shouldn’t we think strategically to increase the costs of those who [disseminate disinformation]?”<sup>10</sup>

---

<sup>10</sup> Interview Bart Groothuis, 20 April 2021, Utrecht.

## *Prognostic Framing*

### Proposed solutions

Because the EC recognizes that this problem is multi-faceted and involves many stakeholders, it emphasizes the different roles that various stakeholders could take in tackling this problem. The social media platforms have an important role to play, as well as media organizations and independent research communities. Therefore, an holistic approach is required. The consequent strategies by the EC may be broadly summarized in three major categories: (1) increasing the long-term resilience of European citizens and organizations by enhancing knowledge and awareness around disinformation; (2) monitoring the evolving nature of the problem constantly, so adequate responses keep developing and can be tested on its efficiency; and (3) pushing social media companies to introduce changes on their platforms. The last point is emphasized by Bart Groothuis:

“Confronting tech companies is core business of Europe. If the Commission doesn’t deliver on that aspect, they will have problem with the citizens, the Parliament and the Council. That’s not OK. Every country faces this problem. And every country knows, you have to solve this via Brussels.”<sup>11</sup>

### Concrete strategies

The first strategy relates to awareness of individual citizens, and is therefore similar to the main approach of the Dutch government. There are deliberate efforts to promote media literacy both at the Union and Member States levels, so European citizens are capable to better identify and deal with disinformation online. The open and plural media landscape of Member States is also reinforced by further support of investigative journalism and independent media. These measures should create long-term resilience.

Second, comprehensive insights on the ever-evolving nature of disinformation and its host technologies will always be required. The EC therefore funds independent research on this subject and supports the creation of non-partisan fact-checking organizations. In this light, the European Digital Media Observatory has been established, which functions as a knowledge hub for fact-checkers, researchers and other stakeholders to support policy makers with up-to-date insights. Additionally, the EC set up programs to quickly recognize disinformation campaigns and spread the information to Member States, such as the Rapid Alert System and EUvsDisinfo.

Third, there are efforts to push social media companies to introduce significant changes to their platforms. This started with the introduction of the Code of Practice on Disinformation, presented by the EC in 2018. “The Code” was signed by major social media platforms – such as Facebook, Twitter, Google and TikTok – and gave them a more tangible responsibility to take measures against disinformation. The signatories of the Code agreed to focus on reducing opportunities and economic incentives for actors that spread disinformation, enhancing transparency around political advertising, taking action against malicious actors, setting up features that prioritize trustworthy information, and collaborating with fact-checkers and the research community. This self-regulatory mechanism is a short-term solution, because it is relatively easy and quick to implement such a measure. It is perceived as a good first step, but as critics and the official assessments note, the Code still has various shortcomings. These concern inconsistent and incomplete application, confusion around definitions, areas that the Code does not cover, and limitations that are inherent to the self-regulatory nature of this code. On this last point, critics such as Bart Groothuis even argue that self-regulatory practices will “never work”, so more distinct sanctions and legal backing would be needed. The Digital Services Act – if done right – may give these aspirations the necessary legal backing. However, at the time of writing it is too early to include this draft legislation into the current analysis.

---

<sup>11</sup> Interview Bart Groothuis, 20 April 2021, Utrecht.

## *Motivational Framing*

The EC pushes for more action on the side of the social media platforms, because it recognizes that the platforms and the advertising industry have a significant role and responsibility in countering the disinformation problem. This is where the EC can go beyond national governments, because of their larger leverage over these powerful technology companies. It can push social media platforms to “take their responsibility”, whether that is through self-regulation or rigorous legislation. However, throughout all documents, the EC makes one thing particularly clear: “any form of censorship – either public or private – should clearly be avoided”. Freedom of expression is a fundamental right that should be protected, inherent to a liberal European society. Besides this normative argument that is repeatedly iterated, the HLEG report also stresses a second-order effect of “censorship”: the possibility of a “backfire effect” that could come from any involvement of governments on the level of content. This involvement can be used by the spreaders of disinformation to create and reinforce an “us versus them” narrative, in which “the establishment” or “the political elite” can be “(mis-)perceived as manipulating the news to their advantage”.

Now, as the Commission itself acknowledges, this creates a delicate balance between wanting to fix the problem of disinformation on the one hand, while preserving the absolute right to freedom of expression on the other hand. The EC does not want “governmental control” on digital media or become the arbiter of truth. However, if the EC is dedicated to tackle the problem of disinformation, it may unknowingly steer itself into tricky territory and face criticism and controversy along the way. Given the novelty and impact of disinformation on social media platforms, this balance is still a contentious topic for discussion:

“There is a difference between freely expressing my opinion, and having my opinion being heard on a massive scale [via social media]. Do I have that right too? What exactly is freedom of expression? That is a new discussion.”<sup>12</sup>

---

<sup>12</sup> Interview Bart Groothuis, 20 April 2021, Utrecht.

## 5. Social media platforms

### 5.1 Facebook

“Overall, I would say that we’re going through a broader philosophical shift in how we approach our responsibility as a company. For the first 10 or 12 years of the company, I viewed our responsibility as primarily building tools – that if we could put those tools in people’s hands, then that would empower people to do good things. What I think we’ve learned now (...) is that we need to take a more pro-active role and take a broader view of our responsibility. It’s not enough to just build tools. We need to make sure that they’re used for good. And that means that we need to means that we need to take a more pro-active view in policing the digital ecosystem.”

– Mark Zuckerberg <sup>13</sup>

In merely fifteen years Facebook grew from a modest and fun website for college students into a larger-than-life business, making billions of dollars of profit annually, employing almost 60.000 people and connecting almost three billion users on its platform. The influence that Facebook has had on the public conversation during the previous decades simply cannot be underestimated. As explained in previous chapters, Facebook has been under increased scrutiny to fix some of the problems that its social media platform faces, including the problem of disinformation. Consequently, Facebook is now willing to come to terms to some of these problems, as the quote by Mark Zuckerberg highlights. The following section discusses the analysis of the relevant documents that Facebook published on the topic of disinformation (n=15). This analysis is supplemented with insights from an public policy expert who now works for Facebook in the Benelux.

#### *Diagnostic framing*

##### Perception of the problem

Facebook understands that it needs to establish clear definitions in order to diagnose the problem. However, the word “disinformation” is sporadically used. Instead, Facebook defines the main problem as Coordinated Inauthentic Behaviour (CIB), which refers to “groups of pages or people work together to mislead others about who they are, and what they’re doing”. The actors behind this coordinated activity use fake accounts – either automated or human-curated – to mislead people about who they are and what they think. Their tactics include spreading of false information and made-up stories, pushing certain narratives, steering public discourse on specific topics. One order of magnitude above CIB is the Influence Operation (IO), which may be synonymous to a “disinformation campaign” as the rest of the world understands them. As such, an IO can consist of a network of different CIB’s. Facebook refers to the Russian interference in the 2016 US presidential election as the most prominent textbook example of an IO.

Importantly, the problem of disinformation is thus always specifically framed in terms of misleading *behaviour*, rather than misleading *content*. When it comes to content, Facebook is referring to “harmful misinformation” rather than disinformation. The scope of harmful misinformation includes fact-checked false information about elections – e.g. false claims about how citizens can vote – or public health – e.g. false claims on COVID-19 or its vaccines. Here, the content of a post is false, which is a reason to act on it. Yet with disinformation, its deceptive behavior and intent is the reason to act on it.

---

<sup>13</sup> A quote of Facebook’s CEO, Mark Zuckerberg, when testifying for the US Senate Judiciary Committee in April 2018. Source: <https://www.youtube.com/watch?v=ByLTX05jSY>

Furthermore, Facebook repeatedly emphasizes that the problem of disinformation on social media is part of a bigger story. The company notes that disinformation never gets disseminated in a vacuum. It is multi-faceted and it involves many stakeholders. Other social media platforms, traditional media, governments and civil society all play an important role in reducing the impact of this problem. Hence these other actors also have a responsibility to work on solutions. This is highlighted by the following quote from the Facebook employee:

“If one says that disinformation on social media is a problem, than I think one’s vision is too narrow. Or at least you don’t get the full picture. Disinformation is a problem for democracy, maybe. But disinformation on social media is just a small piece of the whole puzzle. On its own, [merely focusing on social media] doesn’t say enough to come up with good solutions.”<sup>14</sup>

### Important consequences

The main consequence of deliberate manipulation of the public debate is the negative impact on the quality and honesty of civic discourse on the platform. It is also framed in terms of “corrupting” civic discourse, “deceiving” unsuspecting citizens or “misleading” people in public debates. This is specifically the case in the context of democratic elections, where online discourse can influence the results of an election directly. Facebook recognizes that the integrity of elections and public institutions are in danger when citizens do not have access to accurate information and are misled for a strategic goal.

### Blame and causality

The blame is almost exclusively attributed to “malicious actors” that engage in the dissemination of disinformation for ideological, political or economic purposes, and therefore corrupt the public conversation on the social media platform. Facebook publishes monthly CIB reports in which the company clarifies what coordinated disinformation networks they detected and removed. They are rather extensive about this; these reports include the location, motives and timelines of the perpetrators, visualized in nicely polished graphs and figures. Additionally, Facebook provides screenshots to illustrate examples of disinformation and deceptive behaviour on their platform. In the beginning – around 2017-2018 – these actors mainly originated in Russia. Later, this broadened to include “attackers” from many other states all around the world, which attacked both domestic and foreign targets. Now, this kind of language is also of importance here: *attackers / defenders, malicious actors, fighting, arms race*. In Facebook’s diagnostic framing there is no shortage of these kind of references, indicating a perceived antagonism between Facebook on one side, and these malicious actors on the other side.

### ***Prognostic framing***

#### Proposed solutions

Facebook stresses that it always tries to stay on top of the problem of disinformation. Therefore, gathering information and insights on CIB’s and IO’s to improve their responses in an evidence-based way is deemed to be crucial. Their antagonists keep adapting their tactics, so this information-gathering is a constant “arms race” – because after all, “the more we know about a threat, the better we’ll be at stopping [it] in the future”. Solutions will therefore always be highly *dynamic* by nature, continuously evolving to counter emerging threats or expose weaknesses in Facebook’s defenses.

Facebook emphasizes that they should not be alone in this fight and thus advocates for *cooperation* with all involved stakeholders. As an important example, the company calls for more effective government regulation on social media platforms, especially since 2019. In the words of Mark Zuckerberg himself: “by updating the rules for the Internet, we can preserve what’s best about it, (...) while also protecting society

---

<sup>14</sup> Interview Facebook, 30 April 2021, Utrecht

from broader harms”. Hence the company is working together with lawmakers to come up with effective and fair regulation, as the Facebook employee explained:

“There are conversations on a frequent basis with ministries such as BZK, where (...) we try to share what we see and what we run into, try to encourage them to come up with specific regulations, and also share our knowledge and dilemmas.”<sup>15</sup>

Yet, rather than merely waiting for government regulation, Facebook has also been creating its own policies in order to reduce the impact of disinformation. Hereby the main objective is to – in their own words – “respond to information operations”. The majority of the prognosis is framed in terms of this *response* by Facebook – a response to CIB, a response to fake accounts, a response to false information. The next section discusses these responses, which can be divided into two over-arching strategies: responses to misleading content (the area of misinformation) and responses to misleading behaviour (the area of disinformation). Both policy areas are relevant to this research, because disinformation tactics can include the dissemination of misinformation.

### Concrete strategies

#### 1. Content

Simply put, this is about content moderation. The strategy here is threefold: *remove, reduce, inform*. First, *reduce* involves the removal of posts and accounts that violate the Community Guidelines. Hence, Facebook removes fake accounts – “millions of them every day” – by using automatic detection methods. Additionally, “very specific” harmful misinformation that directly endangers election integrity or public health will also be removed from the platform. Second, *reduce* is about diminishing the impact of other content that is rated as false or misleading by independent fact-checkers. If the content does not cross the threshold of removal, it will get a warning label and its online reach will be artificially limited – appearing lower in people’s News Feeds, losing the ability to share the content, and so forth. Third, *inform* is about giving users more context about certain information. This includes transparency about pages, posts and advertisements. Facebook also invests in enhancing people’s awareness about disinformation via the support of media literacy campaigns, in cooperation with governments and civil society groups.

#### 2. Behaviour

When it comes to creating policies against certain behaviour, things may get more complicated. These policies concern CIB’s and IO’s. When Facebook takes action against one of such networks, “it is because of their deceptive behaviour – it is not because of the content they’re sharing”. The posts themselves do not necessarily have to include misinformation (yet in many cases they do). In order to detect deceptive behaviour in the form of CIB’s, Facebook uses a combination of human employees – such as “election integrity teams”<sup>16</sup> that monitor local elections around the world in real-time and make sure that there are no successful attempts by malicious actors to influence its outcome – and automated detection tools that look for bot activity, spammy behaviour or suspicious behaviour coming from foreign countries. Facebook acknowledges that recognizing the difference between authentic/real behaviour and inauthentic/deceptive behaviour on a large scale is quite tricky. Hence the company stresses that it will keep investing in the development of more sophisticated AI-enhanced detection tools, hoping to eventually catch all inauthentic behaviour in its defensive net.

---

<sup>15</sup> Interview Facebook, 30 April 2021, Utrecht.

<sup>16</sup> During the United States’ mid-term elections and Brazilian presidential elections – both in 2018 – this election integrity team operated from a so-called “war room”.

## *Motivational framing*

For Facebook, freedom of expression is regarded as a “core principle”. The idea of giving everyone around the world a voice is one of the democratic foundations on which this company is built. Yet Facebook now recognizes that there are limits to absolute freedom of expression online, and these limits concern – among other cases such as hate speech and incitement of violence – the topics of mis- and disinformation. The company attempts to make these limits as clear as possible by inscribing it in their Community Guidelines, although they also do not want to become the sole “arbiter of truth”. Therefore, Facebook advocates for solutions that are cooperative by nature. Motivationally speaking, this is rooted in the assumption that the problem of disinformation is not just Facebook’s problem – it is also the responsibility of other relevant actors in society. Hence, these actors must face it together, because “these threats extend beyond any one platform and no single organization can tackle them alone”. As the Facebook employee notes:

“We cannot merely rely on the medium to be the police officer. Obviously we have to make sure that all actors can take their responsibility.”<sup>17</sup>

Simultaneously, Facebook also emphasizes that it certainly recognizes its own responsibility as a company to fix the problem of disinformation (CIB & IO) on its platform. As the company writes: “we have an important role in protecting people and public debate on our platform, and we are focused on that mission”. Bottom line is that Facebook wants to make sure reaffirm to world that the company is committed to shaping online discourse that solely displays authentic behaviour, authentic opinions and authentic engagement. It attempts to reach this goal by setting up robust policies against deceptive behaviour. However, Facebook recognizes that this will nonetheless always be difficult, because when disinformation campaigns “evolve in response to better enforcement, they will continue to blur the boundary between legitimate advocacy and illegitimate manipulation”. In other words, the boundaries between authentic behaviour and inauthentic behaviour will increasingly become more vague which in turn makes enforcement more difficult. This inevitable loop yet again captures the delicate balance between freedom of speech and the response to disinformation: Facebook is committed to enabling online freedom of expression, but what exactly constitutes freedom of expression when seemingly authentic speech can covertly be inauthentic?

---

<sup>17</sup> Interview Facebook, 30 April 2021, Utrecht.

## 5.2 Twitter

“We believe in free expression. We believe in debate and conversation to find the truth. At the same time, we must balance that with our desire to secure our service not to be used to sow confusion, division or destruction. This makes the freedom to moderate content critical to us.”

– Jack Dorsey, 2021 <sup>18</sup>

Although Twitter’s size as a company is smaller than Facebook – in terms of its net-worth and the amount of employees – Twitter has also become a key figure in the disinformation debate during the last five years. It is the company’s self-proclaimed mission to shape the public conversation online. So, when news stories and investigation reports were published about major disinformation campaigns that were unknowingly facilitated by the social media platform, Twitter also found itself under increased scrutiny and endured more public pressure to act on this problem. To illustrate: from October 2017 to March 2021, Twitter’s policy representatives appeared in 17 different Congressional Hearings in Washington DC, many of which were about disinformation and disinformation campaigns. The company’s attitude towards the problem, its introduced solutions and the rationale behind them can be found in Twitter’s relevant blog posts on this subject. The next section discusses the analysis of these documents (n=16). Unfortunately Twitter’s Public Policy department was not available for an interview.

### *Diagnostic framing*

#### Perception of the problem

Frankly the analyzed documents do not provide a clear-cut problem statement about disinformation on Twitter. The company preferably discusses the proposed solutions rather than the underlying problem. Still, particularly during the early phase – 2017, 2018 – the problem was merely framed in terms of *bots*. Twitter defines a bot simply as an automated account, which can – due to its automatic nature – quickly pour large quantities of information into a digital ecosystem. These bots are used on Twitter to spread and amplify particular information, push certain hashtags and trends, generate fake engagement or flood conversations with nonsense. Hence the existence of these bots on Twitter’s platform is perceived to be the main problem that obstructs healthy online conversation. However, during the second phase – 2019, 2020 – Twitter expanded the scope of this problem, now focusing more broadly on “the holistic behavior of an account, not just whether it’s automated or not”. The company calls this “platform manipulation”, and this comes very close to the established term “disinformation” used in this research. It describes the way in which Twitter accounts are used – whether they *behave* in a malicious and coordinated fashion, *intending* to undermine healthy discourse, many times by using false stories. In the third phase, since the unfolding of the COVID-19 pandemic, the scope of the problem has broadened once more. It reached the level of *content*, now including certain misinformation about the coronavirus and its vaccines which is factually wrong – e.g. that the virus would be caused by 5G networks or the vaccines contain microchips – and can therefore be acted upon.

---

<sup>18</sup> A quote by Twitter’s CEO, Jack Dorsey, when testifying for the US Communications and Technology Subcommittee and the Consumer Protection and Commerce Subcommittee in March 2021. Source: [https://www.youtube.com/watch?v=UipFasl\\_nD0](https://www.youtube.com/watch?v=UipFasl_nD0)

### Important consequences

Twitter repeatedly claims that the problem of bots and platform manipulation undermines “healthy conversation” on their platform. If the information ecosystem is overwhelmed with false stories and fake engagement, users cannot have an honest debate. Given the large scale of Twitter, the company acknowledges that this mechanism can also translate to the offline world, which is harmful to civic and political discourse and it puts the integrity of democratic elections in jeopardy. Additionally, in the case of COVID-19 misinformation, the most important consequence is that it directly endangers the health of people in the real, offline world.

### Blame and causality

Initially the blame is mainly attributed to the existence of bots and automated accounts, and to the actors who create them. Twitter recognizes that many of these bots originate from Russia – especially in the context of the US elections, where they were explicitly linked to the Russian government – but the company also reports on large numbers of automated accounts from other countries all over the world. However, after the scope of the problem has been broadened, the perceived cause has shifted as well. It can now be any account that shows malicious behaviour as defined by Twitter’s Rules. Sometimes this behaviour is linked to foreign state actors, sometimes it is linked to domestic coordinated actors or lone wolves with ideological intentions.

### ***Prognostic framing***

#### Proposed solutions

Twitter acknowledges that this is a difficult problem that keeps evolving, so its solutions need to be dynamic and evidence-based. Therefore, Twitter proposes a broad range of solutions to this problem – some are similar to those of governmental institutions, some are specific to Twitter. These policy interventions can be divided into three main categories: (1) fighting bots, (2) improving content moderation and (3) enhancing awareness and transparency.

#### Concrete strategies

First, the main threat to healthy conversation in the early phases was the prevalence of malicious bots on Twitter. Thus, much time and energy was devoted in “fighting” these bots. When these automated accounts are discovered – either by human detection or AI detection tools – they are simply suspended. Twitter still takes down millions of bots every week, and claims it has become very effective at this front. The fight against bots is continuous, and the company has set up clear rules against the use of automation in order to counter “platform manipulation”. Yet since 2019, Twitter has broadened its prognostic frame as a result from a changing diagnosis. It now takes actions against accounts that show deceptive behaviour, not merely against automated accounts. Second, Twitter exercises content moderation to counter false information on the platform. As a private company, Twitter has the authority to determine what kind of content is preferred, allowed and prohibited on their platform. The company was initially quite cautious at this front, but as the problem definition broadened, the toolbox expanded as well. Since 2020, Twitter takes specific action against posts that contain factually false information about election integrity or public health. When this is the case, the post will get a label that it is misleading information – e.g. the earlier-mentioned Donald Trump’s statements about election fraud or Thierry Baudet’s posts about COVID-19 – or the post or account may be removed. This truth-determining is done in cooperation with experts and non-partisan fact-checking organizations globally. The third branch of strategies concerns enhancing awareness, transparency and cooperation. This includes supporting media and information literacy campaigns over the world, imposing more transparent advertisement policies and working in cooperation with many organizations such as governmental institutions, local law enforcement agencies and research communities.

### ***Motivational framing***

The rationale behind these policy interventions is anchored in the mission and vision of Twitter. The company is quite vocal about their motivations, which can be divided into three categories. First, *Twitter aims to serve the public conversation*. This catchphrase is used in almost every document. Twitter wants to create a public digital space in which conversation is healthy and safe – during times of crises, elections, and beyond. The company repeatedly makes statements in the following fashion: “Twitter is committed to serving the public conversation [around this particular event]”, or “it is Twitter’s mission to shape global healthy conversations”. In this context, defensive language is frequently used, such as to “protect” or “defend public conversation”. Misinformation, manipulative behaviour or non-human activity undermines the integrity and honesty of this public conversation – which is why these elements are undesirable on the platform, and should be taken care of.

Following from this, *Twitter recognizes that they also have the responsibility to act*, as the company sees that their online social media platform can have a major impact on the offline world. They explicitly mention that they are willing to play their part, and that is why they perceive their policy interventions to be legitimate. However, this sense of responsibility should – in their framing – not be exclusive to Twitter. The company is keen to point out that other relevant actors also need to acknowledge their role in keeping civic and political discourse healthy and safe, such as governments, media institutions, NGO’s, users, other technology companies, and the public at large. In Twitter’s words: “we’re in this together”.

Third, *Twitter emphasizes its democratic values*. The company is built upon the foundation of free flow of information and the right to free online expression, and it emphasizes that it wants to make a positive impact on the world. Hence, Twitter is “committed to providing a service that fosters and facilitates free and open democratic debate”. It actively encourages people to participate in civic process – for example by voting in democratic elections when they have the opportunity – and be socially engaged citizens. Most policy interventions are rooted in these democratic underpinnings, such as the partnerships with various (democratic) institutions, the support of media literacy programs and the enhancement of transparency. However, although Twitter claims that it does not want to be “the arbiter of truth”, it might be moving into this direction since the rigid content moderation policies on public health or election integrity misinformation were introduced. This delicate balance between protecting freedom of expression and removing harmful disinformation – as the quote by Jack Dorsey at the beginning of this chapter illustrates – creates friction that has caused a fair amount of public controversy.

## 6. Internet Freedom Organizations

This actor consists of three relevant non-profits: Article19, Electronic Frontier Foundation (EFF) and Bits of Freedom. These non-governmental organizations have been concerned with the response to disinformation by governments and social media platforms during the last few years. As important free speech watch dogs they function as a check on the status quo in this realm. Hence their primary concern is to advise, criticize and amend policies that might jeopardize freedom of expression in the digital domain. Additionally, the technical expertise of these organizations – who employ computer scientists, information law attorneys and policy advisors – offers relevant perspectives that may be missed in political discourse.

The three organizations all advocate and lobby for a free and open internet, but they are characterized by different focal points on the topic of disinformation. First, EFF is primarily concerned with content moderation strategies of social media platforms, and what this does to freedom of expression around the globe. Second, Bits of Freedom mainly emphasizes the perverse business models of the major social media platforms, and criticizes the monopoly position that these companies hold in the global digital communication landscape. Lastly, Article19 can be characterized as more social rather than technical, putting the issue in a broad societal context. They are mostly concerned with government regulations to disinformation and examines its adherence to International Human Rights Law. The following section offers an analysis of relevant documents published by Article19 (n=10), EFF (n=10) and Bits of Freedom (n=5). This is combined with insights from an expert interview with Rejo Zenger, senior policy advisor at Bits of Freedom.

### *Diagnostic framing*

#### Perception of the problem

Although these organizations acknowledge the problem of disinformation on social media and the challenges it poses to democracy, they do not spend as much time on defining this problem in itself. Rather they are challenging the policy *responses* to it, which may also create substantial problems to democracy. As EFF puts it: “the abundance of information in the digital world should not be deemed, in itself, a problem. But the responses to the “fake news” phenomenon (...) could be.” These can be categorized in responses by social media platforms and by governments.

On the one hand, EFF is mostly focused on the way that *social media platforms* deal with mis- and disinformation. They are primarily concerned about the platforms’ increasingly strict content moderation strategies that remove specific content and suspend users. The EFF notes that these strategies are (a) too opaque, (b) prone to error – many innocent users and posts are removed by accident – and (c) inconsistent in their appliance. Although many of these strategies are deployed in good faith in order to fight mis- and disinformation, a handful of companies in Silicon Valley without democratic control might nevertheless become “international speech police” by engaging in “private censorship”. This inherently threatens free and democratic societies around the world. Bits of Freedom adds to this perspective by emphasizing the disproportionate power that these few social media platforms have on shaping the digital communication infrastructure. “A few giant corporations dominate the ecosystem” and wield an almost monopoly-like control over the public debate. Furthermore, Bits of Freedom points out that these companies operate on a “toxic business model” which relies on commodifying human attention – with all intention to keep users on their platform as long as possible, and sell this attention to advertisers. More extreme, outraging and possibly misleading content tends to engage more attention. Thus Bits of Freedom emphasizes that this exact mechanism is what makes disinformation thrive on the social media – it is almost natural given their business model.

“I think there certainly is a problem, but it’s not the problem of disinformation in itself. Rather, the problem is that we’re way too dependent on these large companies, who ultimately decide how our public debate is facilitated.”<sup>19</sup>

On the other hand, Article19 puts more focus on *governments* rather than the social media platforms, and it emphasizes that certain governmental responses to disinformation – e.g. by legal measures prohibiting specific types of content – can cause indirect censorship and put freedom of expression in serious jeopardy. They repeatedly emphasize that this would directly infringe on Article 19 of the Universal Declaration of Human Rights (hence the name of the IFO): “Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers”. Certainly this fundamental human right may not be violated on what they call “vague and ambiguous ideas”, arguing that there is not yet a clear and internationally agreed upon definition of “disinformation”. Different countries and different companies make their own distinction in what constitutes as disinformation, misinformation or fake news. They argue that sometimes there is not even a clear distinction between what information would be “true” or “false” – in a lot of instances, there are many plausible shades of grey. All in all, these circumstances make the possibility of fair and effective legislation that focuses on content – that are also in compliance with international human rights norms – quite unlikely.

#### Important consequences

The most important consequence – which is repeatedly mentioned by all organizations – is the risk of losing freedom of expression and free flow of information on the Internet. If governments and social media platforms are over-reacting on the issue of disinformation, “we might be sleepwalking into censorship”, as Article19 puts it. This is already happening in some countries, where questionable regimes have introduced “fake news bans” or “disinformation laws” that provided quasi-legitimate reasons to silence political opponents and marginalized groups. However, at the same time these organizations also recognize that disinformation *in itself* can pose a threat to democracy as well, although this is more scarcely mentioned. The main concern here is foreign election interference by disinformation campaigns. Additionally – in the context of the COVID-19 pandemic – they recognize that disinformation can pose a threat to global public health.

#### Blame and causality

When addressing the root causes of the problem of disinformation on social media, these three actors unanimously agree on an important factor: the way how social media platforms operate. EFF frames this in terms of algorithms “that decide what content users see” and are not transparent. Bits of Freedom refers to this as social media’s “toxic business models” that commodify human attention and select for more extreme content. Article19 stresses the lack of diversity in content that users get to see on social media platforms. All in all, these mechanisms enhance the spread of online disinformation and blur the psychology of the individual user, who is increasingly unable to separate fact from fiction. In short, most of the blame lies with infrastructure rather than actors.

---

<sup>19</sup> Interview Rejo Zenger, 31 May 2021, Utrecht.

## ***Prognostic framing***

### Proposed solutions

These organizations agree that disinformation as a problem has been around for centuries, but the inner workings of our digital ecosystem amplifies this problem to a larger scale. Yet the challenge is to react to the right aspects of the problem and avoid an over-reaction that jeopardizes freedom of expression online. Therefore, as one of the most important solutions, social media companies should (a) adopt *radical transparency* in every possible way. Governments and relevant stakeholders should also invest in (b) *preventive measures*. Furthermore, some form of regulations are needed to properly (c) *regulate the digital communication market*.

### Concrete strategies

This *radical transparency* principle refers to multiple mechanisms. First, social media platforms should adopt very clear and distinct content moderation policies, and be transparent in how they are creating and enforcing these policies. The enforcement should be as minimal as possible. The way how their algorithms work should also be completely transparent, as well as the way how advertisements are served. All in all, these IFO's argue that some transparency improvements by social media platforms "are encouraging, but don't go far enough" – social media platforms should step up their game on this front. Second, governments and other stakeholders should be transparent and participatory in their efforts to counter disinformation. For example, this could be achieved – as Article 19 proposes – in the form of a "Social Media Council", where all stakeholders meet and discuss their strategies in an open, accountable and transparent forum.

*Preventive measures* refer to the empowering of the individual user, to make it more resilient against online disinformation. This can be done by media literacy campaigns and the promotion of a diverse media landscape. Social media platforms should also promote more diversity on their platforms. Furthermore, reliable and accurate information should be easily accessible, especially in the context of the COVID-19 pandemic. These measures should empower users to separate fact from fiction and be exposed to a broader range of perspectives, rather than solely being trapped in one's own online echo chamber.

And if there is to be legal regulation, these laws should regulate the digital communication *market*, not the *content* on that market. The IFO's note that a few Silicon Valley companies and their business models are too dominant in today's communication landscape. The digital ecosystem needs more healthy competition, more accountability and more systemic changes. Especially Bits of Freedom is vocal on this matter. They recognize that laws are necessary – self-regulation is "not effective" – but these laws would require a dynamic and holistic approach. These laws should cover, among other things, mandatory transparency regulations on algorithms and content moderation; mandatory cross-functionality of platforms; and a complete ban on personalized advertisements. When the digital communication infrastructure is properly fixed, the problem of disinformation should naturally diminish. In any case, regulations should never consider "judging information on its content", which creates "a legal duty on truth", which would be essentially censorship in disguise.

"It doesn't make a lot of sense to us to merely focus on a company like Facebook. One can introduce another 100.000 rules for Facebook, but that won't solve those problems such as disinformation. For us, Facebook in itself isn't interesting. We'd rather take a step back, take a look at the entire [digital communication landscape], and try to make a change there. I think that's where the real solution is."<sup>20</sup>

---

<sup>20</sup> Interview Rejo Zenger, 31 May 2021, Utrecht.

### ***Motivational framing***

By now it is clear that all the work of these IFO's is primarily done in the name of protecting freedom of expression online. This fundamental bottom line is repeatedly emphasized in every document and statement. Transparency and democratic accountability are lacking in the few companies that essentially decide what can and cannot be said online, also in the companies' well-intentioned efforts against disinformation on their platforms. So, these IFO's argue that the questions about disinformation should be taken very carefully – it is maybe better to do too little than too much on this regard. If there is something that can be done against the dissemination of disinformation online, it should be tackled at its root cause: the digital infrastructure in which it thrives. As Rejo Zenger stated bluntly in the interview:

“Not the content on Facebook is the problem... Facebook is the problem.”<sup>21</sup>

In order change this digital infrastructure on a more systemic level, legal regulations are still necessary. This is why these organizations work together with governments to give advice for effective and fair legislation that tackles the *right* aspects of the problem. Bottom line is that every diagnosis and/or prognosis is – or should be – always determined in its connection to human rights. In the framing of IFO's, diminishing the impact of disinformation does not have to diminish the impact of online freedom of expression. Society can and should have both. The organizations are convinced that this is true, and the defense of this belief is their main motivational frame.

---

<sup>21</sup> Interview Rejo Zenger, 31 May 2021, Utrecht.

## 7. Discussion

The previous chapter presented the research findings that flowed from the analysis. These findings were guided and structured by the first two sub-questions. This chapter reflects upon these findings and discusses the main similarities differences in framings and motivations by the three different actors, which connects to the third sub-question. Similarities and differences between diagnostic, prognostic and motivational framings of the three relevant actors emerge from six major categories: the (1) shared understanding of the dynamic nature of the problem, (2) difference in attributing causality, (3) difference in understanding transparency, (4) different views on content moderation and fact-checking, (5) shared focus on media literacy and individual responsibility and (6) different priorities with regards to freedom of expression.

First, when it comes to diagnostic framing, there is a wide recognition among governmental institutions and social media platforms that disinformation on social media is a significant problem that needs to be constant attention. It can be harmful to democracy, it can undermine the integrity of elections and erode trust in democratic institutions. Governments strike the most alarming tone on this front. Social media platforms acknowledge the problems but are more focused on the solutions. IFO's are quite alarming in their diagnostic framing too, but they have a different perspective on the nature of the problem, which connects to the point below.

Second, there is one crucial diagnostic frame that is repeatedly emphasized by IFO's, but not echoed by the governmental institutions and social media platforms. This concerns the systemic features of the social media infrastructure itself – its content moderation strategies, business models and market domination in the digital communication landscape – which is framed by IFO's as the main cause of the problem. This perspective is reminiscent of scholarly work such as P. N. Howard (2020) and Zuboff (2019), because it puts a spotlight on the digital infrastructure of social media platforms that ultimately facilitates the spread of disinformation. Although the social media platforms explicitly recognize their own responsibilities in countering the problem of disinformation because it spreads on their platforms, the cause of this problem is framed in terms of the *actors* that make use of their platform. Hence Facebook's and Twitter's main prognostic frame may be summarized as: *if we can just get rid of the bad actors and the fake accounts, we will solve the problem of disinformation*. In other words, social media platforms focus more on the social aspect of the problem rather than the technical aspect (P. N. Howard, 2020). At this point the IFO's would fundamentally disagree, as Rejo Zenger of Bits of Freedom also explicitly emphasized. This is either because the process of “getting rid of the bad actors” by content moderation will inevitably eliminate innocent bystanders and thereby infringe on universal freedom of expression, or because the system on which social media platforms are operated and monetized is inherently going to facilitate the spread of disinformation.

Third, in the prognostic frames of the analyzed actors, transparency is mentioned by all as an important strategy to diminish the impact of disinformation. However, the analysis revealed that, when it comes to “enhancing transparency”, different actors understand this concept differently in practice. Governmental institutions demand that social media platforms enhance their transparency – e.g. via the EU Code of Conduct – mainly by disclosing information about the platforms' efforts against disinformation. Social media platforms agree to this statement, and they are keen to mention that they are already putting in a good effort, for example by publishing monthly disinformation reports. Additionally the platforms emphasize that they enhance their transparency in other ways. For example, by disclosing the information of advertisements, posts and pages to users of their platforms. However, the IFO's are concerned that this is merely PR-talk and that all of these measures are not going far enough. The IFO's threshold for sufficient transparency is certainly higher: they demand “radical transparency” that clearly discloses and legitimizes the content moderation strategies of the social media platforms and opens open the “the black box” of their algorithms. In short, social media platforms emphasize in their prognostic frames that they are doing a fine job, while governments and especially IFO's note that the platforms are not doing enough.

Fourth, the analysis reveals a clash between the prognostic frames of social media platforms and IFO's when it comes to content moderation. On the one hand, social media platforms have been experimenting with different content moderation strategies that involve fact-checking, removing certain content and making changes to algorithms. Facebook and Twitter both invest in better technologies to detect false content and deceptive behaviour, and the companies seem to believe that their work is worthwhile. The platforms see their strategies as a legitimate way to counter disinformation on their platform, which is at least partly supported by scholarly research that shows that these interventions can indeed be effective against the spread of disinformation (e.g. Clayton et al., 2020; Fridkin et al., 2015; Mena, 2020). Yet on the other hand, IFO's are very critical towards this framing. They argue that strict content moderation puts online freedom of expression in serious jeopardy, hence it cannot be the solution to disinformation. This position is also partly supported by scientific research which suggests that prioritizing "the truth" and deleting "falsehoods" – e.g. by fact-checking – is not effective against disinformation (e.g. Chan et al., 2017; Nyhan & Reifler, 2010; Walter & Murphy, 2018). Interestingly, this clear divide in framing between the two actors mimics the division in the scholarly debate on the topic of content moderation. The question of what would be the best way forward is ambivalent both in theory and practice, but social media platforms are nonetheless experimenting – also in cooperation with governmental institutions – with new techniques to counter the spread of disinformation, including stricter content moderation, fact-checking and algorithmic interventions.

Fifth, all actors share another similar prognostic frame, which includes some version of media literacy for citizens in order to enhance social resilience against disinformation efforts. Yet this solution gets prioritized differently by different actors. Governmental institutions see media literacy initiatives as a crucial strategy to solve the problem of disinformation in society. Especially BZK is clear in their ambitions: creating a resilient information society against disinformation in the Netherlands is a prime priority. Additionally, for social media platforms, enhancing media literacy is important but definitely not a first priority. Facebook and Twitter mention that they support various media literacy programs around the world. Yet given the scale of these companies, these efforts are relatively small in comparison to their other policy areas. As for the IFO's, Article19 actively promotes media literacy in their prognostic framing in particular. The organization encourages governments and social media platforms to support media literacy initiatives and reinforce a diverse media landscape, which is also justified by scholarly work such as Möller et al. (2018). This prognosis fits their human rights perspective, as the promotion of media literacy as a strategy will not threaten freedom of expression in any way.

While media literacy initiatives are sustainable strategies that neatly fit the liberal framework of Western states as the Netherlands, it should be questioned if the responsibility to counter disinformation should ultimately lay in the hands of individual citizens. Furthermore, the media literacy strategy seems to contradict important research which shows that – from the receivers' point of view – distinguishing between real and false information and correcting misinformed beliefs will always be a deeply challenging process (e.g. Bode & Vraga, 2015; Lewandowsky et al., 2012; Thorson, 2016). Thus, media literacy initiatives can arguably only work to a certain extent. Nonetheless, in this prognostic frame, the burden of responsibility rather conveniently shifts away from governments and social media platforms towards the individual user. Yet again there is a strong focus on the social aspect of the problem (P. N. Howard, 2020). This strategy avoids any infringement of freedom of expression, but it may also distract from systemic problems that linger under the surface of social media platforms.

Lastly, freedom of expression is mentioned as a crucial motivational frame for all involved actors. However, there is a gradient in the balance between protecting freedom of expression and tackling the problem of disinformation on social media. Governmental institutions emphasize their commitment to freedom of expression many times when proposing any solutions to the problem of disinformation on social media. This is the legitimization of the prognostic frame, which leans heavily on enhancing public resilience against

disinformation through media literacy, cooperating with various actors to strengthen the shared information position, and enhancing transparency of social media platforms. These measures will not infringe on freedom of expression or the free flow of information in any way, otherwise it would be “viewed with suspicion” by civil society, as Bjola (2018, p. 310) also argues.

Social media platforms also repeatedly emphasize their commitment to freedom of expression. In this space, Twitter emphasizes the utter importance of this fundamental human right relatively more often than Facebook. The latter company seems to be more instrumental in its defense of freedom of expression. It is certainly a core value of Facebook, but it recognizes that the distinction between online authentic speech and deceptive speech is becoming more and more vague recently, which corresponds with insights from Mortera-Martinez (2019) and Schick (2020). Malicious actors that attempt to undermine democracy through disinformation can act freely online, while disguised under the banner of free speech. Consequently, both Facebook and Twitter are going through a philosophical shift in their approach towards disinformation. Their initial *laissez-faire* attitude from the early to mid-2010's is making place for a more strict policy approach since 2018, moving the platforms into a direction that some critics such as the EFF call “private censorship”. Yet in fact these private companies are free to determine the rules on their platforms (Gillespie, 2018). Interestingly, this connects to the previous paragraph on governmental institutions. For the most part, governments have expressed their approval of these private social media platforms taking a more proactive role in countering disinformation by stricter content moderation. Yet the government's outside rhetoric condemns any type of censorship – because this is incomprehensible with its liberal values – and mainly emphasizes the earlier-mentioned prognostic frames focused media literacy. This two-faced argument seems contradictory at least to some degree.

IFO's take a firmer stance on freedom of expression. They could be characterized as free speech absolutists, especially EFF and Article19. By their strong framing against government regulations on content and social media content moderation strategies, these organizations see a clear trade-off between “current practices to counter disinformation” and “preserving freedom of expression” right now. The organizations make clear that it does not have to be that way. Although their counter-framing is strong, they also show willingness to participate in establishing the right kind of legislation in both the EU and the US.

## 8. Conclusion

This study examined the way in which the problem of disinformation on social media and its solutions is framed by three important actors in the digital landscape: governmental institutions, social media platforms and Internet Freedom Organizations. Together, these actors are engaged in a powerful interplay of different perspectives and interests. All three recognize that society faces a major challenge, but every actor has its own perspectives and nuances. Fortunately, governmental institutions as BZK and the EC are aware of the novel challenges and problems that the Digital Age can bring to democratic societies and show their willingness to come up with fitting solutions that also protect the right to freedom of expression. However, it can be questioned if the measures that governments bring to the table are sufficient to counter the persistent and ever-evolving problem of disinformation, or if it is fair to shift the responsibility to solve this problem into the hands of individual citizens and private companies. Furthermore, social media platforms are experimenting with their own measures, and also show sincere willingness cooperate with governments to establish fair legislation. On that note, IFO's argue that some of the measures by governments and social media platforms are either going too far or are targeting the wrong aspects of the problem. These organizations thereby function as an important addition to the policy debate around disinformation on social media.

What this study confirms at minimum is that there seems to be no clear-cut solution yet. On the one hand, society will probably see more disinformation campaigns threatening democratic processes and institutions around the world. In the words of Bart Groothuis: "it is going to be a lot worse before it gets better". On the other hand, society may see more rigid actions against disinformation by social media platforms that unavoidably generate much critique and controversy with regards to freedom of expression. Time and time again, the most difficult discussion concerns this balance between preserving freedom of expression on the one hand, while fixing the problem of disinformation on social media on the other. The relevant actors all acknowledge that this is a completely new discussion, hence it requires much debate and deliberation to figure out the right answer in order to strike the right balance between these crucial ambitions.

However, in the long run there seems to be enough room for optimism. Compared to 2016, in 2021 there is certainly a much greater understanding about the problem of online disinformation and its effects on democratic societies. Additionally, the prognostic framings of especially governments and social media platforms suggest that the impact of this problem will once be greatly diminished. Among all actors there is a strong emphasis on cooperation on all levels, and there is confidence that this problem will be solved if everyone works together. Yet as of right now, behind these "can-do" prognostic framings, a recurring theme seems to be uncertainty nonetheless. Uncertainty about how the problem of online disinformation is going to evolve. Uncertainty about which counter-measures or regulations would be the most effective and fair. Uncertainty about how freedom of expression fits into this picture. IFO's are keen to point out these uncertainties. Frankly the academic world is also rather uncertain about a way forward. Still, it should be granted that this problem is relatively novel in today's digital form. A few years down the road, solutions will have been solidified more concretely. Particularly the EC is working on a sustainable solution to this problem, in the form of two comprehensive pieces of legislation: The Digital Markets Act and Digital Services Act. These proposed legislations attempt to merge all perspectives together – including those of IFO's – into an inclusive and fair set of regulations that will establish clear rules and norms for all involved actors. This is, according to all three actors, a very positive prospect.

In the words of Charles Kettering, a reputable industrial innovator: "a problem not fully understood is unsolvable. A problem that is fully understood is half-solved". We might be almost halfway there.

## References

### Academic sources

- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236.
- Awan, I. (2017). Cyber-Extremism: Isis and the Power of Social Media. *Society*, 54(2), 138–149.
- Baele, S. J., Brace, L., & Coan, T. G. (2020). Uncovering the Far-Right Online Ecosystem: An Analytical Framework and Research Agenda. *Studies in Conflict and Terrorism*, 1–21.
- Barberá, P., & Zeitzoff, T. (2018). The new public address system: Why do world leaders adopt social media? *International Studies Quarterly*, 62(1), 121–130.
- Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*, 37(1), 38–54.
- Benford, R. D., & Snow, D. A. (2000). Framing processes and social movements: an overview and assessment. *Annual Review of Sociology*, 26(1), 611–639.
- Benigni, M. C., Joseph, K., & Carley, K. M. (2017). Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter. *PLoS ONE*, 12(12), 1–23.
- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139.
- Bilazarian, T. (2020). Countering Violent Extremist Narratives Online: Lessons From Offline Countering Violent Extremism. *Policy and Internet*, 12(1), 46–65.
- Bjola, C. (2018). The ethics of countering digital propaganda. *Ethics and International Affairs*, 32(3), 305–315.
- Bode, L., & Vraga, E. K. (2015). In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. *Journal of Communication*, 65(4), 619–638.
- Boeije, H. (2010). *Analysis in Qualitative Research*. Thousand Oaks: SAGE Publications.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2),
- Brown, É. (2018). Propaganda, Misinformation, and the Epistemic Value of Democracy. *Critical Review*, 30(3–4), 194–218.
- Brunn, S. D. (1998). A treaty of silicon for the Treaty of Westphalia? New territorial dimensions of modern statehood. *Geopolitics*, 3(1), 106–131.
- Castells, M. (1996). *The Rise of the Network Society*. Oxford: Blackwell.
- Chan, M. pui S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*, 28(11), 1531–1546.
- Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media. *Political Behavior*, 42(4), 1073–1095.
- Conway, M. (2017). Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research. *Studies in Conflict and Terrorism*, 40(1), 77–98.
- Demmers, J. (2016). *Theories of Violent Conflicts: An Introduction* (2nd ed.). London: Routledge.

- Ecker, U. K. H., & Ang, L. C. (2019). Political Attitudes and the Processing of Misinformation Corrections. *Political Psychology, 40*(2), 241–260.
- Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday, 22*(8), 1–33.
- Fridkin, K., Kenney, P. J., & Wintersieck, A. (2015). Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising. *Political Communication, 32*(1), 127–151.
- Ganesh, B., & Bright, J. (2020). Countering Extremists on Social Media: Challenges for Strategic Communication and Content Moderation. *Policy and Internet, 12*(1), 6–19.
- Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven: Yale University Press.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research*. London: Aldine.
- Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the Filter Bubble? *Digital Journalism, 6*(3), 330–343.
- Hameleers, M., Powell, T. E., Van Der Meer, T. G. L. A., & Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication, 37*(2), 281–301.
- Herrera, G. (2008). Cyberspace and Sovereignty: Thoughts on Physical Space and Digital Space. In M. Dunn Caverty, V. Mauer, & S. F. Krishna-Hensel (Eds.), *Power and Security in the Information Age: Investigating the Role of the State in Cyberspace* (pp. 67–94).
- Hodge, E., & Hallgrimsdottir, H. (2020). Networks of Hate: The Alt-right, “Troll Culture”, and the Cultural Geography of Social Movement Spaces Online. *Journal of Borderlands Studies, 35*(4), 563–580.
- Howard, P., Duffy, A., Freelon, D., Hussain, M., Mari, W., & Mazaid, M. (2011). Opening closed regimes: what was the role of social media during the Arab Spring? *Project on Information Technology & Political Islam, 1*–30.
- Howard, P. N. (2020). *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. New Haven: Yale University Press.
- Hwang, T., Pearce, I., & Nanis, M. (2012). Socialbots: Voices from the Fronts. *Interactions, 19*(2), 38–45.
- Jerit, J., & Zhao, Y. (2020). Political Misinformation. *Annual Review of Political Science, 23*, 77–94.
- Jost, J. T. (2017). Ideological Asymmetries and the Essence of Political Psychology. *Political Psychology, 38*(2), 167–208.
- Keller, F. B., Schoch, D., Stier, S., & Yang, J. H. (2020). Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign. *Political Communication, 37*(2), 256–280.
- Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the Currency of Democratic Citizenship. *The Journal of Politics, 62*(3), 790–816.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498.
- Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., Mattsson, C., Benkler, Y., Berinsky, A., Boaden, H., Brown, K., Pennycook, G., Robertson, L., Rothschild, D., Sloman, S., Sunstein, C., Thorson, E., & Watts, D. (2017). Combating Fake News: An Agenda for Research and Action. *Filippo Menczer, May*.
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest, Supplement, 13*(3), 106–131.

- Lewandowsky, S., & Oberauer, K. (2016). Motivated Rejection of Science. *Current Directions in Psychological Science*, 25(4), 217–222.
- Mejias, U. A., & Vokuev, N. E. (2017). Disinformation and the media: the case of Russia and Ukraine. *Media, Culture and Society*, 39(7), 1027–1042.
- Mena, P. (2020). Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook. *Policy and Internet*, 12(2), 165–183.
- Mölder, H., & Sazonov, V. (2019). The Impact of Russian Anti-Western Conspiracy Theories on the Status-Related Conflict in Ukraine: The Case of Flight MH17. *Baltic Journal of European Studies*, 9(3), 95–115.
- Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information Communication and Society*, 21(7), 959–977.
- Morozov, E. (2011). *The Net Delusion: The Dark Side of Internet Freedom*. New York: Perseus Books.
- Mortera-Martinez, C. (2019). What is Europe doing to fight disinformation? *CER Bulletin*, 123.
- Nouri, L., & Lorenzo-Dus, N. (2019). Investigating reclaim Australia and Britain first's use of social media: Developing a new model of imagined political communities online. *Journal for Deradicalization*, 18, 1–37.
- Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, 32(2), 303–330.
- Price, J. H., & Murnan, J. (2004). Research Limitations and the Necessity of Reporting Them. *American Journal of Health Education*, 35(2), 66–67.
- Ragin, C. C., & Amoroso, L. M. (2019). *Constructing Social Research: The Unity and Diversity of Method*. Thousand Oaks: SAGE Publications.
- Reviglio, U. (2019). Serendipity as an emerging design principle of the infosphere: challenges and opportunities. *Ethics and Information Technology*, 21(2), 151–166.
- Schick, N. (2020). *Deep Fakes and the Infocalypse*. London: Monoray.
- Schiffrin, A. (2017). Disinformation and Democracy: The internet transformed protest but did not improve democracy. *Journal of International Affairs*, 71(1), 117–126.
- Shirky, C. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign Affairs*, 90(1), 28–41.
- Somainsi, L., & Pollicino, O. (2020). Online disinformation and freedom of expression in the electoral context: the European and Italian responses. *Sandrine Baume, Véronique Boillet, Vincent Martenet (Eds), Misinformation in Referenda, Forthcoming. Available at SSRN: <https://ssrn.com/abstract=3552680>*.
- Stewart, L. G., Arif, A., & Starbird, K. (2018). Examining trolls and polarization with a retweet network. *Proceedings of WSDM Workshop on Misinformation and Misbehavior Mining on the Web*.
- Swire-Thompson, B., Ecker, U. K. H., Lewandowsky, S., & Berinsky, A. J. (2020). They Might Be a Liar But They're My Liar: Source Evaluation and the Prevalence of Misinformation. *Political Psychology*, 41(1), 21–34.
- Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3), 755–769.
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153.

- Thorson, E. (2016). Belief Echoes: The Persistent Effects of Corrected Misinformation. *Political Communication*, 33(3), 460–480. <https://doi.org/10.1080/10584609.2015.1102187>
- Thorson, K., & Wells, C. (2016). Curated Flows: A Framework for Mapping Media Exposure in the Digital Age. *Communication Theory*, 26(3), 309–328.
- Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stepińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheaffer, T., & Stanyer, J. (2017). Political communication in a high-choice media environment: a challenge for democracy? *Annals of the International Communication Association*, 41(1), 3–27.
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, 37(3), 350–375.
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423–441.
- Weeks, B. E., & Gil de Zúñiga, H. (2021). What's Next? Six Observations for the Future of Political Misinformation Research. *American Behavioral Scientist*, 65(2), 277–289.
- Zeitoff, T. (2017). How Social Media Is Changing Conflict. *Journal of Conflict Resolution*, 61(9), 1970–1991.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

## Empirical sources

- Alexander, J. (2020, February 3). *YouTube lays out disinformation policy ahead of primary vote*. The Verge. Retrieved 5 June 2021, from <https://www.theverge.com/2020/2/3/21120707/youtube-disinformation-election-2020-manipulated-video-channel-policy>
- Article19. (2020, November 19). *What we do*. Retrieved 17 March 2021, from <https://www.article19.org/what-we-do/>
- Bits of Freedom. (n.d.). *Missie*. Retrieved 16 March 2021, from <https://www.bitsoffreedom.nl/missie/>
- BZK [Ministerie van Binnenlandse Zaken en Koninkrijksrelaties]. (2021, February 9). *Kamerbrief voortgang maatregelen tegen desinformatie richting Tweede Kamerverkiezing*. Rijksoverheid. Retrieved 21 February 2021, from <https://www.rijksoverheid.nl/documenten/kamerstukken/2021/02/09/kamerbrief-voortgang-maatregelen-tegen-desinformatie-richting-tweede-kamerverkiezing>
- Chen, A. (2015, June 2). *The Agency*. The New York Times. Retrieved 17 February 2021, from <https://www.nytimes.com/2015/06/07/magazine/the-agency.html>
- Conger, K., & Isaac, M. (2021, January 16). *Inside Twitter's Decision to Cut Off Donald J. Trump*. The New York Times. Retrieved 21 February 2021, from <https://www.nytimes.com/2021/01/16/technology/twitter-donald-trump-jack-dorsey.html>
- Electronic Frontier Foundation. (2021, March 9). *About EFF*. Retrieved 16 March 2021, from <https://www.eff.org/about>
- European Commission. (2018, March). *A multi-dimensional approach to disinformation: Final report of the High Level Expert Group on Fake News and Online Disinformation*. Shaping Europe's Digital Future. Retrieved 3 March 2021, from <https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation>

- European Commission. (2020a, September 10). *Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement*. Shaping Europe's Digital Future. Retrieved 4 March 2021, from <https://ec.europa.eu/digital-single-market/en/news/assessment-code-practice-disinformation-achievements-and-areas-further-improvement>
- European Commission. (2020b, July 7). *The European Digital Media Observatory*. Shaping Europe's Digital Future. Retrieved 11 March 2021, from <https://ec.europa.eu/digital-single-market/en/european-digital-media-observatory>
- Harris, T. (2021, February 10). *Tristan Harris in conversation with Professor Victoria Nash*. Oxford Internet Institute. Retrieved 11 February 2021, from <https://www.oii.ox.ac.uk/videos/oii-wednesday-webinar-week-4-tristan-harris-in-conversation-with-professor-victoria-nash/>
- Howell, L. (2013). *Digital Wildfires in a Hyperconnected World*. World Economic Forum. Retrieved 17 February 2021, from <http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/>
- I&O Research. (2017, December 23). *Desinformatie leidt tot verwarring bij nieuwsconsument*. Retrieved 15 February 2021, from <https://www.ioresearch.nl/actueel/desinformatie-leidt-tot-verwarring-bij-nieuwsconsument/>
- Instituut voor Informatierecht. (2019, October 18). *De verspreiding van desinformatie via internetdiensten en de regulering van politieke advertenties: Tussenrapportage*. Rijksoverheid. Retrieved 28 February 2021, from <https://www.rijksoverheid.nl/documenten/rapporten/2019/10/18/de-verspreiding-van-desinformatie-via-internetdiensten-en-de-regulering-van-politieke-advertenties>
- Jain, R. R. (2019, April 9). *In India's Election, Voters Feed on False Information*. Retrieved 16 February 2021, from <https://www.bloomberg.com/News/Articles/2019-04-09/in-India-s-Election-Voters-Feed-on-False-Information>.
- Kamerstuk 26.643, nr. 508. (2017, December 18). Retrieved 2 March 2021, from <https://zoek.officielebekendmakingen.nl/kst-26643-508.html>
- Kamerstuk 30.821, nr. 91. (2019, October 28). Retrieved 2 March 2021, from <https://zoek.officielebekendmakingen.nl/kst-30821-91.html>
- Kelly, J. (2021, February 18). *How 'fact-checking' can be used as censorship*. Financial Times. Retrieved 24 March, from <https://www.ft.com/content/69e43380-dd6d-4240-b5e1-47fc1f2f0bdc>
- Kelly, M. (2021, February 24). *TikTok removed more than 300,000 videos for election misinformation*. The Verge. Retrieved 5 June 2021, from <https://www.theverge.com/2021/2/24/22298024/tiktok-election-misinformation-disinformation-transparency-report>
- Lippman, D. (2016, March 13). *Trump's Week of Errors, Exaggerations and Flat-out Falsehoods*. Politico Magazine. Retrieved 16 February 2021, from <https://www.politico.com/magazine/story/2016/03/trump-fact-check-errors-exaggerations-falsehoods-213730/>
- NOS. (2020, November 10). *Indonesisch botleger voert beïnvloedingscampagne in het Nederlands*. Retrieved 16 February 2021, from <https://nos.nl/nieuwsuur/artikel/2355987-indonesisch-botleger-voert-beïnvloedingscampagne-in-het-nederlands.html>

- Paul, K. (2020, October 16). *YouTube announces plans to ban content related to QAnon*. The Guardian. Retrieved 12 March 2021, from <https://www.theguardian.com/technology/2020/oct/15/youtube-ban-qanon-content-technology>
- Roose, K., Isaac, M., & Frenkel, S. (2021, January 7). *Facebook Struggles to Balance Civility and Growth*. The New York Times. Retrieved 12 March 2021, from <https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html>
- Roth, Y., & Pickles, N. (2020, May 11). *Updating our approach to misleading information*. Twitter Blog. Retrieved 12 March 2021, from [https://blog.twitter.com/en\\_us/topics/product/2020/updating-our-approach-to-misleading-information.html](https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html)
- Sheth, S., & Bertrand, N. (2017, July 16). *Trump Jr.'s meeting with a Russian lawyer shed new light on the extent of Russia's election interference*. Business Insider. Retrieved 12 April 2021, from <https://www.businessinsider.nl/evidence-russia-meddled-in-us-election-2017-6?international=true&r=US>
- The Economist. (2017, November 23). *Russian Twitter trolls meddled in the Brexit vote. Did they swing it?* Retrieved 15 February 2021, from <https://www.economist.com/britain/2017/11/23/russian-twitter-trolls-meddled-in-the-brexit-vote-did-they-swing-it>
- Wardle, C. (2018, December 10). *Fake news. It's complicated*. Medium. Retrieved 16 February 2021, from <https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79>

## Appendix I – Data overview

### List of conducted interviews

Date	Organization	Respondent
20-04-2021	European Parliament	Bart Groothuis, member of the liberal Renew Europe party, the Committee on Foreign Affairs and the Special Committee on Foreign Interference and Disinformation.
21-04-2021	BZK	Anonymous, policy director
30-04-2021	Facebook	Anonymous, public policy
31-05-2021	Bits of Freedom	Rejo Zenger, senior policy advisor

### List of documents

Date	Organization	Name document
18-12-2017	BZK	Kamerstuk 26.643, nr. 508 - Informatie- en communicatietechnologie
19-12-2017	Tweede Kamer	Aanhangsel vergaderjaar 2017-2018, nr. 733 - Kamervragen desinformatie
15-01-2018	EZK	Kamerstuk 33.009, nr. 48 - Innovatiebeleid
06-03-2018	AIVD	Jaarverslag 2017
06-03-2018	BZK	Kamerstuk 26.643, nr. 291 - Informatie- en communicatietechnologie
16-03-2018	BZK and J&V	Kamerstuk 30.821, nr. 42 - Nationale veiligheid
13-12-2018	BZK	Kamerstuk 30.821, nr. 51 - Nationale veiligheid
02-04-2019	AIVD	Jaarverslag 2018
18-04-2019	NCTV	Χίματα, een duiding van het fenomeen 'hybride dreiging'
07-06-2019	NCTV	Geïntegreerde risico analyse Nationale Veiligheid
18-10-2019	BZK	Kamerstuk 30.821, nr. 91 - Beleidsinzet bescherming democratie tegen desinformatie
12-10-2020	BZK	Kamerbrief - tegengaan digitale inmenging Tweede Kamer verkiezingen 2021
29-10-2020	BZK	Kamerbrief - beantwoording nadere vragen over de bestrijding van online desinformatie
09-02-2021	BZK	Kamerbrief - ontwikkelingen maatregelen om desinformatie richting de Tweede Kamer verkiezingen tegen te gaan

30-04-2018	European Commission	Report of the independent High Level Expert Group on fake news and online disinformation
26-09-2018	European Commission	EU Code of Practice on Disinformation
05-12-2018	European Commission	European Action Plan against Disinformation
14-06-2019	European Commission	Verslag over de uitvoering van het actieplan tegen desinformatie
08-03-2021	European Commission	Assessment of the Code of Practice on Disinformation
27-04-2017	Facebook	Information Operations and Facebook
17-05-2018	Facebook	Announcing New Election Partnership With the Atlantic Council
09-08-2018	Facebook	Hard Questions: Where Do We Draw the Line on Free Expression
19-10-2018	Facebook	Q&A on Upcoming US and Brazil Elections
14-11-2018	Facebook	How Does Facebook Investigate Cyber Threats and Information Operations?
36-12-2018	Facebook	Coordinated Inauthentic Behavior Explained
30-03-2019	Facebook	Four Ideas to Regulate the Internet
21-10-2019	Facebook	Helping to Protect the 2020 US Elections
16-04-2020	Facebook	An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19
08-10-2020	Facebook	Recommended Principles for Regulation or Legislation to Combat Influence Operations
08-10-2020	Facebook	Removing Coordinated Inauthentic Behavior
21-10-2020	Facebook	Inauthentic Behavior Report: Update on Our Work Against Deceptive Behavior
31-03-2021	Facebook	You and the Algorithm: It Takes Two to Tango
26-05-2021	Facebook	Taking Action Against People Who Repeatedly Share Misinformation
26-05-2021	Facebook	Threat Report: The State of Influence Operations 2017-2020
14-06-2017	Twitter	Our approach to bots and misinformation
28-09-2017	Twitter	Update: Russian interference in the 2016 US presidential election
31-10-2017	Twitter	US Senate Committee on the Judiciary: Opening Remarks
17-10-2018	Twitter	Enabling further research of information operations on Twitter

01-11-2018	Twitter	Five days until #ElectionDay 2018
19-02-2019	Twitter	The world's window into the European election conversation
20-03-2019	Twitter	Celebrating #EUMediaLiteracyWeek
13-06-2019	Twitter	Information operations on Twitter: principles, process and disclosure
24-10-2019	Twitter	Twitter builds partnership with UNESCO on media and information literacy
04-03-2020	Twitter	Stepping up our work to protect the public conversation around COVID-19
18-05-2020	Twitter	Bot or Not? The facts about platform manipulation on Twitter
10-10-2020	Twitter	Expanding our policies to further protect the civic conversation
12-10-2020	Twitter	An update on our work around the 2020 US Elections
12-01-2021	Twitter	Coronavirus: Staying safe and informed on Twitter
09-02-2021	Twitter	#SaferInternetDay 2021: Together for a better Internet
01-03-2021	Twitter	Updates to our work on COVID-19 vaccine misinformation
03-03-2017	Article19	Free speech mandates issue Joint Declaration addressing freedom of expression and "fake news"
18-01-2018	Article19	Free speech concerns amid the "fake news" fad
09-10-2018	Article19	EU: New Code of Practice on Disinformation fails to provide clear commitments, or protect fundamental rights
03-2020	Article19	Viral Lies: Misinformation and the Coronavirus
05-2020	Article19	Ensuring the Public's Right to Know in the COVID-19 Pandemic
11-06-2020	Article19	Europe: EU Communication on tackling coronavirus disinformation
19-10-2020	Article19	Inside Expression: September 2020 – We need to use our Right to Know, right now
12-01-2021	Article19	US: A Capitol riot and Big Tech takes a stand: but is it the one we want?
12-02-2021	Article19	Submission to UN Special Rapporteur on freedom of expression and 'disinformation'
30-04-2021	Article19	EU: Disinformation guidance must protect freedom of expression
10-10-2017	Electronic Frontier Foundation	With Facebook, Twitter in the Crosshairs of Investigators Probing Russian Interference, Let's Consider The Risks of Applying Election Ad Rules to the Online World

11-04-2018	Electronic Frontier Foundation	Despite What Zuckerberg's Testimony May Imply, AI Cannot Save Us
27-06-2018	Electronic Frontier Foundation	UN Report Sets Forth Strong Recommendations for Companies to Protect Free Expression
02-28-2018	Electronic Frontier Foundation	Fake News and Elections in Brazil: Several Initiatives, No Easy Answer
28-02-2019	Electronic Frontier Foundation	EFF to the inter-American System: If You Want to Tackle "Fake News," Consider Free Expression First
29-04-2019	Electronic Frontier Foundation	Content Moderation is Broken. Let Us Count the Ways
01-05-2019	Electronic Frontier Foundation	Censorship Can't Be The Only Answer to Disinformation Online
04-05-2020	Electronic Frontier Foundation	New Laws Banning False News Threaten the Free Flow of Information Worldwide
24-12-2020	Electronic Frontier Foundation	How COVID Changed Content Moderation: Year in Review 2020
07-04-2021	Electronic Frontier Foundation	India's Strict Rules For Online Intermediaries Undermine Freedom of Expression
19-06-2019	Bits of Freedom	Fix the system, not the symptoms
21-06-2019	Bits of Freedom	Brief aan BZK: Dominantie van platformen in communicatielandschap
07-11-2019	Bits of Freedom	Wat we nodig hebben is regulering van de markt, niet van onze communicatie
27-11-2020	Bits of Freedom	Verkiezingsmanipulatie los je niet op met vriendelijk vragen
29-03-2021	Bits of Freedom	Healthy Public Debate Online