Beyond Regulation: Approaching the challenges of the new media environment

Rafael Goldzweig, Marie Wachinger, Daniela Stockmann, and Andrea Römmele
Dahrendorf Forum IV
Working Paper No. 06
11 December 2018
About the Authors

Rafael Schmuziger Goldzweig is a Research Associate at the Dahrendorf Forum, contributing to the working group on Societal Change, Politics, and the Public Sphere. His research interests lie at the intersection of technology and politics, focusing on new media and elections. He was a Google Policy Fellow and a DAAD Helmut-Schmidt-Programme Scholar. Prior to joining the Dahrendorf Forum, he worked as a political consultant. Rafael holds a BA in International Affairs and a Master of Public Policy (MPP) from the Hertie School of Governance. Author email: schmuziger@hertie-school.org

Marie Wachinger has been a Research Associate at the Dahrendorf Forum since September 2017 after two years as Research Associate at the Hertie School of Governance. She is currently working on a PhD in Political Science at the Free University Berlin. Her research focuses on solidarity, statehood, and the EU. She previously worked at a newspaper and at a political think tank. Marie holds a BA from Amsterdam University College and an MSc in Political Theory from the London School of Economics and Political Science. Author email: wachinger@hertie-school.org

Daniela Stockmann is Professor of Digital Politics and Media at the Hertie School and a member of the Dahrendorf Forum working group on Societal Change, Politics, and the Public Sphere. She holds degrees from the University of Rochester, the School of Oriental and African Studies at the University of London, and a PhD from the University of Michigan, Ann Arbor (2007). She was previously an Associate Professor of Political Science at Leiden University. Her book, Media Commercialization and Authoritarian Rule in China received the 2015 Goldsmith Book Prize. Author email: stockmann@hertie-school.org

Andrea Römmele is a Professor of Communication in Politics and Civil Society at the Hertie School and Co-Chair of the Dahrendorf Forum working group on Societal Change, Politics, and the Public Sphere. She held the Fulbright Distinguished Chair in Modern German Studies at the University of California, Santa Barbara, and has been a visiting fellow at Johns Hopkins University and the Australian National University in Canberra. She is co-founder and Editor-in-Chief of the Journal for Political Consulting and Policy Advice. She holds a master’s degree from the University of California, Berkeley, a PhD from Heidelberg University, and a habilitation from the Free University of Berlin. Author email: roemmele@hertie-school.org

Contact: Hallie Detrick, h.detrick@lse.ac.uk
Abstract

The spread of false information and hate speech has increased with the rise of social media. This paper critically examines this phenomenon and the reactions of governments and major corporations in Europe. Policymakers have turned towards national regulation as a means to manage false information and hate speech. This article looks into the legislative frameworks on the issue in Germany, France, the UK, the Czech Republic, and Italy and compares them. In response to such regulatory pressure, tech companies have been changing aspects of their platforms to deal with this trend, for example through content moderation. We propose tentative alternatives to this current approach towards reinforcing boundaries for freedom of expression.

Keywords

regulation; social media; tech companies; false information; hate speech; elections

Acknowledgements

We would like to express our gratitude to our colleagues at the Dahrendorf Forum for the feedback given.

The Dahrendorf Forum

The Dahrendorf Forum is a joint initiative by the Hertie School of Governance, the London School of Economics and Political Science, and Stiftung Mercator that recognises expert knowledge and public debate can each benefit from mutual exposition.

The Dahrendorf Team generates and disseminates social science research that is both policy relevant and of the highest standard. The researchers concentrate on impacting high-level policymakers and practitioners close to the centres of political action and decision-making.
1. Introduction

The spread of false information and hate speech has increased with the rise of social media. The phenomenon of false information dates back to the pre-digital age, but there is a fundamental difference of scale when we compare a pamphlet or newspaper that circulates locally with an online article that can go viral worldwide with a simple click of a button. Social media allows ordinary citizens to become ‘prosumers’—producers and consumers of news. Such information can go viral at an extraordinary speed and funnel to the centre of public discourse, transcending national boundaries. In the early times of the internet, this phenomenon was regarded as bearing tremendous opportunity for strengthening democracy and building trust across social groups that are often difficult to reach. More recently, however, observers started to realise that the same traits that made new media a force for more-perfect information can also catalyse the spread of false information, creating distrust, contributing to polarisation, and thus challenging democratic values and destabilising society and entire political systems.

Especially during elections, false information and misleading content have emerged as a challenge to policymakers, tech companies, and citizens. Elections are strategic periods for democratic societies. They are times when politicians have to stand by their promises and citizens have the opportunity to reflect, evaluate, and voice their opinions about the past and future. In recent years, election periods in the United States, Germany, and other democracies around the world have experienced campaigns during which false information and anti-democratic messages originated and quickly spread online through various channels and following unprecedented dynamics. Widespread perceptions about what constitutes false information are often coupled with ‘uncivil’ interactions between people who identify themselves a belonging to opposing, increasingly polarised left- and right-wing groups incapable of interacting in a calm, respectful, and tolerant manner that constitutes the foundation of democratic discourse.

How do we make sense of the role of social media in this phenomenon, and what solutions can we offer to bring back evidence and truth to the centre of the debates happening in the public sphere—whether online or offline? How can these solutions bring back trust in information spread by social media?

This paper explores the problem of false information online and analyses reactions by public and private stakeholders to manage, control, and channel information, thus imposing limits on freedom of expression for the sake of stabilising and preserving democracy. While governments have attempted to solve this problem with laws and regulations that hold tech companies accountable for the content they facilitate, companies have looked for solutions in the design of their products, their guidelines for buying advertising, and the calibration of algorithms to decrease the reach of potentially harmful content.
Our intention is to draw attention to current policy approaches to the problem of false information and hate speech and to spark a debate about fundamental questions that we believe need to be discussed more broadly as part of public discourse: Are current approaches the best way to tackle these challenges? Who should have the authority to decide the boundaries of public discourse—governments, companies, or citizens? And, in a broader sense, does a healthy democracy require limits on freedom for expression in cyberspace?

In order to explore this field, we start from a definition of the problem, going beyond the one-size-fits-all distinction of 'fake news' to understand the specifics of the information disorder, focusing our attention on the interaction between government and tech companies. We also explore the current state of the media landscape to understand how and where online and offline meet, and what the main features are that allowed this phenomenon to gain strength in recent years. We then look at legislative measures taken so far. We outline and compare the legal approaches of five European countries as well as the EU on the matter in order to describe the status quo. Subsequently, we present self-regulation initiatives, trying to evaluate their impact so far and where they still fall short.

2. Forget Fake News. We are dealing with a bigger problem.

'Fake news' is a label that contains different aspects of what is broadly known as 'information disorder'. The term fake news is used to discredit information with which one does not agree, and is often used to define a broad spectrum of false or misleading information, without considering the nuances behind this phenomenon. The concept describes a variety of content that range from satirical news sites to false pieces of information aimed at creating political diversion. The vagueness and misuse of this label makes the term fake news academically useless (Marwick, 2018).

Hence, we base our conceptualisation of information disorder on the framework of the Council of Europe (Wardle & Derakhshan, 2017) to define the problem and explore its potential solutions. The framework divides the source of the problem into mis-, dis- and mal-information: Mis-information refers to false information shared without the intention of causing harm. Dis-information refers to false information shared with the intention to cause harm. Mal-information deals with true information shared with the intent of causing harm via leaks and through harassment.

Figure 1 below summarises this framework. While mis-information (left), which can also be described by the catchy term “bullshit”, refers to false information spread without the intention of harm, dis-information (middle) refers to false or manipulated content that is used deliberately with the intention to cause harm. Mal-information (on the right) refers to genuine information that is used with the purpose of harming an individual or organization, such as through leaks or hate speech. This working paper will focus on the attempts to cope with false information (mis- and dis-information).
Changing the incentives to avoid the spread of false information alone is not enough to make people believe in evidence and facts. It is important to understand why people engage with fake content. Here, we are dealing with a question that goes beyond the lack of digital literacy that allows one to assess whether content is reliable or not. Understanding the reasons why people engage with fake content—and even knowingly share it—is important to make sense of the information disorder.

Communication plays two roles for the ones who exercise it: the role of transmitting information and a ritualistic role (Carey, 1989). In the social media era, ‘transmission view’ is what we take for granted when we engage with news pieces and articles—sharing means that we want to pass information to others and disseminate knowledge. The ritualistic role, which is more important to the question we aim to answer, deals with the fact that people share news to reinforce their values, beliefs, and world views. From this perspective, the purpose of communicating is to build an image based on the values and questions one supports. The ritualistic function of communication is deeply connected to the inherent relation it has to emotions.

Jürgen Habermas argues that any constative act of communication usually contains an expressional dimension: if an agent promises something, it is connected to a feeling of responsibility; if an agent apologises, the apology stems from a feeling of regret, and so on (Habermas, 1995). The strong interdependence of affection and transmission of information combined with the ritualistic intention make communication very complex. The three are virtually inseparable from each other. The information aspect may be secondary and perhaps even not important under this perspective, which explains why people sometimes do not care about whether the facts they distribute are reliable or not. This view is important to understand why partisan media outlets flourish in groups with a specific set of beliefs. Truth becomes secondary to values.
The information disorder and communicative practices cannot be analysed separately from the social, political, and technological circumstances in which they find themselves in. Hence, we need to understand how false or biased online content interacts with the changing media landscape and why it finds its way into social groups.

3. The Bigger Picture

3.1 Media Landscape: Fragmentation and media consumption

With recent examples of algorithmic platforms amplifying fake election coverage (Caplan et al., 2018), the viral aspect of how content spreads on the internet has led to uncertainty and a sense of helplessness. In order to understand this phenomenon and come up with solutions for creating a healthier public sphere, it is important to take a closer look at the environment in which this misleading information spreads. In the public sphere, the media landscape is arguably the sector that has most changed in light of increasing digitalization.

Andrew Chadwick (2007 and 2013) coined the term “hybrid media system” in order to describe these profound alterations. Hybridisation of the media system describes the changed relations within and between media. Older, established media outlets have entered a very complex interdependence and competition with “new” digital media. In some instances their dynamics complement each other catering different needs and habits of consumers, while in others traditional media struggle to adapt to the new environment. Ultimately, it is a struggle for power: technologies, consumers, and communicative practices change, and with them the influences of different actors in the media landscape.

While digital tools allow for more immediate and direct interaction and are increasingly popular, they are also prone to the challenges of the information disorder. Traditional media like newspapers, radio, and television, in turn, are frequently accused of representing a political “establishment” and struggle to uphold their influence in society. They also adapt to the new circumstances, creating a third category of “renewed” media that attempts to accommodate the need to increase citizen participation (Hoskins & O’Loughlin, 2010). The boundaries between what can be considered mainstream become increasingly blurred. The term “hybrid” is apt because it allows for the intersection of all types of media. On the one hand they can be said to be increasingly fragmented, but on the other hand they create new dynamics and new patterns of consumption. What are these dynamics and how are they connected to the problem of mis- and dis-information?

In this notably fragmented and hybrid media environment, it is likely that mis- and dis-information are prone to different dynamics in their spread. One of the most prominent arguments in this regard concerns the risks associated with echo chambers and filter bubbles. The assumption is that
Guess et al. (2018) found that in the 2016 US presidential election campaign, “fake news website production and consumption was overwhelmingly pro-Trump in its orientation”. They also found that there was some problem of selective exposure: “A narrow subset of Americans with the most conservative information diets was disproportionately likely to visit fake news websites. These results contribute to the ongoing debate about the problem of ‘filter bubbles’ by showing that the ‘echo chamber; is deep [...] but narrow (the group consuming so much fake news represents only 10% of the public)” (ibid). This means the segments of the population that are prone to consuming dis-information are likely to be affected by a lack of exposure to other news sources. However, this part of the population is relatively small. The authors found Facebook to be a main contributor to this phenomenon and also concluded that fact-checking was ineffective to counter the problem, since those in the “filter bubble” would not read or trust them.

A BuzzFeed News analysis (Silverman, 2016) found that top fake election news stories generated more total engagement on Facebook than top election stories from 19 major news outlets combined in the same time span. Specifically, they were able to show that the “top-performing fake election news stories on Facebook generated more engagement than the top stories from major news outlets such as the New York Times, Washington Post, Huffington Post, NBC News, and others […]. During [the] critical months of the campaign, 20 top-performing false election stories from hoax sites and hyperpartisan blogs generated 8,711,000 shares, reactions, and comments on Facebook. Within the same time period, the 20 best-performing election stories from 19 major news websites generated a total of 7,367,000 shares, reactions, and comments on Facebook”. This means that false content spread much faster on the internet than traditional media content.

Flaxman et. al. (2016) conducted a more general study, analysing the two opposing theories that exposure to new media would either lead to more diverse perspectives and informed opinion or create filter bubbles. They did a large-scale study of online news consumption. Their findings were not clear-cut for either sides: “We showed that articles found via social media or web-search engines are indeed associated with higher ideological segregation than those an individual reads by directly visiting news sites. However, we also found, somewhat counterintuitively, that these channels are associated with greater exposure to opposing perspectives. Finally, we showed that the vast majority of online news consumption mimicked traditional offline reading habits, with individuals directly visiting the home pages of their favorite, typically mainstream, news outlets.” This may indicate some support for Guess et al.’s findings that effects are overall not so significant.
by consuming certain sources of media on social media networks, users will find themselves exposed only to similar sources and arguments because of the underlying algorithms, thereby causing opinions to move to the extremes. The lack of exposure to contradicting arguments works to confirm existing convictions and hinder balanced argumentation. Moreover, the described dynamics serve to accelerate the spread of deliberate mis-information. Empirical research about the danger of polarisation and mis-information on social media only seems to confirm the described dangers to some extent.

The empirical evidence suggests that while dis-information seems to spread faster on the internet, the observable effects of filter bubbles and echo chambers are relatively modest and limited to small segments of the population. Long-term studies will have to provide more reliable findings but for now, the phenomenon cannot be considered one that involves the majority of the population.

However, those parts of the population that do contribute to the problems attached to mis- and dis-information can be assumed to be driven in part by the phenomenon of confirmation bias. Confirmation bias is “the tendency to acquire or evaluate new information in a way that is consistent with one’s preexisting beliefs” (Allahverdyan & Galstyan, 2014). Studies have found that confirmation bias can also occur in traditional media environments, by simply eschewing those media outlets that do not represent one’s political views (Iyengar & Hahn, 2009). However, a media environment with hybrid characteristics is arguably an even better breeding ground for confirmation bias. The ritualistic function of communication can be exercised much more easily online due to the immediacy and openness of the forum for discourse. This, combined with a general tendency towards confirmation bias, may lead to further polarisation in the future, even though evidence suggests that it is not (yet) a mass phenomenon.

3.2 Societal Rifts and the Media Landscape as a Mirror of Them

A look at the consumption habits of individuals or groups should take into account the social and political circumstances in which they find themselves. Given the strong ritualistic function of media consumption, what makes people share misinformation? Why did/When did the value of distributing the truth become inferior to the need to advance their world views and opinions? A complete picture of societal circumstances can provide insight into a more systemic approach to regulation of new media and potential alternatives.

In part, the success of mis- and dis-information can be explained by increasing polarisation, which drives certain societal groups farther apart. This phenomenon is evident in certain understandings of populist discourse that promotes a juxtaposition of “us” vs. “them” (e.g. in the work of Müller, 2016). Societal polarisation, more specifically, can be attributed to increased social cleavages. Cleavages are not a new phenomenon. In the 1960s, political scientists Seymour Martin Lipset and Stein Rokkan (Lipset & Rokkan, 1967) formulated a theory on cleavages in society that form along
certain lines: between labour and capital, between the church and the state, between centre and periphery, and between urban and rural areas. Their idea was to show how the cleavages were resulted in the citizens’ identification with corresponding political parties.

In part, these juxtapositions retain their validity. However, the complexity of modern societies and their placement in a globalised context makes the cleavage theory seem too simple. The ‘societal rifts’ we observe today are sometimes less clearly defined, more globalised and intertwined. So what are today’s societal rifts based on and how do they relate to the developing media landscape?

Studies like the World Inequality Report (Alvaredo et al., 2018) continuously point toward the pressing problem of material inequality. More specifically, “the top 1% richest individuals in the world captured twice as much growth as the bottom 50% individuals since 1980. Income growth has been sluggish or even zero for individuals with incomes between the global bottom 50% and top 1% groups. This includes all North American and European lower- and middle-income groups” (p.11). While material inequalities differ on a global level and depend on regional levels of economic regulation, this finding indicates an increasing divide between the rich and poor.

However, inequalities are multidimensional, extending to access to education and healthcare and discrimination based on gender, ethnicity or other factors. To borrow Bourdieu’s classification: societal rifts run along the lines of economic, social, and cultural capital (Bourdieu, 2011). They are exacerbated by globalisation, which allows large flows of money and knowledge to accumulate in few parts of the world and among few people. Global threats like climate change and war and armed conflict will also be unevenly distributed in their consequences, favouring the rich and harming the poor disproportionately. Looking at the European case, Alberto Martinelli argues: “The combined effect of the global economic crisis (low growth, high unemployment, huge sovereign debt) and Middle Eastern wars (massive immigration of asylum-seekers, terrorist attacks by fundamentalist groups in European cities) exacerbate the rifts within the EU–creditor countries versus debtor countries, neoliberal Europe versus social Europe, national sovereignty versus federal union, free movement versus limited movement ... and create a rich breeding ground for national-populist parties and movements” (Martinelli, 2016: chapter 1).

4. Regulation: the challenge to go beyond legislation

The previous sections illustrate that the ritual of consuming and sharing information goes beyond the simple click of a button. To understand the motivations and the environment where information travels online is central to arriving at solutions that will address the root of the problem, rather than simply diminish its symptoms. Legislation has been offered as a way to make tech companies accountable for the negative externalities that their products bring, and they have been adjusting such products to regain credibility as platforms that connect people and allow discussions to happen in a healthy environment—at the price of losing the trust of users if they fail in doing so.
While some propositions seem interesting, many disregarded the problems elaborated in the previous sections of this paper, focusing on approaches that are unlikely to solve the problem. The next section critically assesses such proposals by both governments and tech companies, before elaborating recommendations on how to move forward on which solutions should be at the centre of the debate.

4.1 Country Overview: Which attempts have been made so far, and where?

Policymakers have often used regulation to address rising challenges, and their response to the ones posed by the use of social media for political purposes have proved no different. The first section of this article explored the environment in which mis- and dis-information flourishes. In light of this situation, legislation has been approved, discussed or rejected in Europe, aiming to hold tech companies behind these platforms accountable for such challenges. This section will explore these attempts, bringing to the debate the cases where legislation is desirable, where it falls short, and how it can negatively affect the problem. It is essential to use such cases as the baseline for future discussions, since in many cases policymakers write legislation before fully understanding the issue.

When it comes to solving pressing issues related to technology, it is common to look at policy solutions that define tomorrow in terms of what we know today. However, it is inadequate to take an approach designed for another era, hoping it fits the realities of tomorrow (Wheeler, 2018). Hence, legislation is not always the way to go, and if put in place, it needs to be carefully designed to address specific issues, contrary to what has been suggested by many European legislators. In this context, we can see a clash of approaches to this question in Europe. While Germany, France, and Italy have chosen a legislative approach, the UK and the Czech Republic have concentrated on building a government task force rather than punishing companies that act as distributors of content online (Serhan, 2018). Some legislation and actions have been approved, while others have been rejected. Our overview aims to identify when legislation can be beneficial, and when it is undesirable. Figure 2 summarizes the findings.

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>METHOD</th>
<th>IN FORCE</th>
<th>CHANGE EXISTING LAW</th>
<th>TARGET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>Legislation</td>
<td>Yes</td>
<td>No</td>
<td>Tech Companies</td>
</tr>
<tr>
<td>France</td>
<td>Legislation</td>
<td>No</td>
<td>Yes</td>
<td>Tech Companies</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>Task Force</td>
<td>Yes</td>
<td>No</td>
<td>Tech Companies, users</td>
</tr>
<tr>
<td>Italy</td>
<td>Legislation</td>
<td>No</td>
<td>Yes</td>
<td>Users</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>Task Force</td>
<td>Yes</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>European Commission</td>
<td>Recommendations</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In the German case, the Act to Improve Enforcement of the Law in Social Networks (BMJV, 2017)—known as NetzDG—has been in force since October 2017 and does not address the problem of information disorder in its entirety, as defined in the first section of this paper. The law innovates by putting the responsibility to control the content shared via online platforms on the companies that own them. It simply translates the legal obligations of the German Criminal Code to the online environment, obligating social networks to take down unlawful content that is shared by its users. Since information disorder is not in the Criminal Code, it is not subject to this law.

The law does not turn the companies into what some critics call a “Ministry of Truth”, making them the only actors deciding what kind of content is legal or illegal: it places the responsibility to judge whether content is lawful or not to a recognized third-party institution. While it may work in Germany, the line between hate speech and freedom of expression is often blurred. The risk with this law is that some countries could use the same approach to covertly curtail freedom of speech.

In France, a proposed law to fight against false information (National Assembly, 2018) went beyond applying existing laws to the digital sphere and proposed amendments to the Electoral Code, the Freedom of Communication law, and the Trust in Digital Economy law. The proposal was a reaction to the Emmanuel Macron campaign’s allegation that Russia had interfered in France’s 2017 presidential election with the intention of influencing the outcome in the final days before the vote. It was rejected by the French Senate in July 2018, but it brings interesting insights about how the State attempts to regulate social media.

In general, while raising important points related to the impact of false information on elections and democratic institutions, the proposed legislation shows a great lack of understanding of the complexity of the problem. The law deals with information disorder during electoral periods, but it lacks a more complex definition of what false news is. It goes as far as excluding parodic or satirical content, but it doesn’t specify how to spot false content and where to draw the line between fake and partisan, for example.

Once fake content is spotted, the law would authorise the state to take down content based on a judge’s decision. According to the proposal, “the judge, ruling in urgency (48 h), may order the dereference of the site, the removal of the content in question and the prohibition of its release online, the closure of the account of a user who has repeatedly contributed to the dissemination of this content, or even blocking access to the website” (National Assembly, 2018). It seems to ignore the feasibility of such a proposal, and the possibility that self-regulation of tech companies could address this issue in a more effective and targeted manner. Critics of the law said such legislation would be inefficient and potentially dangerous, opening space for a future government in France with a more radical approach to use it to increase censorship (The Straits Times, 2018).

Following the same line, a legislative proposal was submitted in Italy on 7 February 2017 (“DDL Gambaro”, Senato della Republica, 2017) in the Senate of the Republic, after the President of the Italian Competition Authority called for new rules on fake news. As in the French case, the bill proposes changes to existing legislation, introducing provisions that criminalise the circulation of
fake news. The proposal imposes a fine of up to €5,000 on anyone who publishes or circulates fake news; exaggerated or biased information on manifestly false facts; or any false information likely to cause public alarm, threaten public interests, or have a misleading impact on public opinion. Additionally, it punishes anyone who carries out a hate speech campaign against certain individuals or against the democratic process with a fine up to €10,000.

Unlike the French and German cases, however, the Italian proposal targets the users and not the tech companies, treating the fight against false information as an individual criminal offence. Still, all three cases take a top down approach to the problem (see summary in Figure 3), assuming that regulation will pressure platforms to take responsibility for the content they facilitate and users for the content they share. On the one hand, we have a public sector that tries to regulate industries of which it lacks specific knowledge. On the other hand, tech companies innovate without fully considering the impacts of their products on society. Hence, if regulation tends to fall short, and tech companies are not fully getting things straight with their internal rules and regulations, how can we move forward?

The United Kingdom has approached the question of how to stop the spread of fake news in a different way. Last July, the Digital, Culture, Media and Sport Select Committee, a cross-party committee of MPs appointed to scrutinise the government, set out 53 recommendations and conclusions on how to fight disinformation and “fake news” (Parliament, 2018). The recommendations do not only propose regulation, but also provide alternative approaches to such a
complex problem. It also does not fall into general easy answers such as digital literacy as a main solution to the problem—increasing digital literacy only accounts only for three of all the recommendations.

The final report—with non-binding recommendations at this stage—calls for a closer relationship between government and tech companies to set the right tone when dealing with such challenges. It recommends that the government rejects the term ‘fake news’, and instead puts forward an agreed definition of the words ‘misinformation’ and ‘disinformation’. As mentioned in our introduction, it is essential not to approach the challenge with a one-size-fits-all perspective, and a nuanced understanding can help frame a legal definition of the problem.

The Committee report continues, urging the UK to come up with the legal definition for the legal gap that will exist after Brexit, specifically concerning the General Data Protection Regulation in force in the European Union since May 2018, acknowledging users’ privacy is an important pillar to such challenges. Privacy laws are important to protect how personal data is handled by companies, containing provisions that could restrict bad actors from making disinformation campaigns effective by targeting susceptible individuals (Kornbluh, 2018).

An interesting innovation in the recommendations is how the Committee sees the role of regulators with regard to future challenges. It proposes that the duties of the Information Commissioner should be expanded, following the same logic of how the government audits and scrutinises the finance sector. It would impose a levy on tech companies operating in the UK to strengthen the IC into a regulatory body able to understand and provide solutions to what happens in such platforms—including the capacity to perform algorithmic auditing and identification of how false accounts are artificially boosting content. In order to do so, it proposes a legal definition of what tech companies are, making them responsible for the content they facilitate and increasing legal responsibilities with regard to transparency. This brings to the public sector the capacity to understand what happens in such companies and to regulate the challenges that arise accordingly.

As in the French case, the Committee worries about whether electoral rules are appropriate for the challenges of the social media era. However, contrary to the general approach used by French legislators, the Committee urges the government to update electoral laws to reflect changes in campaigning techniques, including micro-targeted political campaigning, increasing the length of the regulated period, definitions of what constitutes political campaigns, and absolute transparency of online political campaigning, among others. It does not give the government the power to force content to be taken down by the judiciary, as does the French proposal.

Unlike in other European countries, the alternative news business is thriving in Czech Republic (Schultheis, 2017). According to a poll conducted in the country, a quarter of voters in the country read and believe in alternative facts, while over one third of those who read do not care whether this information is a result of foreign intervention. The Czech case shows a worrisome trend, and reveals a factor that most legislation has not taken into account: rules alone are unlikely to change
the problem. Reading, sharing, and believing in false information is a result of ideological and cognitive beliefs that are exacerbated by the lack of knowledge of algorithms in the online environment.

Hence, using a similar approach to the UK, the Czech Republic chose to establish a task force before offering legislation to such challenges (see summary in Figure 4). As a result of the preliminary conclusions of the National Security Audit, which identified terrorism, radicalisation, and foreign disinformation campaigns as hybrid security threats, the Centre Against Terrorism and Hybrid Threats (Ministerstvo vnitra České republiky, 2018) was established within the Ministry of the Interior. Operating since the beginning of 2017, the centre acts as a specialised analytical and communications unit. Based on its monitoring work, the centre will evaluate challenges and come up with legislative proposals.

At the EU level (European Commission, 2018b), the Commission set up a high-level group of experts (HLEG) to advise on policy initiatives to counter fake news and disinformation spread online. The report with recommendations was aimed at establishing a framework for policy solutions, which was published in March 2018. The recommendations depart from a multi-dimensional approach with mutually reinforcing responses to enhance transparency of online news, protect data privacy, increase media and information literacy, and develop tools to empower users and journalists to tackle disinformation and safeguard the diversity and sustainability of the European news media ecosystem. The Commission’s approach aims to support the countries in following their own strategies, providing a framework for the debate but leaving regulatory remedies under EU or national laws.
National laws, however, seem to be falling short when it comes to disinformation (see Figure 5 for an overview of issues tackled by the countries mentioned). The different aspects of the information disorder seem to be prioritised differently in the countries analysed. Defining the problem is crucial to choosing the appropriate legislative tools to tackle them.

We need regulation, but we need governments and policymakers to understand what they are regulating, first. The UK’s step towards establishing a task force that brings all such details into the discussion is a great example of a long term approach to the problem. However, to see it turning into the ideal type of legislation is a step yet to be taken.

4.2 Self-Regulation, Fact-Checking, and Other Reactions to Regulatory Approaches

While legislation has so far seemed to fall short, tech companies are changing their platforms to address the challenges their products bring to society in their own way. As Hintz (2016; 333) argues: “Social media are sites through which the state enforces regulations, but they also formulate and enforce their own rules for acceptable user behaviour. In the area of content control, they have created platform-specific policies for accepting and rejecting content, which emerge from their commercial goals and are subject, to varying degrees, to influence by both user communities and the state.” In the aftermath of Russian interference in the 2016 US presidential election, Twitter, Google, and Facebook, the biggest tech companies that had their platforms used to spread disinformation, were summoned to Congressional hearings to explain how their platforms allowed these campaigns to flourish. Facebook and its subsidiaries WhatsApp and Instagram together with Twitter and Google’s YouTube account for the majority of internet users worldwide (Smart Insights, 2018).
The process of accepting the role of their platforms in harming electoral processes did not come overnight. Not even one week went by after Trump was elected before Facebook’s CEO Mark Zuckerberg came forward to express his views on the role of his company in the elections (Zuckerberg, 2016). His Facebook post on 19 November 2016 came after a series of criticisms over the lack of action in controlling the quality of the content shared in social networks. According to Zuckerberg at the time, “more than 99% of what people see is authentic”, and, of what people are exposed to, “only a very small amount is fake news and hoaxes”. While he was right in stating that “identifying the truth is complicated” when it comes to content shared, he clearly underestimated the impact of misinformation on the elections.

The position of his company and other tech giants has changed drastically since. Facing legislative attempts at regulation in many countries, they changed their platforms to address questions related to their influence in recent political processes in their own way. This section will analyse these recent attempts and critically assess whether they have been successful.

4.2.1 Facebook

In the aftermath of the US Elections, Facebook quickly prepared to avoid the same problems in the French and the German ones, which took place in April and September 2017. In those countries, the platform partnered with fact-checking organisations to find ways of improving the content shared by users on their platforms. In a recent statement entitled “Preparing for Elections” (Zuckerberg, 2018), Facebook’s CEO Mark Zuckerberg shared the recent efforts to tackle false information, fake accounts, and hate speech on the platforms under his control ahead of the midterm elections in the US and the general elections in Brazil.

Facebook’s struggles are not only a problem in established democracies. In 2013, the platform launched the Internet.org project, renamed Free Basics in September 2015, aiming to increase the connectivity of people in developing countries (Mohan, 2015). In a white paper at the time, Zuckerberg set out a vision of connectivity as a human right (Zuckerberg, 2013). The project is part of Facebook’s business plan to increase its user base in the future, but the way this expansion was done did not account for the specific political situation of the countries it operated. For example, the group responsible for spreading hatred online against the Rohingya in Myanmar was shut down by Facebook only eight months later (Serrato and Meyer-Resende, 2018), showing the lack of political awareness and attention of Facebook to what was happening at the time in countries it operates. In another case, users claimed not to have Internet access even though they were active on Facebook, which showed a deep lack of systemic understanding (Mirani, 2015).

On the positive side, the platform launched the Facebook Journalism Project to collaborate with the news industry and work towards a more responsible approach to informing readers in the digital age (Facebook, 2017). The idea is to come up with new storytelling formats, work closely with independent media to promote local news, and adapt business models of media outlets—some of the issues that are directly or indirectly related to the spread of false information. More
recently, the company decided to allow independent researchers to access their data, looking deeper into the impact of social media on elections and democracies (Schrage and Ginsberg, 2018) and, via WhatsApp, offered grants up to $50,000 for research with special attention to misinformation (WhatsApp, 2018).

In light of the misuse of social media during recent elections, Facebook implemented tools to ensure more transparency behind ads they allow in their platforms (Facebook Newsroom, 2018) and showed a more active role in taking down networks of fake accounts to avoid misinformation to spread. However, it is not clear from reading their transparency report (Facebook, 2018) whether the actions taken were enough to fight the spread of false information, hate speech, and other influences. Organised groups found instant-messaging apps such as WhatsApp an effective tool to spread false news, and it is still not clear how the platform is dealing with such challenges. Among other omissions, the report does not give us a clear picture of how the platform was used in past elections and whether this impact is diminishing in current electoral periods.

Another issue the company has struggled to resolve is the psychological pressure its content moderators suffer in their work of taking down content that goes against the platform’s guidelines. Non-disclosure agreements and harsh working conditions make this job very controversial (Solon, 2017).

With regard to news sharing, the company announced in January 2018 that it would alter the News Feed algorithm in order to prioritise content produced by friends and family above content shared by media and businesses (Wong, 2018). While this indeed was helpful to stop spreading false information, it also harmed the business model of established news outlets. Folha de São Paulo, the largest-circulation Brazilian newspaper, decided to stop publishing content via Facebook on 8 February 8 2018, accusing the platform of diminishing the visibility of professional journalism. According to the newspaper, Facebook user interactions with the 10 biggest news sources in Brazil dropped 32 percent compared with the same month in the previous year after the algorithmic change (Folha de São Paulo, 2018).

4.2.2 Twitter

The news and online networking service Twitter has been associated with problems of mis- and dis-information mainly through its dynamics of spreading content. Tweets not only spread uncontrollably though the online behaviour of users, but the service is particularly prone to the misuse by programmed ‘bots’. These automatic accounts have been accused of spreading and disseminating false and misleading information, thereby influencing major political events. The platform has been subject to criticism for evaluating cases of misinformation too favourably for the right to freedom of expression, thereby facilitating political influence on election campaigns, most notably ahead of the US election 2016.
Twitter has made some changes to ensure a healthier debate online. In recent months the platform has been increasing the rate of account suspension as a result of pressure from the US Congress (Timberg & Dwoskin, 2018). The company has also increased its effort to identify and block bots from tweeting. Alcott et al. (2018) observe in their study that while the absolute number of misinformation cases is much higher on Facebook, Twitter has not managed to lower theirs as effectively after the US election. In the Twitter blog, several changes in the design of Twitter were announced in February 2018, limiting the possibilities of posting, retweeting, or liking content through multiple accounts at the same time (Roth, 2018). The announcements as such, however, remained vague regarding specific prohibitions or consequences.

With regard to advertising, Twitter is in the process of creating stricter standards for election-related advertisements, while Facebook has begun testing a system in which viewers of disputed ads are pointed to “alternative recommended content” (for more information on these and other examples, see European Commission, 2018a). As in the Facebook case, it is still unclear whether the pace at which the company is making changes is fast enough to avoid the negative impacts of disinformation campaigns in the future. Because of recent restrictions on publicly available data, it is hard for independent third parties to back up whether the actions the companies are taking are indeed effective responses to such challenges.

4.2.3 YouTube

The video platform YouTube plays a slightly different role in the information disorder. The main issues associated with it are false or reprehensible content in videos, as well as the “Recommended” and “Autoplay” functions, which are guided by an algorithm (Tufekci, 2018). Autoplay can lead users to problematic content that may even be illegal in some jurisdictions, for example Holocaust denialism in Germany. The algorithm has been accused of guiding users towards more and more...
extreme content, including conspiracy theories (ibid). Since the platform’s business model is based on the maximisation of views per video, this strategy of luring users into more sensationalist content makes sense.

In 2017, YouTube announced that it would hire up to 10,000 human reviewers, work harder to ensure compliance with its own guidelines, and fight “violent extremism” (YouTube, 2017). These commitments seem fairly vague, and YouTube was heavily criticised for not taking a stronger stance. Also, as we have seen in Facebook’s case, the role of human reviewers is psychologically challenging, which is why YouTube has announced they will limit reviewing activity to four hours per day (Statt, 2018).

Another change was the more-effective removal of ads from content that is attacking or harassing people based on their religion, race, gender and other categories (Schindler, 2017). In any case, the challenges ahead remain: combining a profitable business model with the need to avoid problematic and radicalising content through self-regulation.

With regard to hate speech, in 2016, Facebook, Twitter, YouTube, and Microsoft signed on to the European Commission’s Code of Conduct on Countering Illegal Hate Speech Online, committing to take measures to monitor and remove illegal postings in a timely manner. Yearly evaluations by the Commission indicate that significant progress has been made on this front since the start of the initiative, with the four companies now removing 70 percent of reported instances of online hate speech, up from only 28 percent in 2016 (European Commission, 2018c). Moreover, the initiative has recently expanded to include Google+ and Instagram as well.

4.3. The Path Forward

In light of the pressure that legislation is imposing to tech companies, some companies have gone beyond individually changing their platforms and have agreed on a code of conduct to fight false information in their platforms (European Commission, 2018). Many of the recent changes made by the platforms individually addressed similar aspects of the problem, but the code of conduct summarizes a comprehensive set of actions to be taken by such companies. The code presents actions on five different fronts:

1. Increase the scrutiny of ad placements in order to reduce revenues of the purveyors of disinformation;
2. Increase transparency about political and issue-based advertising;
3. Guarantee the integrity of services through intensifying and demonstrating the effectiveness of efforts to close fake accounts;
4. Empower consumers by giving them access to different news sources and providing them with easily-accessible tools to report disinformation, and
5. Empower the research community by providing them with relevant data on the functioning of their services.

The code was welcomed by the European Commission, but it also received criticisms from media advisors and lobby groups for the advertising industry for failing to present more concrete and measurable actions (Chee, 2018). Similar to the proposals offered individually by these companies, it is very hard to assess whether such measures are having a meaningful impact in taking care of the problem. Taking down networks of fake accounts and adapting algorithms are in principle positive developments, but it is still unclear to what extent they are doing enough to really solve the problem.

5. Conclusion

By focusing so much on taking down untrustworthy news, the tech companies have overlooked other forms of disinformation: more subtle partisan media, images and videos, audio sent via WhatsApp, and more. Also, taking down content does not tackle the problem at its root. In this sense, without comprehensive monitoring across platforms and more concrete checks and balances that ensure tech companies take action, which in some cases may go against their business models, such proposals may fail to go beyond a marketing strategy to safeguard their image. We also need tech companies to play an active role in preventing their platforms from being used against democratic values and institutions in the future, proactively in understanding the threats that may come from the development of new platforms, apps, and technologies.

On the other hand, we see a strong regulatory approach from politicians who seem to fail to understand what they are regulating in the first place. On the proposed legislation analysed in this article, we see attempts to address the effects of the problem, but not attacking their roots, as with the tech companies’ approaches. The complexity behind those platforms clashes with the current regulatory approach of many policymakers that often uses one-size-fits-all definitions to tackle fake news. Such a one-size-fits-all approach won’t work in practice.

Legislation so far has shown different approaches to the problem of false information online. The key question seems to be: who should be in charge of dealing with false information and hate speech? The German model outsources the question to the courts and companies by making them accountable for hate speech in their platforms; the French proposal outsources it to courts and companies with regard to false information during elections. Italy’s proposal would make individuals responsible for the content they share. In other parts of the world, we see the US outsourcing regulation to companies; China to the government; and Sweden to civil society (Mattsson and Parthasarathi, 2018). With many proposals out there, it is still early to assess their efficacy in recent electoral processes, but the myriad of solutions shows that tackling the spread of false information associated with democratic political processes is still an ongoing task with more questions to answer than solutions to offer.
A complex problem brings complex solutions. When talking about the problem of computational propaganda, Secretary Madeleine K. Albright (2018) said, “citizens are speaking to their governments using 21st century technologies, governments are listening on 20th century technology and providing 19th century solutions”. Tech companies, governments, citizens, media, and civil society organisations need to cooperate towards a greater understanding of the impact of new technologies and come up with clever and complementary approaches to solve the challenges brought by its use. It is time to look for 21st century solutions to 21st century problems.
Bibliography


