



Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe

Florian Saurwein & Charlotte Spencer-Smith

To cite this article: Florian Saurwein & Charlotte Spencer-Smith (2020): Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe, Digital Journalism, DOI: [10.1080/21670811.2020.1765401](https://doi.org/10.1080/21670811.2020.1765401)

To link to this article: <https://doi.org/10.1080/21670811.2020.1765401>



Published online: 27 May 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

ORIGINAL ARTICLE



Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe

Florian Saurwein  and Charlotte Spencer-Smith

Institute for Comparative Media and Communication Studies (CMC), Austrian Academy of Sciences and University of Klagenfurt, Vienna, Austria

ABSTRACT

Online disinformation poses a challenge to democratic societies and has become a prominent issue on the research and political agenda. While many analyses focus on patterns of distribution and reach of disinformation, this article contributes to the analysis of strategies to counter disinformation. Employing a governance perspective, it provides a descriptive analysis of the emerging mix of governance responses in the European system of multilevel governance and on the continuum between market and state. Results of the analysis show that the proliferation of disinformation on social media has developed from a socio-technical mix of platform design, algorithms, human factors and political and commercial incentives. Actors and technologies involved provide a starting point for targets of governance within an accountability network. In practice, national governance responses are uneven across the EU, but individual countries pressing for stronger regulation of internet platforms and a weakening of liability protections. In addition, the European Commission has intensified its efforts to combat disinformation and put additional pressure on platforms to take action and provide some level of transparency. However, clarity about the effects of these measures is blurred by contradicting evidence and barriers for research to access platforms and relevant data.

KEYWORDS

Disinformation; social media; internet platforms; governance; self-regulation

Introduction

Online disinformation poses a challenge to democratic societies that depend on public debate and well-informed citizens who express their free will in political processes. It has therefore become a prominent issue on the scientific and political agenda. “Fake news publishers” exploit social media to distribute disinformation, influence opinion and interfere in elections, posing a threat to democracy. The Ukraine conflict of 2014 and the US presidential election in 2016 have been focal points of concerns about the use of disinformation to underhandedly serve political ends. In Europe, the refugee crisis of 2015, the Brexit referendum in 2016, and a series of major elections, notably the

German federal and French presidential elections in 2017, raised concerns about the scale of disinformation and its threat to European democracy.

“Fake news” has been the most popular term used to problematize the subject, particularly since journalistic investigations into economically-motivated disinformation on social media during the US elections in 2016 (Subramanian 2017). However, the politicization and emotionalization of the term (Brummette et al. 2018) and its appropriation to attack established news media (Ross and Rivers 2018), have led to attempts at a more appropriate problem definition (Wardle and Derakhshan 2017; Sänglerlaub 2017). The European High level expert group on fake news and disinformation (HLEG 2018, 5) thus defines disinformation as “forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit.”

Scholarship has produced insightful studies of online disinformation, focusing particularly on its threat to democracy. Studies of phenomena termed “fake news” have covered a broad spectrum, including news satire, news parody, fabrication, manipulation, advertising and propaganda, which vary in the author’s immediate intention to deceive and level of facticity or truthfulness (Tandoc, Lim and Ling 2018). Of these, fabrication, manipulation and propaganda have received significant attention because of the threat they pose to democracy. These are also characterized by their low levels of facticity and a high level of intention to deceive. Wardle and Derakhshan (2017) distinguish between misinformation, false information that is not created to cause harm; disinformation, false information that is created to cause harm; and malinformation, information that is not false but is intended to cause harm. Disinformation by this classification includes fabrication, but also content that has been manipulated, given a false context or comes from a source posing as someone else.

Further studies chart the rise of disinformation in the age of social media platforms: Bakir and McStay (2018) situate the rise of “fake news” in the context of interconnected crises in journalism that have arrived with platformisation; the weakening of legacy news media, the accelerated round-the-clock news cycle, mis- and disinformation being shared horizontally through user-generated content, the emotionalization of online media, and the online advertising that benefits from this, for example, through clickbait. Allcott and Gentzkow (2017) point to disinformation as a tool of foreign interference by investigating “fake news” websites active during the US presidential elections, identifying profit-driven actors, and also noting social media’s disproportionate role in distribution, compared to other media. These observations have since been expanded with Marwick and Lewis’ (2018) broader review of the landscape of information manipulation beyond classic “fake news” on the American Internet. In a case study of the Russia-Ukraine conflict, Mejias and Vokuev (2017) have situated disinformation within a larger political context and found that states do not have to pull all the strings if they “can rely on citizens’ do-it-yourself disinformation campaigns.” The Oxford Internet Institute has produced a series of national case studies covering not just the US, Russia and the Ukraine, but also Canada, Poland, Taiwan, Brazil, Germany and China (Woolley and Howard 2018). While some analyses focus on scope, motivations and techniques of disinformation campaigns, others try to assess the reach of disinformation as measured by metrics such as visits and time spend on

website, and social media interaction (Fletcher et al. 2018). In recent years, the disinformation threat has also elicited governance responses on national levels, and in the case of the EU, at an international level. However, so far these governance responses to disinformation have gained less academic attention than its distribution. This article contributes to closing this gap with an analysis of the current governance of disinformation on social media in Europe.

The paper begins by setting forth its theoretical framework drawing upon the governance approach, alternative modes of regulation and ideas of distributed accountability. It then examines disinformation as a socio-technical assemblage of risk. The paper further analyzes governance responses within the European Union. First, it looks at government responses at EU level and at national level, focusing on France, Germany and the United Kingdom. Then it examines non-governmental responses by fact-checking organizations and by self-regulation by major social media companies. The final section summarizes main results of the descriptive analyses of an emerging accountability network and reflects upon on involved accountability problems.

Governance and Accountability: A Theoretical Perspective

From a theoretical perspective, the analysis builds on multidisciplinary concepts of governance research. It takes a risk-based approach and combines it with descriptive, institutional analyses of governance arrangements, considering the existing institutions in the system of multilevel governance and on the continuum between the market and the state. This allows for a comprehensive analysis of the emerging mix of governance responses to disinformation.

The risks caused by the distribution of disinformation provide justification for regulation and governance. These include libelous claims, increasing social distrust, political polarization, and election interference. From a public-interest point of view, governance should minimize such risks. Accordingly, a “risk-based approach” (Black 2010) identifies and examines these risks, their causes and opportunities to reduce them.

When combating risks, it is tempting to immediately consider hard regulation by national governments as the first port of call. However, the governance approach (e.g., Rosenau and Czempel 1992; Rhodes 1996) expands this rather narrow view and considers both horizontal and vertical dimensions of governance (Engel 2004; Puppis 2010). Looking vertically, national government is part of a multi-level governance structure (Hooghe 1996), embedded and complemented by inter- and supra-governmental institutions, such as the European Union. Looking horizontally, statutory regulation by law is complemented by alternative modes of governance. These comprise self-organization by individual companies, collective self-regulation by industry branches and co-regulation – regulatory cooperation between state authorities and the industry. From an institutional perspective, governance options can be located on a continuum ranging from market mechanisms at one end, to command-and-control regulation by state authorities at the other (Latzer et al. 2003; Bartle and Vass 2005). Alternative modes of governance are generally preferred for the media sector due to concerns that direct regulation may lead to censorship. Print journalism, for instance, has a long tradition of self-regulatory bodies. Additionally, media governance in

Europe has also introduced co-regulatory approaches for the protection of minors, advertising rules and for measures to combat hate speech, for example, most recently by means of the Audiovisual Media Services Directive (AVMSD). Co-regulation enables self-regulation within a legal framework and public oversight and creates an arrangement where state and industry share regulatory competencies and responsibilities.

Hence, the analysis of risks and governance arrangements also directs us to questions of accountability. Disinformation on social media involves several types of actors (e.g., platforms, users, public institutions, professional journalism) who may serve both as objects and subjects of governance strategies. Some types of actors (e.g., users, fake news publishers, social media platforms) are directly involved in the production of risks and thus serve as objects for governance. Other types of actors (e.g., states, but also social media platforms) have regulatory capacities to set and enforce rules to reduce risk. Together, involved actors have competencies and capacities to act in relation to risks and therefore form “accountability networks” (Neuhäuser 2014; Sombetzki 2017). A multi-dimensional approach to disinformation (HLEG 2018) will allocate accountability in a shared, distributed and cooperative structure (Helberger, Pierson, and Poell 2018; Saurwein 2019).

This article explores the different institutions and types of actors involved in the spread of disinformation (The Problem of Online Disinformation as a Socio-Technical Assemblage) and governance reactions (Governance Reactions on the Continuum between States and Markets). In combination, these analyses allow for mapping of the emerging accountability network.

The Problem of Online Disinformation as a Socio-Technical Assemblage

The rise of social media has promoted an explosion of disinformation that is cheap and easy to produce, disseminate internationally and test on multiple audiences. The problem can be conceptualized as an assemblage of social media platforms, actors and big data (Woolley and Howard 2016). Its proliferation on social media has developed from a socio-technical mix of platform design, algorithms, human factors and political and commercial incentives. Just as journalism encompasses social actors, technological actants, as well as work-practice activities and audiences, so too does fake news production and distribution (Lewis and Westlund 2015). For the purposes of this analysis, we identify actors at different levels, including social media companies who determine the policies and technological platform design, “fake news publishers” who use these systems to disseminate disinformation, and users who fuel distribution by clicking, liking, sharing and watching. In addition, in a socio-technical process, non-human technological actants, such as interface design and news feed algorithms, play a key role in the “fake news process.”

Producers and sharers of disinformation: “Fake news publishers” have economic or political motives, varying in tactics depending on their goals. Economically motivated publishers have much in common with the producers of clickbait: they earn money from pay-per-click advertising on their websites (Subramanian 2017). In contrast, the primary goal of politically motivated actors is to manipulate public opinion and disrupt politics. These operations are known to use fake accounts, pretending to be real

people who connect with real users and then share disinformation (Schuster and Iramova 2018). They also use botnets to exploit social media algorithms by re-posting content so that it appears popular and therefore relevant to the algorithms (Osipova and Byrd 2017). Tactics are by no means fixed but change over time. For example, in the US, Russian disinformation actors have posed as civil rights activists on social media to organize real-life demonstrations (Broderick 2019). They have also moved away from producing their own disinformation, which is costly and can appear inauthentic, to promoting disinformation and hyperpartisan propaganda that originates in the target country (Scott 2019).

Social media users: Social media users hold two roles; first they are media audiences who receive information and disinformation, and as such are also the key battleground in the fight against disinformation. Disinformation becomes harmful only once it is received and consumed by social media users. Second, the technological actants and affordances of social media have blurred the line between audience and actor, so that media consumers can also become media distributors (Lewis and Westlund 2015). The interactive design of platforms and algorithms gives ordinary users a role in amplifying and distributing disinformation. As users engage with content, they signal to recommendation algorithms that this content is interesting and relevant, and it may be recommended to other users. Responsibility for algorithmic boosting is therefore diffused across multiple users, whose roles and contributions are obscured, and who may not be aware that they are contributing to a harm.

Social media companies: Platform design choices that enable the growth of disinformation can partly be explained by the corporate goal of maximizing growth. Thompson and Vogelstein (2018) argue that competition between Facebook and Twitter in news-sharing led to design choices at Facebook that fueled the disinformation problem on the platform. Similarly, critics of YouTube point to its video autoplay function and complex recommendation algorithm, both of which keep users on the site, but also recommend conspiracy theory and other misinformation videos (Lewis 2018). While changing the media landscape in ways that provide opportunities for disinformation to flourish, social media platforms have also weakened professional news and investigative journalism that would otherwise be well-placed to challenge disinformation (Frenkel, Casey, and Mozur 2018).

Platforms and technology: The design of popular social media platforms makes it difficult to differentiate between media sources and enables the fast dissemination of disinformation. Weblinks are displayed in a standardized format, so that a fake news link will appear the same as a link to an established media organization (Thompson and Vogelstein 2018). Both Facebook and Twitter offer “share” and “retweet” buttons under every post, fueling rapid sharing between users. Hidden within platform interfaces, algorithms reward content that attracts user engagement, such as clicks, likes, shares, comments and views. User activity signals relevance to algorithmic recommendation systems such as news feeds (Thompson and Vogelstein 2018), “trending topics” features (Manjoo 2017), and video autoplay recommendations. Algorithms therefore promote sensational content, providing an advantage to clickbait and fictitious but curiosity-inducing stories. It has been hypothesized that online disinformation attracts attention through outrage and novelty, compelling users to click (Vosoughi, Roy, and Aral 2018).

Governance Reactions on the Continuum between States and Markets

The risks and the diverse causes of disinformation have elicited various governance reactions by different entities in the multilevel governance system. One can observe the development of governance at the level of the European Union (Reactions by the European Union) in major European countries such as France, Germany and the UK (Reactions in Selected European Countries) and by means of industry self-regulation and self-organization of individual companies (Responses from the Market, Industry and Civil Society).

Reactions by the European Union

Disinformation has been on the political agenda in Europe since at least the Ukraine conflict of 2014 when Russia was accused of conducting an “information war” on the Internet. Accordingly, the European Commission developed an action plan to counter Russia’s disinformation campaign (European Council 2015) and established the East StratCom Task Force to act against Russian disinformation. The Task Force runs the website, “EU vs Disinfo,” which identifies and refutes Russian disinformation.

Ahead of the European Parliament elections in May 2019, the EU Commission intensified its activities (European Commission 2018a). This started in early 2018 with the establishment of the High-level expert group on fake news and online disinformation (HLEG), a public consultation, a multi-stakeholder conference, a colloquium and a Eurobarometer survey. Subsequently, the future strategy was laid out in a communication on “Tackling Online Disinformation: A European Approach” (European Commission 2018b). It aimed at strengthening the efforts of internet platforms in combating disinformation and at supporting independent fact-checking. The approach also included technical measures, such as protecting elections from cyberattacks, promoting reliable authentication systems and artificial intelligence, blockchain and cognitive algorithms to identify disinformation and improve transparency. Furthermore, the communication suggested promoting media literacy, quality journalism and strategic communication to counter disinformation.

Among the central addressees of the Commission’s strategy are major internet platforms because of their central role in the distribution and amplification of disinformation. According to the Commission, platforms have long failed to react appropriately to disinformation. It therefore called upon them to strengthen their efforts, noting that, “self-regulation can contribute to these efforts, provided it is effectively implemented and monitored” (European Commission 2018b, 7). In October 2018, the Commission introduced an EU Code of Practice on Disinformation. Signatories, including internet and advertising companies such as Google, Facebook, Twitter and Mozilla, voluntarily agree to fight disinformation and manipulative election advertising. Among other things, they commit to making the origin and scale of political advertising more transparent, preventing “fake news” publishers from profiting from advertising revenue and removing fake accounts faster. They also commit to set clear rules for the misuse of bots on their platforms. Signatories also promised to empower users by helping them make more informed decisions and increasing transparency around political ad targeting. In addition, they pledge to support and not to prohibit independent

research on disinformation, and to report on implementation and progress for third party review. Ahead of the European election in 2019 platforms submitted monthly reports (European Commission 2019).

In addition to the Code of Practice, the Commission has taken further measures (see European Commission 2018c). These include supporting an independent European network of fact-checkers, establishing a European platform on disinformation through linking national organizations within the “Connecting Europe” program, and supporting the work of the Network Information Security Cooperation Group to identify best practices for protecting elections from disinformation. 40 million euros have been committed for research and innovation projects to identify disinformation through research funding programs such as Horizon 2020. However, supportive measures for quality journalism remain comparatively weak. In its communique, the Commission simply encourages member states to counteract market failures that harm the sustainability of quality journalism, references the compatibility of funding with state aid rules and raises the prospect of supporting initiatives that promote quality journalism and media pluralism. Since then, projects promoting investigative journalism, the modernization of newsrooms and quality news content on European affairs have been funded, but on a modest scale. The call for projects promoting quality content on European affairs, for example, committed 1.9 million euros (European Commission 2018c).

To build resilience against disinformation, the Commission supports media literacy, including through existing media literacy programs and initiatives, such as #SaferInternet4EU and “Media Literacy for All.” In the area of strategic communication against disinformation, it focusses on internal coordination of communications and the development of awareness-raising. For example, an internal Network against Disinformation has been set up to “better detect harmful narratives, support a culture of fact-checking, provide fast responses and strengthen more effective positive messaging” (European Commission 2018c, 12). The budget of East StratCom Task Force has also been increased from 1.9 million to five million euros for 2019.

In December 2018, the Commission published the first report on the implementation of the communication on “Tackling Online Disinformation” (European Commission 2018c). At almost the same time, an Action Plan on disinformation was introduced, concretizing EU measures into four thematic pillars (European Commission and High Representative of the Union for Foreign Affairs and Security Policy 2018). Alongside implementation of the Code of Practice and awareness-raising measures, the Action Plan aims at enhancing analytical capabilities (additional digital tools, data analysis capabilities, expert personnel) for better identification of disinformation and evaluation of its scope and effects. To improve cooperation, the Plan calls for an early warning system, national contact points and encourages exchange information between member states.

Effects and Evidence: Disinformation around the European Parliament Elections 2019

A brief overview of the EU approach reveals a high level of engagement in its fight against disinformation, driven by serious concerns of the risk of online disinformation

and election interference ahead of the European Parliament elections in May 2019. This also presented an opportune window to investigate the extent and nature of online disinformation in the EU. However, investigations into suspicious activity produce different views of the scale of the problem. An Oxford Internet Institute report found that less than 4% of tweets in Europe were “junk news,” “ideologically extreme, misleading, and factually incorrect information,” or from identifiable Russian sources. The same analysis found that while junk news received more engagement on Facebook, professional news sources had far greater visibility on the platform (Marchal et al. 2019). However, the left-leaning activist organization Avaaz (2019) published that it had reported disinformation networks to Facebook comprising over 500 pages and groups that had posted content viewed half a billion times. Differing messages about the scale of online disinformation in Europe can be attributed to different research focuses, and illustrate challenges in attempting to establish a complete picture of the problem. A further complicating factor is the suspicion that much disinformation circulates on encrypted messaging services like WhatsApp, an issue that has also appeared in Europe during the coronavirus crisis (Apuzzo and Satariano 2019; Delcker, Wanat, and Scott 2020). Other studies suggest that disinformation is moving away from classic “fake news” websites to other tactics, such as selectively amplifying real news stories and emphasizing polarizing topics (Krasodonski-Jones et al. 2019). Furthermore, information manipulation is increasingly being carried out by domestic groups, such as right-wing populist activists (Apuzzo and Satariano 2019; Avaaz 2019). This suggests that online manipulation activities are a dynamic challenge that can evolve faster than attempts to govern them.

Reactions in Selected European Countries

Overview of National Efforts in the EU

On a national level, efforts across EU member states are uneven, encompassing a spectrum from legislating a higher level of platform responsibility than provided for than the EU Code of Practice, through efforts to improve media literacy and establish government monitoring units, to a seemingly total reliance on EU-level efforts. On one end, France and Germany are the only countries to have taken direct legislative steps to increase platform responsibility in specific areas, although the idea of a more expansive legal “duty of care” is being floated in the United Kingdom. Bills in Ireland, Italy and Lithuania that would have sanctioned the production or distribution of disinformation have been considered but not materialized (O’Halloran 2017; Verza 2018; Gerdziunas 2019). The exception to this has been the successful passing of a law in Hungary that imposes a five-year prison sentence for spreading false information that alarms the public or prevents government efforts to protect people in response to the coronavirus crisis (Walker 2020). This, however, can be better understood as part of an authoritarian turn within Hungary masquerading as a measure to prevent disinformation.

One key difference between Europe and the United States is that some European countries have preexisting criminal laws against hate speech, libel and defamation. Supplementary laws and enforcement could update these laws so that they can be re-

purposed to combat disinformation, as has been the case in Germany. However, in part in reaction to concerns about censorship, several countries have chosen to pursue an approach that focuses on user awareness. Belgium, Finland, Luxembourg and Sweden have already launched awareness and media literacy initiatives, while the Czech government has established a monitoring unit that informs the public of pro-Kremlin disinformation efforts (Schultheis 2017). The Baltic countries Estonia, Latvia and Lithuania have a more complex history of targeted disinformation tactics sponsored by the Russian state, including through television broadcasting, which has prompted stricter regulation of broadcast media (Gerdziunas 2017). As disinformation moves online, the NATO Strategic Communications Centre of Excellence and the EU East StratCom Taskforce monitor disinformation efforts, and the Lithuanian military is collaborating with media and civil society volunteers known as “Baltic elves” through a platform used to debunk disinformation stories (Gerdziunas 2018). However, national initiatives seem to be less prominent in other countries. An Austrian government response to a parliamentary question confirms significant reliance on EU-level measures and a lack of national initiatives (Parlament der Republik Österreich 2019), while Bulgaria has been criticized for also lacking a national plan (Meta 2019). This places them in a category of countries that have neither national measures nor coverage under the Facebook fact-checking program, meaning that both governance and industry measures are weak compared with other European countries.

The following cases focus on the three largest countries in the EU by population, Germany, France, and the United Kingdom. While this focus remains on the Western European context, it serves to illustrate notable emerging changes to governance. Taken together, these initiatives indicate a regulatory trend away from the liability protections of the European e-commerce Directive (2000/31/EG), and increased focus on platform responsibility.

France

France is seen to be on the front line in the area of platform regulation in Europe (Kayali, Momtaz, and Vincour 2019). It has initiated a national digital tax and is pushing for a law against online hate speech. A recent report suggested regular audits of social media platforms and more transparency around their internal processes for handling harmful content and hate speech (see Desmaris, Dubreuil, and Loutrel 2019).

Moreover, France was the first European country to introduce the law against the manipulation of information (no. 2018-1201, 2018-1202), in reaction to indications of Russian efforts to influence the presidential elections in 2017 (Noack 2018), in order to undermine the later president Emmanuel Macron (Cichowlas 2017). The new law passed in November 2018, allowing judges to order the immediate removal of online articles deemed to be disinformation during election campaigns. The law states that users must be provided with information on usage of personal data, that online political campaigns must disclose financiers and amounts spent, and empowers the national broadcasting agency to suspend television channels under foreign influence which “deliberately disseminate false information likely to affect the sincerity of the ballot” (Fiorentino 2018). Sanctions include one year in prison and a fine of €75,000.

French opposition parties and journalist associations have criticized the law as an attempt by President Macron to suppress unfavorable information (Zeit Online 2019).

United Kingdom. In the United Kingdom, the aftermath of the Brexit referendum in 2016 brought concerns about online manipulation and the role of social media in electoral interference to the fore. This was problematized in the House of Commons Fake News Enquiry which did not find conclusive evidence of a large-scale Russian “fake news” campaign, but emphasized the UK’s vulnerability to online manipulation and a lack of regulation around digital political campaigning (House of Commons 2019). Furthermore, data released by Twitter shows that fake accounts linked to Russia posted thousands of tweets about Brexit before the referendum (Field and Wright 2018).

Although legislation has not yet been initiated, the 2019 Online Harms White Paper provides an insight into the government’s plans to establish a new statutory duty of care and a new regulator for platforms (HM Government 2019). One concern about introducing an online harms regulator is that it could quickly become a “regulator for everything” (Miller et al. 2018). A duty of care would represent a departure from the existing regime of wide-ranging liability protections and imply a requirement for proactive measures by platforms. The proposed regulator would be responsible for developing codes of practice addressing harms. In particular, the paper envisages a “code of practice that addresses disinformation” that would “ensure the focus is on protecting users from harm, not judging what is true or not” (HM Government 2019, 72). Its suggestions for the content of the code can be broadly divided into platform measures to improve social media literacy and transparency, promoting diverse news content and quality news media, developing and enforcing policies against bad actors, and involving fact-checking organizations, especially during election periods. Notably, it does not mention any obligation to remove disinformation content, instead preferring platforms to reduce the visibility of disinformation and increase the visibility and accessibility of quality news.

Germany. In Germany, revelations about disinformation during the 2016 US presidential elections raised the fear of interference in the 2017 German federal elections. In the preceding years, Russia was suspected of cyberattacks and the spread of fake news through the use of bots, trolls, and pro-Russian TV channels (Snegovaya 2018). However, analysis shows very little Russian online interference around the German elections, with most disinformation originating from domestic far-right actors (Sängerlaub, Meier, and Rühl 2018). Fears of a major Russian interference were therefore overestimated (Fürstenau 2017).

The most significant development in platform governance in Germany so far has been the introduction of the Network Enforcement Act in 2017, primarily to combat online hate speech. The act obliges social networks to remove or block access to manifestly unlawful content within 24 h of receiving the complaint or within 7 days in less clear-cut cases. They must also maintain an effective and transparent complaints procedure. Failure to comply with the act can result in a fine of up to 50 million euros. The act itself does not explicitly refer to disinformation, but the Ministry explicitly

states that the act can be used against disinformation where it constitutes insult, defamation or slander, which are criminal offenses in Germany (Bundesjustizamt 2018). The act has proven controversial: proponents see it as a strong instrument to enforce existing laws on social media, while critics point to instances where platforms have deleted more content than necessary out of fear of incurring large fines (Krempel 2018).

Responses from the Market, Industry and Civil Society

Fact-Checking

Fact-checking is one of the most important non-governmental responses to disinformation and has developed significantly in Europe in recent years, with around 30 active organizations. These organizations, however, face significant challenges in countering online disinformation (Graves and Cherubini 2016). First of all, debunking online disinformation is only one of the roles performed by fact-checkers, and verifying claims made by politicians in public continues to be a central function. The limited time, funding and staff available to fact-checkers are thus not devoted to online disinformation alone. Second, many NGO organizations rely on grants and struggle to find sustainable models of funding. Third, fact-checking websites do not typically command large readerships of their own. They aim instead for stories to be re-reported in mainstream media. This means that, unless working within a partnership with Facebook, fact-checkers have little chance of combating “fake news” directly at its source.

Self-Regulation by Platforms

Public and political pressure have stimulated platforms to combat disinformation by means of individual self-organization and within the framework of EU Code of Practice on Disinformation. Overall, self-regulation by the three major platforms in Europe, Facebook, Twitter and YouTube, is shaped by a reluctance to ban “fake news” outright. Instead, platforms prefer to sanction certain kinds of behaviors, and, in the case of Facebook, provide less visibility to disinformation, while allowing it to stay online. Platforms work with third-party fact checkers or human moderators to identify disinformation content to de-prioritise, but not to ban it under content moderation rules (Caplan, Hanson, and Donovan 2018). Two key areas in which platforms are willing to take stronger action are electoral disinformation and medical disinformation, which has been thrown into stark relief by the Covid-19 crisis. In more detail, self-regulation can be analyzed at the levels of policy, implementation, and oversight.

Platform policies: Due to concerns about free speech and liability for content, platforms are reluctant to ban disinformation outright. Instead, they pursue policy responses that can be identified in three broad tranches. First, platforms prefer to prohibit certain suspicious-looking user behaviors, which could be “manipulative” or “inauthentic” (Gleicher 2018; Twitter 2019). This covers a range of practices including malicious use of bots or impersonation of others (Twitter 2019; Google 2019). To give an impression of scale, Facebook took down 2.19 billion fake accounts in the first quarter of 2019 alone (Facebook Inc 2019). Second, platforms make problematic

content less visible, without banning it completely. Facebook and YouTube call this “borderline content” and have changed their algorithms so that it will be recommended less often (Zuckerberg 2018a; YouTube 2019). Third, platforms take special measures during major elections (e.g. Zuckerberg 2018b). For example, Facebook places restrictions on political adverts, such as requiring purchasers to verify their identity. This includes one of the two exceptions to the reluctance to ban disinformation: all three platforms ban election disinformation, such as false or misleading information about when, where and how to vote. While Twitter and YouTube only do this for specific election seasons, Facebook has implemented this full-time. Similarly, major platforms remove medical disinformation that could lead to physical harm, which they have enforced during the Covid-19 crisis. Disinformation about the disease not only fulfills the criterion of threatening physical harm, it is also an area in which, according to Mark Zuckerberg, it is “easier to set policies that are a little more black and white and take a much harder line,” not least because the World Health Organization is available to set an international standard for reliable information (Smith 2020).

Implementation: To implement these policies, companies rely on a combination of artificial intelligence and human moderation to identify and mitigate problematic content and behaviors. As a simple response, the messaging app WhatsApp has introduced a limit on the spread of information by preventing users from forwarding a message to more than five of their contacts at a time. As a response to the spread of Covid-19 disinformation, WhatsApp added more friction by allowing users to forward “highly forwarded messages” (that have already been forwarded more than five times) to only one person at a time (Newton 2020). In a more complex response, Facebook reduces the spread of disinformation through “downranking.” Facebook uses technology to identify quantitative indicators that suggest potential “fake news,” such as if a link is not often shared after being read (Mosseri 2017). It also allows users to report “fake news.” These flagged posts are then sent to professional fact-checking organizations for verification. If marked as false, a post will not be deleted, but become less visible on Facebook, for example by being shown lower down on News Feed (Zigmond 2018). To a certain extent, this can be viewed as an appropriate solution for how to identify disinformation without giving platforms extra journalistic responsibilities of distinguishing fact from fiction. However, introducing the human element of fact-checkers also has costs in terms of speed and scalability. The fact-checking process takes three days on average, by which point, most stories have already spread far (Zigmond 2018). Fact-checkers do not operate globally, but nationally, so Facebook must find fact-checking partners on a country-by-country basis. Currently, only half of EU countries are covered by Facebook fact-checking partnerships.

A second important feature of implementation is that platforms have taken special measures in particularly high-risk situations, such as the European parliamentary elections and the Covid-19 crisis. Facebook set up a temporary “war room” for the EU elections, similar to the one for the US mid-terms (Graham-Harrison 2019). Twitter also set up a dedicated complaints tool for reporting voter disinformation specifically for the EU elections (Twitter 2019). Each platform also had a political ads library ready in Europe prior to the European election in 2019, showing ads run by authorized advertisers. The Covid-19 crisis underlines that platforms are prepared to take much

stronger action against disinformation under exceptional circumstances. Both Facebook and Twitter deleted posts from Heads of State (BBC 2020), while YouTube has taken the step of banning content that promotes disinformation connecting Covid-19 and 5G telecommunication networks (Kelion 2020). Furthermore, Facebook has taken the additional step of showing a link to WHO information to users who “have liked, reacted or commented on harmful misinformation about Covid-19” (Rosen 2020).

Third, platforms have also introduced measures to help users make better judgments about news and guide them toward more authoritative sources. Facebook has introduced a context button to links, which shows Wikipedia information about the publisher and where the link has been shared and by how many users. YouTube has also introduced the “breaking news shelf,” which automatically highlights news videos from authoritative sources during major news events. The feature was available in nine countries prior to the EU elections, with roll-outs planned in other European countries (Google 2019).

Settlement of disputes: While major platforms name freedom of expression as a value that they seek to uphold, nuanced mechanisms aimed at balancing the right to free expression with the urgency of combating disinformation are lacking. Reducing algorithmic recommendations of content on YouTube and Facebook preserves freedom of speech without guaranteeing “freedom of reach” (DiResta 2018). However, downranking also removes potential disputes from appeals mechanisms that are available to users when content has been removed. On Facebook, this also means that such cases would also escape the attention of the planned Oversight Board, which will make decisions on complex cases (Douek 2019). At the same time, Facebook exempts politicians and political advertising from fact-checking outright, on the grounds that it is important for users to scrutinize what politicians are saying. These examples indicate that while there are attempts, albeit controversial ones, to assert other interests amidst the fight against disinformation, this is represented at the level of policy-setting and implementation, but lacking in procedural mechanisms that would weigh interests and resolve disputes.

Oversight: All three major platforms publish their own transparency reports on content moderation and submit publicly available implementation reports on efforts against disinformation to the European Commission in the framework of the EU Code of Practice on Disinformation. This can be considered ‘soft’ oversight aimed at improved transparency, but does not involve consequences such as institutional sanctions. However, while reporting provides information about which measures platform companies are taking, the effectiveness of these measures remain a dark area, especially as media reports and academic studies about the scale of online disinformation seem to contradict each other.

Summary and Reflection on Implications for Accountability

Online disinformation poses a major challenge to democratic societies that depend on public debate and well-informed citizens who express their free will in political processes. Evolving social, economic and technical conditions are promoting the spread of

disinformation. The intensification of social conflicts, political polarization and antagonism is fertile ground for polarizing, partisan or even misleading information. At the same time, “structural and economic changes in the news media, increased fragmentation and personalization, and algorithmically dictated content dissemination and consumption, affect the production and flow of news and information in ways that may make it more difficult to assume that legitimate news will systematically win out over misinformation” (Napoli 2019, 82). It has therefore become a prominent issue on the scientific and political agenda.

While many analyses focus on trends in subject matter, patterns of distribution and the reach of disinformation on the Internet, this article contributes to the analysis of strategies to counter disinformation. Employing a governance perspective, it analyzes the emerging mix of governance responses to disinformation in the European system of multilevel governance and on the continuum between market and state (Latzer et al. 2003). It provides an overview of EU initiatives, national legislation and enforcement in selected European countries, as well as efforts by platforms and third-party responses to disinformation. In combination, these analyses show the distribution of accountability in an emerging accountability network and enable reflection on accountability problems. The analysis and conclusions can be summarized as follows:

Disinformation as a socio-technical assemblage: Disinformation is a complex problem that does not allow for a simple one-size-fits all governance solution. Its proliferation on social media has developed from a socio-technical mix, in the form of a combination of platform design, algorithms, human factors and political and commercial incentives. It involves social media companies, which determine the policies and technological design of platforms, “fake news publishers” who use these systems to disseminate disinformation and adapt strategies to changes, and users who fuel distribution by clicking, liking, sharing and watching. Actors and technologies involved point to potential targets for governance, as well as groups of actors to whom accountability can be assigned.

Central role of Internet platforms: Among the many layers involved, Internet platforms play a central role in distribution and reduction of disinformation. These platforms are reluctant to ban disinformation outright, preferring instead to act against suspicious behavior, and to reduce the visibility of disinformation content. Policies are implemented through processes supported both by humans and machines. In particular, Facebook’s experience with third-party fact-checkers highlights how humans can help improve quality, but face limitations in terms of scope and scale. Meanwhile, platforms pay particular attention to disinformation during elections by introducing policies and tools specifically for major election periods. A successful approach to disinformation must take into account the central role of Internet platforms in distribution and reduction of disinformation.

Different policy responses across European countries: Disinformation has become a prominent issue on the political agenda. Political awareness of disinformation started to gain traction with the Ukraine conflict and achieved further prominence during the US presidential election. Awareness peaks during important election periods, sometimes leading to regulatory responses, as evidenced in individual countries in Europe. In France, social media platforms must comply with court orders to delete fake news

during elections, while in Germany, there is no explicit ban on disinformation, but the Network Enforcement Act applies to disinformation that constitutes criminal offenses such as insult, defamation and hate speech. The United Kingdom plans to set up a regulator, establish a duty of care and oblige platforms to commit to combating disinformation. Governance responses vary from country to country, and most of the measures taken are rather recent, so it is too early to conduct comprehensive performance assessments and draw conclusions on how governance can best move forward. However, national governance responses lead to stronger regulation of internet platforms and a weakening of long-established liability protections and demonstrate that individual European countries are taking individual steps beyond pan European approaches.

Multiple initiatives of the European Commission: In the European system of multi-level governance, the European Commission looms large. Ahead of elections to the European Parliament, the Commission attempted to combat disinformation with a broad approach and multiple initiatives. These included financial and coordinative instruments such as establishing and funding infrastructures and organizations (East Stratcom Task Force, European Network of Fact-Checkers, early warning system), as well as projects and awareness-raising campaigns within existing programs (e.g., Connecting Europe, #SaferInternet4EU). Furthermore, the Commission has established the EU Code of Practice on Disinformation, which puts additional pressure on platforms to take action and provide some level of transparency.

Emergence of an accountability network: Altogether, the analysis demonstrates that different types of actors and institutions are involved in the spread of disinformation and the implementation of governance reactions. Together these actors and institutions have competencies and capacities to act against the spread of disinformation and therefore can be described as an accountability network. These groups of actors have the potential to adopt a variety of distinct accountability concepts, comprising legal liability (fake news publishers, platforms), self-responsibility (users), accountability by design (technologies, algorithms), as well as corporate social responsibility (companies, platforms). Moreover, the accountability network involves political meta accountability (the state), which comprises the duty of the government to ensure an adequate regulatory framework to protect freedom of speech, support pluralism and media quality, to mitigate risks like disinformation and to allocate and assign rights and duties with regard to accountability to particular types of actors.

Moderate investments in journalism and fact-checking: A multi-dimensional approach to disinformation that takes into account all groups in the accountability network should be able provide resources, incentives and sanctions for involved actors and institutions to fulfill their responsibilities adequately. This includes, for instance, public efforts and investments for the promotion of public awareness of the problem of disinformation and educational initiatives to increase media literacy and critical media reception. Moreover, discovering disinformation and counteracting it are essential for a healthy information ecosystem. However, current socio-economic trends endanger institutions that are equipped to perform fact-checking functions systematically. Scientific research, quality journalism and fact-checking organizations are all threatened by the lack of sustainable business models and revenues. Consequently, from a

public interest perspective, the OSCE (2017) considers support for fact-checking entities and support for a free, independent and diverse communications environment key means of addressing disinformation and propaganda. In practice, however, so far steps taken by the EU have drawn criticism. Supportive measures for journalism remain comparatively weak and projects promoting investigative journalism and quality news have been funded only on a modest scale. The EU has admitted that its independent network of fact-checkers has not achieved adequate geographical coverage and capacity for analyzing sources and patterns in disinformation (European Commission 2018c, 7).

Light-touch oversight with Code of Practice: A multidimensional approach to disinformation must consider the central role of Internet platforms in the distribution of disinformation. The European Commission has decided to hold platforms to greater account by means of a voluntary code of practice. The code presents a set of commitments that allocates accountability within the social media industry by setting expectations in combination with reporting requirements and light-touch oversight. However, the code has been criticized for lacking a common approach, measurable objectives, KPIs, meaningful commitments and enforcement tools (Sounding Board of the Multistakeholder Forum on Disinformation Online 2018). On a broader scale, platform companies may have attempted to influence the preparatory work for the code to ward off measures that “would have allowed the EU competition commissioner to examine the platforms’ business models to see whether they helped misinformation to spread” (Schmidt and Dupont-Nivet 2019). Furthermore, while reporting provides information about which measures platform companies are taking, the effectiveness of these measures remains a dark area, especially as media reports and academic studies about the scale of online disinformation seem to contradict each other. To achieve a broad and accurate view of the distribution and effects of disinformation, there is a need for better access, more transparency and more in-depth cooperation between platforms and academic research in the analysis of disinformation.

Danger of confusion around accountability: A multi-dimensional approach to disinformation should also clearly allocate accountability in a shared, distributed and cooperative structure. The variety of involved parties and the distribution of action point to a structure of distributed accountability in the domain of disinformation. This variety risks confusion around accountability and shirking of responsibility. In practice, the distribution of accountability as given by the EU Code of Practice on Disinformation between public authorities and private platforms has led to criticism and uncertainty.

Danger of bypassing judicial review: First, as an instrument of self-regulation, the voluntary Code of Practice bypasses review of potential violations of fundamental rights by the constitutional courts (Rudl 2018). This fuels the fear that private enforcement measures to combat disinformation do not sufficiently weight freedom of communication on the one hand and protection of public discourse from misleading information on the other. While the communication on Tackling Online Disinformation and the Code of Practice mentions the need to balance the fight against disinformation with the fundamental right to freedom of expression and an open internet, they do not put forward measures to balance and manage these interests.

Uncertainty regarding liability protections: Second, with regard to the governance architecture, the relationship between the Code of Practice on Disinformation and the e-commerce Directive remains unclear. The e-commerce Directive protects platforms from liability for unlawful content posted on their platforms as long as they do not have knowledge of it. However, the Code of Practice on Disinformation sets down specific commitments for platforms with regard to content moderation. Since moderation leads to knowledge of content, platforms could be held to account and lose their liability privilege. Besides providing an incentive for platforms to either not moderate at all or to remove more content than necessary, this leads to uncertainty about accountability (see e.g., European Commission 2016, 9). Leaked plans for a “Digital Services Act” that would replace the e-Commerce Directive suggest that the EU intends to continue to uphold liability protections for platforms, but would seek to counter platform reluctance to monitor content by specifying that platforms that enact proactive monitoring measures are not liable (Fanta and Rudl 2019).

Conditional immunity to counter procedural shortcomings: Finally, policies and operations of Internet platforms have been subject to criticism in terms of accountability. Critiques point to a lack of clear, transparent and non-discriminatory definitions and standards for dealing with disinformation, the opacity of algorithmic selection in advertising and news feeds that may promote disinformation, a lack of due process and options for appealing automated and manual content moderation decisions and the reluctance of platforms to demonstrate genuine openness and cooperation in independent evaluation processes. These issues point to *procedural shortcomings* that are partly addressed by the voluntary Code of Practice, but also mentioned in analyses of platform content policies beyond disinformation. If self-regulatory approaches to mitigate procedural shortcomings fail, governments could “use the offer of intermediary immunity as a lever to get social media companies to engage in public-regarding behaviour” (Balkin 2019, 21). With regard to future policy and allocation of accountability, a possible path forward would be to make adherence to *procedural requirements* – such as standards definition, due process, transparency, evaluation and oversight – a condition in order for platforms to further enjoy immunity from liability for third party content.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The paper presents selected results from the project “The automation of the social: Algorithmic selection in social online networks” funded by the Vienna Anniversary Fund for the Austrian Academy of Sciences.

ORCID

Florian Saurwein  <http://orcid.org/0000-0001-8018-6439>

References

- Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–236.
- Apuzzo, Matt, and Adam Satariano. 2019. "Russia Is Targeting Europe's Elections. So Are Far-Right Copycats." <https://www.nytimes.com/2019/05/12/world/europe/russian-propaganda-influence-campaign-european-elections-far-right.html>
- Avaaz. 2019. "Far Right Networks of Deception." <https://avaazimages.avaaz.org/Avaaz%20Report%20Network%20Deception%2020190522.pdf>
- Bakir, Vian, and Andrew McStay. 2018. "Fake News and the Economy of Emotions: Problems, Causes, Solutions." *Digital Journalism* 6 (2): 154–174.
- Balkin, Jack. 2019. "How to Regulate (and Not Regulate) Social Media." Keynote, Association for Computing Machinery Symposium on Computer Science and Law, New York City, October 28, 2019.
- Bartle, Ian, and Peter Vass. 2005. *Self-Regulation and the Regulatory State*. Bath: University of Bath.
- BBC. 2020. "Coronavirus: World Leaders' Posts Deleted Over Fake News." <https://www.bbc.com/news/technology-52106321>
- Black, Julia. 2010. "Risk-Based Regulation: Choices, Practices and Lessons Learnt." In *Risk and Regulatory Policy: Improving the Governance of Risk*, edited by OECD, 185–224. Paris: OECD Publishing.
- Broderick, Ryan. 2019. "Here's Everything the Mueller Report Says About How Russian Trolls Used Social Media." <https://www.buzzfeednews.com/article/ryanhatesthis/mueller-report-inter-net-research-agency-detailed-2016>
- Brummette, John, Marcia DiStaso, Michael Vafeiadis, and Marcus Messner. 2018. "Read All about It: The Politicization of 'Fake News' on Twitter." *Journalism & Mass Communication Quarterly* 95 (2): 497–517.
- Bundesjustizamt. 2018. "Netzwerkdurchsetzungsgesetz – Häufig gestellte Fragen." https://www.bundesjustizamt.de/DE/Themen/Buergerdienste/NetzDG/Fragen/FAQ_node.html#faq10018922
- Caplan, Robyn, Lauren Hanson, and Joan Donovan. 2018. *Dead Reckoning: Navigating Content Moderation after 'Fake News'*. New York: Data & Society.
- Cichowlas, Ola. 2017. "Russia Launches Operation 'Anyone but Macron'." <https://www.themoscowtimes.com/2017/02/09/russia-launches-operation-anyone-but-macron-a57104>
- Delcker, Janosch, Zosia Wanat, and Mark Scott. 2020. "The Coronavirus Fake News Pandemic Sweeping Whatsapp." <https://www.politico.com/news/2020/03/16/coronavirus-fake-news-pandemic-133447>
- Desmaris, Sacha, Pierre Dubreuil, and Benoit Loutrel. 2019. "Creating a French Framework to Make Social Media Platforms More Accountable." <http://thecre.com/RegSM/wp-content/uploads/2019/05/French-Framework-for-Social-Media-Platforms.pdf>
- DiResta, Renee. 2018. "Free Speech Is Not the Same as Free Reach." <https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/>
- Douek, Evelyn. 2019. "Facebook's 'Draft Charter' for Content Moderation: Vague, But Promising." <https://www.lawfareblog.com/facebooks-draft-charter-content-moderation-vague-promising>
- Engel, Christoph. 2004. "A Constitutional Framework for Private Governance." *German Law Journal* 5 (3): 197–237.
- European Commission. 2016. "Communication from the Commission on Online Platforms and the Digital Single Market Opportunities and Challenges for Europe." COM(2016)288, May 25.
- European Commission. 2018a. "Fake News and Online Disinformation." <https://ec.europa.eu/digital-single-market/en/fake-news-disinformation>
- European Commission. 2018b. "Communication from the Commission 'Tackling online disinformation: A European Approach.'" COM(2018)236, April 26.
- European Commission. 2018c. "Report from the Commission on the implementation of the Communication 'Tackling Online Disinformation: A European Approach.'" COM(2018)794, December 5.
- European Commission. 2019. "Code of Practice against Disinformation." Statement, May 17.

- European Commission and High Representative of the Union for Foreign Affairs and Security Policy. 2018. "Action Plan against Disinformation." JOIN(2018)36, December 5.
- European Council. 2015. "European Council Meeting (19 and 20 March 2015) – Conclusions." <https://www.consilium.europa.eu/media/21888/european-council-conclusions-19-20-march-2015-en.pdf>
- Facebook Inc. 2019. "Facebook Press Call, May 23." <https://fbnewsroomus.files.wordpress.com/2019/05/cser-press-call-5.23.19.pdf>
- Fanta, Alexander, and Tomas Rudl. 2019. "Leaked Document: EU Commission Mulls New Law to Regulate Online Platforms." <https://netzpolitik.org/2019/leaked-document-eu-commission-mulls-new-law-to-regulate-online-platforms/>
- Field, Matthew, and Mike Wright. 2018. "Russian Trolls Sent Thousands of Pro-Leave Messages on Day of Brexit Referendum, Twitter Data Reveals." <https://www.telegraph.co.uk/technology/2018/10/17/russian-iranian-twitter-trolls-sent-10-million-tweets-fake-news/>
- Fiorentino, Michael-Ross. 2018. "France Passes Controversial 'Fake News' Law." <https://www.euro-news.com/2018/11/22/france-passes-controversial-fake-news-law>
- Fletcher, Richard, Alessio Cornia, Lucas Graves, and Rasmus Kleis Nielsen. 2018. "Measuring the Reach of 'Fake News' and Online Disinformation in Europe." <https://reutersinstitute.politics.ox.ac.uk/our-research/measuring-reach-fake-news-and-online-disinformation-europe>
- Frenkel, Sheera, Nicholas Casey, and Paul Mozur. 2018. "In Some Countries, Facebook's Fiddling Has Magnified Fake News." <https://www.nytimes.com/2018/01/14/technology/facebook-news-feed-changes.html>
- Fürstenau, Marcel. 2017. "Fears of Fake News Overshadowed Its Effect – In Germany." <https://www.dw.com/en/fears-of-fake-news-overshadowed-its-effect-in-germany/a-41132306>
- Gerdziunas, Benas. 2017. "Baltics Battle Russia in Online Disinformation War." <https://www.dw.com/en/baltics-battle-russia-in-online-disinformation-war/a-40828834>
- Gerdziunas, Benas. 2018. "Lithuania Hits Back at Russian Disinformation." <https://www.dw.com/en/lithuania-hits-back-at-russian-disinformation/a-45644080>
- Gerdziunas, Benas. 2019. "Lithuania Set to Ban Fake News from Russia." <https://www.dw.com/en/lithuania-set-to-ban-fake-news-from-russia/a-47409350>
- Gleicher, Nathaniel. 2018. "Coordinated Inauthentic Behavior Explained." <https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/>
- Google. 2019. *EC Action Plan on Disinformation: Google April 2019 Report*. Brussels: European Commission.
- Graham-Harrison, Emma. 2019. "Inside Facebook's War Room: The Battle to Protect EU Elections." <https://www.theguardian.com/technology/2019/may/05/facebook-admits-huge-scale-of-fake-news-and-election-interference>
- Graves, Lucas, and Federica Cherubini. 2016. *The Rise of Fact-Checking Sites in Europe*. Oxford: Reuters Institute for the Study of Journalism.
- Helberger, Natali, Jo Pierson, and Thomas Poell. 2018. "Governing Online Platforms: From Contested to Cooperative Responsibility." *The Information Society* 34 (1): 1–14.
- HLEG (High Level Group on Fake News and Online Disinformation). 2018. *A Multi-Dimensional Approach to Disinformation*. Luxembourg: European Commission.
- HM Government. 2019. *Online Harms White Paper*. United Kingdom: Department for Digital, Culture, Media & Sport.
- Hooghe, Liesbet. ed. 1996. *Cohesion Policy and European Integration: Building Multi-Level Governance*. Oxford: OUP.
- House of Commons. 2019. *Disinformation and "Fake News": Final Report*. HC 1791. London: House of Commons.
- Kayali, Laura, Rym Momtaz, and Nicholas Vincour. 2019. "Emmanuel Macron's Plan to Fix Facebook, YouTube and Twitter." <https://www.politico.eu/article/emmanuel-macrons-plan-to-fix-facebook-youtube-and-twitter/>
- Kelion, Leo. 2020. "Youtube Tightens Covid-19 Rules after Icke Interview." <https://www.bbc.com/news/technology-52198946>

- Krasodonski-Jones, Alex, Josh Smith, Elliot Jones, Ellen Judson, and Carl Miller. 2019. *Warring Songs: Information Operations in the Digital Age*. London: Demos.
- Krempf, Stefan. 2018. "Erste Löscherichte bestätigen Gefahr von Overblocking." <https://www.golem.de/news/netzdg-kritiker-erste-loeschberichte-bestaetigen-gefahr-von-overblocking-1807-135758.html>
- Latzer, Michael, Natascha Just, Florian Saurwein, and Peter Slominski. 2003. "Regulation Remixed: Institutional Change through Self- and Co-Regulation in the Mediamatics Sector." *Communications and Strategies* 50 (2): 127–157.
- Lewis, Paul. 2018. "'Fiction Is Outperforming Reality': How YouTube's Algorithm Distorts Truth." <https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>
- Lewis, Seth C., and Oscar Westlund. 2015. "Actors, Actants, Audiences, and Activities in Cross-Media News Work." *Digital Journalism* 3 (1): 19–37.
- Meta, M. K. 2019. "Ivan Georgiev: The Fight against Disinformation in Bulgaria Is almost Nonexistent." <https://meta.mk/en/ivan-georgiev-the-fight-against-disinformation-in-bulgaria-is-almost-nonexistent/>
- Manjoo, Farhad. 2017. "How Twitter Is Being Gamed to Feed Misinformation." <https://www.nytimes.com/2017/05/31/technology/how-twitter-is-being-gamed-to-feed-misinformation.html>
- Marchal, Nahema, Bece Kollanyi, Lisa-Maria Neudert, and Philip N. Howard. 2019. *Junk News during the EU Parliamentary Elections; Lessons from a Seven-Language Study of Twitter and Facebook*. Oxford: Oxford Internet Institute.
- Marwick, Alice, and Rebecca Lewis. 2018. *Media Manipulation and Disinformation Online*. New York: Data & Society.
- Mejias, Ulises A., and Nikolai E. Vokuev. 2017. "Disinformation and the Media: The Case of Russia and Ukraine." *Media, Culture & Society* 39 (7): 1027–1042.
- Miller, Catherine, Jacob Ohrvik-Stott, and Rachel Coldicutt. 2018. *Regulating for Responsible Technology: Capacity, Evidence and Redress: A New System for a Fairer Future*. London: Doteveryone.
- Mosseri, Adam. 2017. "Working to Stop Misinformation and False News." <https://www.facebook.com/facebookmedia/blog/working-to-stop-misinformation-and-false-news>
- Napoli, Philip M. 2019. *Social Media and the Public Interest*. New York: Columbia University Press.
- Neuhäuser, Christian. 2014. "Roboter und moralische Verantwortung." In *Robotik im Kontext von Recht und Moral*, edited by Eric Hilgendorf, 269–286. Baden-Baden: Nomos.
- Newton, Casey. 2020. "How Whatsapp Is Making It More Expensive to Spread Misinformation." <https://www.theverge.com/interface/2020/4/8/21212110/whatsapp-forward-limit-encryption-apple-imessage-signal>
- Noack, Rick. 2018. "Everything We Know So Far about Russian Election Meddling in Europe." https://www.washingtonpost.com/news/worldviews/wp/2018/01/10/everything-we-know-so-far-about-russian-election-meddling-in-europe/?noredirect=on&utm_term=.7144f9b5d9b5
- O'Halloran, Marie. 2017. "Government Defeated on Online Advertising and Social Media Bill." <https://www.irishtimes.com/news/politics/oireachtas/government-defeated-on-online-advertising-and-social-media-bill-1.3327979>
- OSCE. 2017. *Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda*. Vienna: OSCE.
- Osipova, Natlia V., and Aaron Byrd. (2017). "Inside Russia's Network of Bots and Trolls." <https://www.nytimes.com/video/us/politics/100000005414346/how-russian-bots-and-trolls-invade-our-lives-and-elections.html>
- Parlament der Republik Österreich. 2019. Maßnahmen gegen Desinformation (2760/AB)." https://www.parlament.gv.at/PAKT/VHG/XXVI/AB/AB_02760/index.shtml
- Puppis, Manuel. 2010. "Media Governance: A New Concept for the Analysis of Media Policy and Regulation." *Communication, Culture & Critique* 3 (2): 134–149.
- Rhodes, Roderick Arthur William. 1996. "The New Governance: Governing without Government." *Political Studies* 44 (4): 652–667.

- Rosen, Guy. 2020. "An Update on Our Work to Keep People Informed and Limit Misinformation about COVID-19 – About Facebook." <https://about.fb.com/news/2020/04/covid-19-misinfo-update/>
- Rosenau, James N., and Ernst Otto Czempiel, eds. 1992. *Governance without Government: Order and Change in World Politics*. Cambridge: CUP.
- Ross, Andrew S., and Damian J. Rivers. 2018. "Discursive Deflection: Accusation of "Fake News" and the Spread of Mis- and Disinformation in the Tweets of President Trump." *Social Media + Society* 4 (2): 205630511877601.
- Rudl, Tomas. 2018. "EU-Kommission will von Plattformen "freiwillige" und weitreichende Internetzensur." <https://netzpolitik.org/2018/eu-kommission-will-von-plattformen-freiwillige-und-weitreichende-internetzensur/>
- Sängerlaub, Alexander. 2017. *Deutschland vor der Bundestagswahl: Überall Fake News?! Berlin: Stiftung Neue Verantwortung*.
- Sängerlaub, Alexander, Miriam Meier, and Wolf-Dieter Rühl. 2018. *Faktenstatt Fakes*. Berlin: Stiftung Neue Verantwortung.
- Saurwein, Florian. 2019. "Emerging Structures of Control for Algorithms on the Internet: Distributed Agency – Distributed Accountability." In *Media Accountability in the Era of Post-Truth Politics. European Challenges and Perspectives*, edited by Tobias Eberwein, Susanne Fengler, and Matthias Karmasin, 196–211. London: Routledge.
- Schmidt, Nico, and Daphné Dupont-Nivet. 2019. "Facebook and Google Pressured EU Experts to Soften Fake News Regulations, Say Insiders." <https://www.opendemocracy.net/en/facebook-and-google-pressured-eu-experts-soften-fake-news-regulations-say-insiders/>
- Schultheis, Emily. 2017. "The Czech Republic's Fake News Problem." <https://www.theatlantic.com/international/archive/2017/10/fake-news-in-the-czech-republic/543591/>
- Schuster, Simon, and Sandra Irimova. 2018. "A Former Russian Troll Explains How to Spread Fake News." <http://time.com/5168202/russia-troll-internet-research-agency/>
- Scott, Mark. 2019. "Half of European Voters May Have Viewed Russian-Backed 'Fake News'." <https://www.politico.eu/article/european-parliament-russia-mcafee-safeguard-cyber/>
- Smith, Ben. 2020. "When Facebook Is More Trustworthy than the President." <https://www.nytimes.com/2020/03/15/business/media/coronavirus-facebook-twitter-social-media.html>
- Snegovaya, Maria. 2018. "Russian Propaganda in Germany: More Effective than You Think." <https://www.the-american-interest.com/2017/10/17/russian-propaganda-germany-effective-think/>
- Sombetzki, Janina. 2017. "Verantwortung und Roboterethik – ein kleiner Überblick." *Humboldt Forum Recht* 3: 10–30.
- Sounding Board of the Multistakeholder Forum on Disinformation Online. 2018. "The Sounding Board's Unanimous Final Opinion on the So-Called Code of Practice." <https://www.euractiv.com/wp-content/uploads/sites/2/2018/10/3OpinionoftheSoundingboard-1.pdf>
- Subramanian, Samanth. 2017. "Meet the Macedonian Teens Who Mastered Fake News and Corrupted the US Election." <https://www.wired.com/2017/02/veles-macedonia-fake-news/>
- Tandoc, Edson, Zheng Wei Lim, and Rich Ling. 2018. "Defining "Fake News." *Digital Journalism* 6 (2): 137–153.
- Thompson, Nicholas, and Fred Vogelstein. 2018. "Inside Facebook's Two Years of Hell." <https://www.wired.com/story/inside-facebook-mark-zuckerberg-2-years-of-hell/>
- Twitter. 2019. *Twitter April Update: Code of Practice on Disinformation*. Brussels: European Commission.
- Verza, Sophia. 2018. "An Overview of Italian Online and Offline Political Communication Regulation." https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2018/02/policy-brief_def1.pdf
- Vosoughi, Sorousch, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–1151.
- Walker, Shaun. 2020. "Hungary to Consider Bill That Would Allow Orbán to Rule by Decree." <https://www.theguardian.com/world/2020/mar/23/hungary-to-consider-bill-that-would-allow-orban-to-rule-by-decree>

- Wardle, Clare, and Hossein Derakhshan. 2017. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. Strasbourg: Council of Europe.
- Woolley, Samuel C., and Philip N. Howard. 2016. "Political Communication, Computational Propaganda, and Autonomous Agents." *International Journal of Communication* 10: 4882–4890.
- Woolley, Samuel C., and Philip N. Howard. 2018. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. New York: OUP.
- YouTube. 2019. "Continuing Our Work to Improve Recommendations on YouTube." <https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html>
- Zeit Online. 2019. "Frankreich beschließt Gesetz gegen Fake-News." <https://www.zeit.de/politik/ausland/2018-11/frankreich-pressefreiheit-fake-news>
- Zigmond, Dan. 2018. "Machine Learning, Fact-Checkers and the Fight against False News." <https://newsroom.fb.com/news/2018/04/inside-feed-misinformation-zigmond/>
- Zuckerberg, Mark. 2018a. "A Blueprint for Content Governance and Enforcement." <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>
- Zuckerberg, Mark. 2018b. "Preparing for Elections." <https://www.facebook.com/notes/mark-zuckerberg/preparing-for-elections/10156300047606634/>