ENABLING DIGITAL
RIGHTS AND GOVERNANCE

Auditing Big Tech: Combating Disinformation with Reliable Transparency

Ben Wagner

Johanne Kübler

Lubos Kuklis

Carolina Ferro

February 2021

Table of Contents

Executive summary

Disinformation on social media platforms has become a major concern during elections and the COVID-19 pandemic. Under increased governmental and societal pressure, social media platforms attempt to limit the spread of disinformation. However, their exact procedures to identify and regulate the spread of disinformation on their platforms and whether they are successful or not in this endeavor are currently impossible to verify. To counter the lack of transparency of social media platforms and give governments and regulators an independent way to assess these efforts, we propose creating 'independent auditing intermediaries.' An auditing mechanism of this kind would ensure the accuracy of 'transparency data' provided by large online platform providers about the content on their services.

Transparency data that have been audited would be considered 'verified data' in this context. Such an institution could be created within the context of the proposed European Digital Services Act (DSA), but it should be a distinct legal entity to guarantee its independence. This institution would be responsible for collecting and verifying data, producing verified data and making them available only to authorities endowed with the legal competence to use them, to a legally specified extent for a legally-specified purpose. The collection and verification of the data on the one hand (responsibility of the independent auditing intermediaries), and their use for regulatory purposes on the other (obligation of the regulation authorities), would be distinct processes. This model would further enhance the independence of the institutions involved and guarantee the security of the data in question.

Acknowledgments

This policy paper was commissioned by Omidyar Network and benefited from reviews and inputs by external experts. The authors wish to thank the Omidyar team, especially Agustin Rossi, without whom this paper would not have been possible. We would also like to acknowledge the valuable contribution of Mathias Vermeulen and Wafa Ben-Hassine to a previous version of this paper.

1. Introduction

In recent years, social media companies have come under scrutiny due to the increased circulation of disinformation on their platforms. With its reported 1.79 billion daily users in June 2020, Facebook has become the main arena for disinformation during election campaigns worldwide. In 2019, the Mueller report investigating the potential Russian interference in the 2016 US Presidential election outlined the extensive efforts by Russian actors to manipulate the electorate using Facebook (Mueller 2019). A year earlier, the social network faced significant pressure after it emerged that Myanmar's military used it to incite violence against the Rohingya minority beginning in 2017 (Mozur 2018). WhatsApp, the messaging app owned by Facebook, was used as a major channel for rumors, doctored images and fake news stories in the run-up of the recent elections in Brazil and India (Funke 2018; Murgia, Findlay, and Schipani 2019). Meanwhile, an allegedly doctored video on Google's video platform YouTube sparked an ultimately unsuccessful coup in Gabon, highlighting how disinformation could wreak the most havoc in developing countries, often home to fragile governments and populations with nascent digital literacy (Breland 2019). Finally, Twitter faces calls to reign in the United States president, Donald J. Trump, who uses the platform to bypass the media and spread falsehoods and inflammatory messages.

Disinformation is regarded as a major concern during the 2020 US election campaign, as is misinformation circulating on social media about COVID-19. The latter is a source of political concern during the current global health crisis when the circulation of misleading information can be particularly dangerous. Disinformation about a public health emergency of this scale, which could contribute to the demise of hundreds of thousands of people, elevates the question of how to deal with it from a primarily political matter to one of life and death, forcing governments and social media platforms into action. However, how these companies identify and regulate the spread of disinformation on their platforms and whether they are successful or not in this endeavor is, currently, impossible to verify for governments and the public at large.

In light of the considerable amount of harmful online content, European governments and the European Union are rushing to introduce measures to fight the spread of false information, ranging from government task forces and media literacy campaigns to legislation proposals. The following paper offers an alternative approach. It provides an overview of the problems associated with the lack of transparency of social media platforms' efforts to reduce the spread of disinformation, before proposing a response in the form of auditing intermediaries. After discussing which problems the concept of auditing intermediaries is designed to address, it will then discuss some of the main challenges associated with access to data, the potential misuse of intermediaries, and the general lack of data provision standards by large online platforms. In conclusion, the article suggests an urgent need for an auditing mechanism to ensure the accuracy of 'transparency data' provided by large online platform providers about

the content on their services. Transparency data that have been audited would be considered 'verified data' in this context. Without such a transparency verification mechanism, existing public debate is based merely on a whim, and digital dominance is likely only to become more pronounced.

2. What is the current state of the art in disinformation globally?

Disinformation can be defined as the dissemination of false information with the aim of influencing public opinion, groups, or individuals in the interests of political or economic interests. Contrary to misinformation, whose inaccuracies are unintended, disinformation is deliberately false information spread intentionally. "Fake news," on the other hand, is a political expression used to criticize a news story or media outlet.

Misinformation, disinformation and propaganda tactics date as far back as Ancient Rome and was commonly used by intelligence services and during conflicts, notably in the Cold War period (Goldman 2006:43). Recent developments in technology and media have amplified the impact of misleading information on modern international politics. The move towards internet-based communication, especially social network platforms, enables politicians and other actors to engage with a global audience directly. Traditional quality media outlets are struggling amid the competition with digital media and other traditional media outlets and the 24-hour news cycle, resulting in the gradual erosion of editorial standards and increasingly sensationalized news coverage.

The demise of the traditional gatekeepers is accompanied by changing media consumption habits of the population. According to a Pew Research Center survey of eight Western European countries conducted in late 2017, television remains the most important news source. However, online news consumption comes second, and in two countries, Sweden and Denmark, it even surpasses television as the primary news source (Matsa 2018).

As concerns about the increased salience of disinformation rose, social media companies have sought to alleviate the situation in numerous ways. Whereas social media companies initially denied holding any accountability over the content published on their platforms, they have since established a combination of human-driven and automated editorial processes to promote or filter certain content types. Strategies to combat the spread of disinformation ranged from fact-checking suspicious information, the deletion or reduction of the reach of suspicious profiles, and alerting users that they might unwittingly have shared disinformation.

2.1. Fact-check, delete, slow spread, repeat

Fact-checking is the most widespread strategy to counter disinformation on social media platforms. Fact checks and dedicated sections have become a regular feature of many established media outlets. There are entire websites dedicated to fact-checking, such as Snopes and PolitiFact in the United States.

Many social media sites have entered partnerships with third-party partner organizations. For instance, Facebook outsources the debunking of viral posts flagged by users through cooperation with over 60 third-party fact-checking organizations worldwide. Their conclusions are used to flag misinformation and downrank the posts in the News Feed algorithm to limit their spread (Constine 2020). In addition to human fact-checking, Facebook has invested heavily in Artificial Intelligence to supplement the scrutiny done by human eyes, mainly to identify modified imagery (Sumbaly et al. 2020). The messaging app WhatsApp also collaborates with organizations of the International Fact-Checking Network (IFCN), encouraging users suspicious of information received on their app to contact the IFCN fact-checking chatbot.

The video service YouTube started showing US viewers informational panels from fact-checkers such as FactCheck.org and PolitiFact in April 2020, although it does not share a full list of partners. First introduced in Brazil and India in 2019, the panels appear on searches for topics where fact-checkers have published relevant articles (Newton 2020b).

Twitter, whose executives at one time referred to the platform as "the free-speech wing of the free-speech party," broke with its previous low interference approach when it introduced a blue exclamation point label and warning messages for tweets containing disputed or misleading information related to COVID-19, in May 2020. The label also displays a "get the facts" tag linking to more information. However, contrary to other platforms, Twitter did not partner with independent fact-checking organizations. Earlier that year, the service had introduced a label for Tweets containing "synthetic and manipulated media" that had been "significantly and deceptively altered or fabricated" (Shapiro and Juhasz 2020).

Another common strategy to limit the spread of disinformation is the suspension of suspicious accounts. Facebook's "Community Standards," a list of official rules first published in 2018, spell out objectionable activities liable to removal and potentially a ban from using the platform. Nudity, graphic violence, child abuse, and hate speech content, for example, flagged either by users or algorithms to review, are removed or escalated by one of roughly 15,000 contractors all over the world. Likewise, Facebook routinely deletes accounts that engage in "inauthentic behavior," for example accounts, which conceal the true identity, purpose and ownership or control of a page. Similarly, Twitter enforces violations of its manipulation policies by removing accounts. These ban users from "artificially amplify or suppress information or engage in behavior that manipulates or disrupts people's experience on Twitter" (Twitter n.d.).

Besides labeling fake news and suspending accounts, social media platforms have also employed more interventionist approaches to limit the spread of misleading information and to alert their users that they engaged with it. In the wake of a surge of mis- and disinformation on COVID-19, WhatsApp introduced new measures to limit the reach of viral messages. The platform had previously faced criticism when it was used by electoral campaigns to spread false rumors, doctored photographs, and hoaxes in the run-up to elections in Brazil and India (Murgia et al. 2019). In contrast to other platforms, WhatsApp's end-to-end encryption means that it cannot see the contents of messages sent on the service. In an attempt to reduce the spread of misleading information, it set limits to the forwarding of viral messages. Messages sent through a chain of five or more people, considered "highly forwarded," can only be forwarded to one more user, instead of to the previous limit of five (Newton 2020a). Furthermore, in August 2020, it enables users to search the contents of messages that have been forward through a chain of five or more people (J. Porter 2020).

As for Facebook, content rated false by independent third-party fact-checkers is not removed, but instead, its distribution is reduced by appearing lower in the News Feed. This approach is justified by the need to keep its users "informed without stifling productive public discourse" as well as a "fine line between false news and satire or opinion" (Facebook n.d.). In a similar line, a metric called Click-Gap downranks posts that link to websites that are disproportionally popular on Facebook compared with the rest of the web in an attempt to promote more trustworthy news sources and reduce the spread of misinformation (Dreyfuss 2019). Furthermore, Facebook has trialed several iterations of tools meant to alert users that they engaged with dis- and misinformation with a message on their News Feed. About a year after the 2016 US presidential election, the company introduced a tool to alert users if they had liked the page created by the Internet Research Agency, the Russian propaganda group active during the elections. During the COVID-19 pandemic, a similar tool invites users who shared disinformation to visit a page created by the World Health Organization debunking popular COVID-19 myths (E. Porter 2020).

2.2. Does it work?

In reports and statements, social media companies are keen to stress the effectiveness of their measures in limiting the prevalence of misleading information on their platforms. Most companies publish regular reports outlining the results of their efforts in enforcing their rules. Besides these regular reports, companies also publish statements to announce their clampdown's success on specific actors.

Thus, in May 2020, Facebook announced that warning labels were added to about 50 million pieces of content related to COVID-19, based on around 7,500 articles by their independent fact-checking partners (Rosen 2020). Furthermore, in its August 2020 quarterly Community Standards Enforcement Report, the company claims to have deleted 1.5 billion fake accounts in the second quarter of 2020, representing approximately 5% of its active users (Facebook

2020). As for fact-checking, Facebook claims that the lower ranking in the News Feed of an article labeled "false" reduces future views by 80% on average (Levin 2018). Similarly, Twitter publishes its biannual Transparency Report since July 2012, in which it outlines the number of accounts locked or suspended for violating its rules. In a separate statement in June 2020, Twitter announced it had suspended 23,750 accounts plus 150,000 “amplifier” accounts run by groups connected to the Chinese government, linked to a “manipulative” disinformation campaign concerning the Hong Kong protests and coronavirus (Murphy and Yang 2020). Finally, on 27 April 2020, WhatsApp claimed its forwarding limits cut the spread of viral messages by 70 percent in weeks after introducing a new restriction three weeks earlier (Singh 2020).

Despite large online platforms reporting that their measures against disinformation are working, there remain significant challenges. For one, the actual efficacy of the current fact-checking process remains uncertain. The main complicating factors are volume and speed. For example, it is estimated that 500 million tweets are sent on Twitter every day. Moreover, the process of fact-checking takes time, so when fake news stories are identified as potentially false, the label is often added after the story has already been shared widely and the damage has been done (Levin 2017).

Facebook also entertains a difficult relationship with its own fact-checkers, with journalists working as fact-checkers for Facebook complaining that the company ignored their concerns and failed to use their expertise to combat disinformation (Levin 2018). One of the frustrations of fact-checkers with Facebook is that the company has used vague warning labels on false information to alert users to potential interactions with disinformation. The platform has justified these with backfire, i.e., the psychological effect when contradictory evidence does not change but actually increases an individual's belief in the initial misconception (Nyhan and Reifler 2010). Facebook claims that some of its interventions had the opposite effect of the one intended. For example, when it began labeling false posts as "disputed" after the 2016 election, users shared them more (Levin 2017). However, scientific research on the backfire effect is disputed, with recent experimental research indicating that presenting individuals with corrections is effective (E. Porter 2020). Thus, it may well be that a social media company like Facebook has found clear evidence of the backfire effect that external researchers have been unable to demonstrate. It is, however, impossible to verify this claim unless the company publishes internal research and allows an inspection of the underlying data.

This highlights the central challenge policymakers and regulars face, namely that the kind of information large platforms are required to make public in transparency reports remains largely unregulated. To date, there is a lack of meaningful regulation defining which data should be provided as well as the concrete form transparency data should take. The regular reports published by social media companies certainly highlight the scale of the problem of mis- and disinformation on their platforms. However, as of today, the social media companies’ claims regarding the scale of the problem and the enforcement and effectiveness of the

measures undertaken to stifle the spread of disinformation cannot be independently verified due to a lack of independent oversight.

As companies vying for future investment and under increased public pressure to tackle the problem of false information, social media platforms have an interest in demonstrating the efficacy of their measures. For instance, the forwarding limits imposed by WhatsApp in April 2020 appear to have a significant impact on the spread of viral messages, assuring that the platform remains “a place for personal and private conversations,” according to a spokesperson (Singh 2020). However, there is currently no way to verify the accuracy of this claim. Similarly, Facebook has refused to publicly release any data to support its claims that the dissemination of articles labeled as "false" is reduced by 80%.

Even in cases where transparency is mandated by law such as the German Network Enforcement Act (NetzDG), researchers and regulators alike have found the transparency data provided by Facebook to be highly problematic, with Facebook fined 2 million Euros for miscategorizing and misreporting data required in its government reporting requirements under the NetzDG (Wagner et al. 2020). This is due in part to Facebook prioritizing its own internal content moderation policy over external legal constraints systematically, but also to a lack of a joint industry standard by which data about content moderation is published. There is neither a standardized format provided by NetzDG that the resulting transparency data provided by either Facebook or any other online platform could be considered comparable. This lack of standardized reporting cannot just be blamed on states alone. It is equally due to the failure of large online platforms to standardize the manner in which they report their content moderation practices.

Finally, policymakers not only do not know how to resolve the policy issues at hand, but they are even unable to gain a basic understanding of what the core problems associated with it might be. Private companies’ voluntary provision of data in transparency reports is not just problematic because that data is unverified, but because their own presentation of categories and standards for transparency data allows them to shape the debate's dimensions extensively. The way in which private sector platforms like Google or Facebook provide transparency reports under public disclosure requirements such as the German Network enforcement law or the EU General Data Protection Regulation (GDPR) is as a mechanism to manage the visibility of certain categories and obscure visibility from others (Albu and Flyverbom 2019; Flyverbom 2016; Flyverbom, Christensen, and Hansen 2015).

Regulators and the general public are thus unable to make accurate determinations about what is happening in online platforms because they are currently unable to access accurate data about them. This, limits both effective decision-making about the nature of existing policy problems policymakers are aware of, as well as the ability to be able to respond to policy challenges they are not yet aware of.

In all of these contexts, the dominance of large transnational online platforms exacerbates this problem. Large platforms can more easily 'play' existing national jurisdictions against each other, for example, by threatening to switch their head offices' locations if substantial regulatory burdens are increased. This was one of the key reasons why Tesla built their first European office in the Netherlands, and it seems a plausible way to explain low rates of implementation of EU data protection law GDPR in by the Irish data protection authority. As one leading international election observer noted, that "we're running after the tech companies, they have enormous resources, and they're playing us" (Wagner et al. 2020). The dominance of large online platforms also contributes to limiting the ability of any regulatory jurisdiction to gain access to relevant data.

3. What else is being done?

Confronted with social media companies' apparent incapacity to effectively stop the spread of mis- and disinformation on their platforms, various other stakeholders such as civil society, governments, and security services have stepped in. For instance, local civil society organizations provide social media platforms with information about fraudulent accounts attempting to influence elections. This was the case in Moldova in 2019 when Facebook removed more than 160 fake accounts and pages after they were reported by the civil society organization Trolless (Gleicher 2019). On the other hand, in the Czech Republic, a group of PR professionals created a blacklist of online media platforms disseminating disinformation to deprive a media ecosystem plagued by a blend of pro-Russian propaganda and anti-EU rhetoric of advertisement revenue (Brokes 2020). Yet, civil society groups and political party members have pointed out that it can be challenging to attract Facebook's attention to notify them of fake news and profiles (Berzina 2019).

The spread of dis- and misinformation is also a significant concern for governments. As an example, we will look at the German and UK governments' reaction to the proliferation of false news. Following reports that dis- and misinformation circulating on social media in the run-up to the US Presidential Election may have affected voting, German Chancellor Angela Merkel declared a clampdown on 'fake news' on social media in November 2016. In a speech in the Bundestag, Merkel contended that "opinions aren't formed the same way as they were 20 years ago […] Today we have fake sites, bots, trolls – things that regenerate themselves, reinforcing opinions with certain algorithms, and we have to learn to deal with them. […] We have regulations that allow for our press freedom, including the requirement for due diligence from journalists. Today we have many that experience a media that is based on very different foundations and is much less regulated" (Cockburn 2016). Besides the 2016 US elections, the COVID-19 pandemic was another catalyst for government attention to dis- and misinformation. For instance, in April 2020, the UK government denounced Russian-backed media for disseminating “disinformation” about the country’s Prime Minister’s health during

the coronavirus crisis. The news outlet Sputnik had reported that Boris Johnson was “receiving lung ventilation," citing a "health source." The UK government also expressed its dismay at the "crazed conspiracy theory" that 5G masts spread the coronavirus. The government's spokesperson said that the UK government works with social media companies to “press for further action to stem the further spread of falsehood and rumors” (McGuinness 2020).

The German and the UK governments have introduced a number of initiatives to help counter misinformation distributed online. Partly in response to concerns about insufficient regulation of misinformation on social media platforms, Germany drafted its "Netzwerkdurchsetzungsgesetz" (Network Enforcement Act, NetzDG) in 2017. While the law primarily targets hate speech, unlawful content in this legislation also covers misinformation. The law, aimed at large social network websites with more than 2 million members, imposes high fines for noncompliance with existing legal duties to remove illegal content within 24 hours and requires social media companies to publish a biannual report on how they deal with complaints. However, the attempt to tackle the spread of disinformation through legal means is controversial because it privatizes the policing and censorship of what is and is not acceptable forms of expression (European Commission 2018; Wagner and Ferro 2020). Laws obliging social networks to take down posted content also likely leads to overzealous blocking of content by platforms afraid of fines, threatening freedom of speech (Human Rights Watch 2018).[1] Regulation of disinformation thus represents “a blunt ant risky instrument” (European Commission 2018).

The United Kingdom does not currently have any legislation that regulates news posted on online platforms. However, several government-sponsored reports have recommended introducing laws mandating social media companies to remove content identified as harmful or face fines. In the meantime, the Cabinet Office created a Rapid Response Unit in April 2018. Composed of analyst-editors, data scientists, media and digital experts, it aims to "monitor […] news and information being shared and engaged with online to identify emerging issues with speed, accuracy and integrity" (Government Communication Service 2018). In addition, following Russian disinformation after the poisoning of Sergei Skripal and the circulation of disinformation about the Brexit referendum, disinformation is perceived as a potential national security threat, with foreign actors seeking to influence UK citizens. In line with this outlook, the government initiated several initiatives, such as the National Security Communications Team. Charged with “combating disinformation by state actors and others,” the unit is part of the March 2018 National Security Capability Review and was approved at a meeting of the National Security Council (NSC). The unit’s goals to “more systematically deter [the UK’s] adversaries and help […] deliver on national security priorities” (BBC 2018) echo previous British initiatives during the Cold War period, in which Foreign Office’s Information

[1] A review of the German parliament's law in 2020, which compels platforms to report hate crimes to the police, is currently pending signature by the German president, as part of it is deemed to potentially violate the constitution.

Research Department (IRD) was set up to counter Soviet propaganda (Lomas 2018). The National Security Communications Unit was paired with the Fusion Doctrine, which, according to news reports, charged intelligence services to use social media to disrupt misinformation (Mccann and Farmer 2018).

4. Why is the current securitized institutional response to disinformation a problem?

The approach of the United Kingdom of responding to propaganda with counterpropaganda highlights the increasing securitization of the core responses to disinformation. However, the focus on disinformation as a geopolitical conflict between nation-states is highly problematic from a democratic governance perspective. As a result, social media platforms have found themselves at the crossroads of geopolitical battles, with incessant pressure being put on them to clamp down on false information. However, as several speakers underlined at the Freedom House's Second Annual Media Policy Forum in Moldova, governments should not dictate media, including social media, what to say (Berzina 2019). In the current situation, social media platforms are torn in a politicized battle they cannot win, and they inevitably will choose the side of whoever is the most powerful.

While investigations into disinformation campaigns during the 2016 US election have shown that there clearly is a geopolitical dimension, there is an urgent need to devise appropriate responses to disinformation outside of the realm of (geo)politics. Independent arbiters are most suitable for this task, and they are also more likely to take a human rights-based approach. Furthermore, a key reason for the securitization of government responses is that public institutions charged with responding to disinformation lack access to verifiable data on the scope and scale of disinformation. At the moment, if an election board or a media regulator wants to know what types of digital content are being shared in their jurisdiction, they have no effective mechanisms for finding this data or ensuring its veracity. Without accurate and verified data, policymakers and media regulators face limits both in terms of effective decision-making about the nature of existing policy problems they are aware of, as well as the ability to be able to respond to policy challenges they are not yet aware of. Only access to verified data from large online platforms offers the chance to develop effective responses to disinformation. The following sections will look at some of the challenges around unverified data and propose possible solutions on how to respond to them.

5. Verified data as a response to an ongoing problem[2]

[2] Parts of sections 5 to 8 are based on Wagner and Kuklis (2021).

The most helpful way to ensure that public institutions can make accurate determinations about what is happening on online platforms would be to develop an institutionalized mechanism for the verification of platform data. “Verified data” is understood as verified data provided as part of transparency reports by platforms or similar public disclosures. Independent auditors subsequently check this data to ensure it is an accurate representation of the platform's state. While a great number of regulatory bodies already exist, giving all regulators competencies and capacities to verify data essential for the exercise of their duties individually would create considerable redundancies. As a result, a separate institutionalized mechanism that provides a verification function for data provided by platforms to regulators would be the most effective response to this problem.

For independent auditors to check and verify the data provided by social media platforms, they will need access to relevant data. However, access in this context does not entail access to all data at all times by all regulators. We do not advocate for the creation of 'direct access' to online platforms by any government regulator. For instance, the Chinese government appears to have direct access to relevant data on popular online platforms such as Sina Weibo, TikTok or WeChat (Hong and Harwit 2020; Jiang 2019; Jiang and Fu 2018; Kloet et al. 2019). The government's capability to correct what they consider disinformation on these platforms in near real-time and heavily influence platform developments in the area of content moderation suggests a great deal of access to data and a close relationship between government regulators and large online platforms. However, in this highly authoritarian solution with unfettered access to what citizens do online, it is not possible to safeguard key human rights such as freedom of expression or privacy, which makes a direct access style solution untenable for democratic governments.

Following legal theorist Max Weber (Weber 1980), the signature of legitimate governmental authority is that any power exerted by governments is fundamentally limited and includes legitimate reasons for those individuals towards whom power is being exercised to refuse to comply. Such ‘limited’ approaches to the exertion of power by governments are difficult to square with direct access approaches that provide access to all of the data all of the time.

At the same time, the current opacity and lack of accountability of social networks can be said to encourage authoritarian approaches to the governance of the internet because they leave policymakers with few viable policy options. The temptation to resort to authoritarian methods to control disinformation is particularly strong when social media platforms' unchecked power can sway elections or other key democratic goods. It is thus imperative to provide governments with alternative non-authoritarian means to hold dominant online platforms accountable. The first step to achieve this is by providing access to accurate and verified data.

6. Auditing intermediaries

The appropriate institutional accountability mechanism (Bovens 2010) to warrant the accuracy of data made available by online platforms is to create an “auditing intermediary." Independent auditors are widely relied upon in the financial sector to independently assess financial records and business transactions of companies they are not affiliated with. Independent auditors are typically used to avoid conflicts of interest and to ensure the audit’s integrity. In this case, the intermediary would be in charge of auditing data provided by large online platforms upon request. The creation of such an institution would result in several advantages, notably to limit the exposure of private data to as few actors as possible, to reduce potential security risks, to ensure regulators remain within the scope of their mandate, to streamline the process, to provide access to data to smaller regulators lacking the capacity to organize audits themselves, and to limit the risk of institutional capture of the regulators.

First, channeling the auditing process through centralized auditing intermediaries limits sensitive private data exposure to as few actors as possible. Privacy and data protection are central concerns for organizations that wish to provide transparency, with existing privacy laws such as the GDPR limiting mechanisms disclosure (Bankston 2018; Keller 2018). Using 'auditing intermediaries’ limits challenges associated with privacy and data protection, as it can ensure that a more limited subset of verified data is provided to both regulators and the general public. It also follows the principle of data minimization, which is enshrined in Article 5 of the GDPR.

Second, distancing the audit process from the regulator asking for data ensures that regulatory action does not overstep its bounds (Hodge 2015; Viscusi 1996). Ensuring that regulators remain within the scope of their mandate is particularly important given the diversity of regulators interested in regulating online platforms and the considerable power that can be drawn from access to their data (Becker 2013; Yan 2018).

Third, by limiting the number of points through which the online platforms need to interact with outside intermediaries, it limits potential security risks that could arise from providing a broad set of different regulators access to a wide variety of systems. It is to be assumed that any kind of access provided to data by large online platforms is highly likely to constitute a considerable security risk. As a result, limiting the number of individuals with access limits the potential exposure to this specific risk.

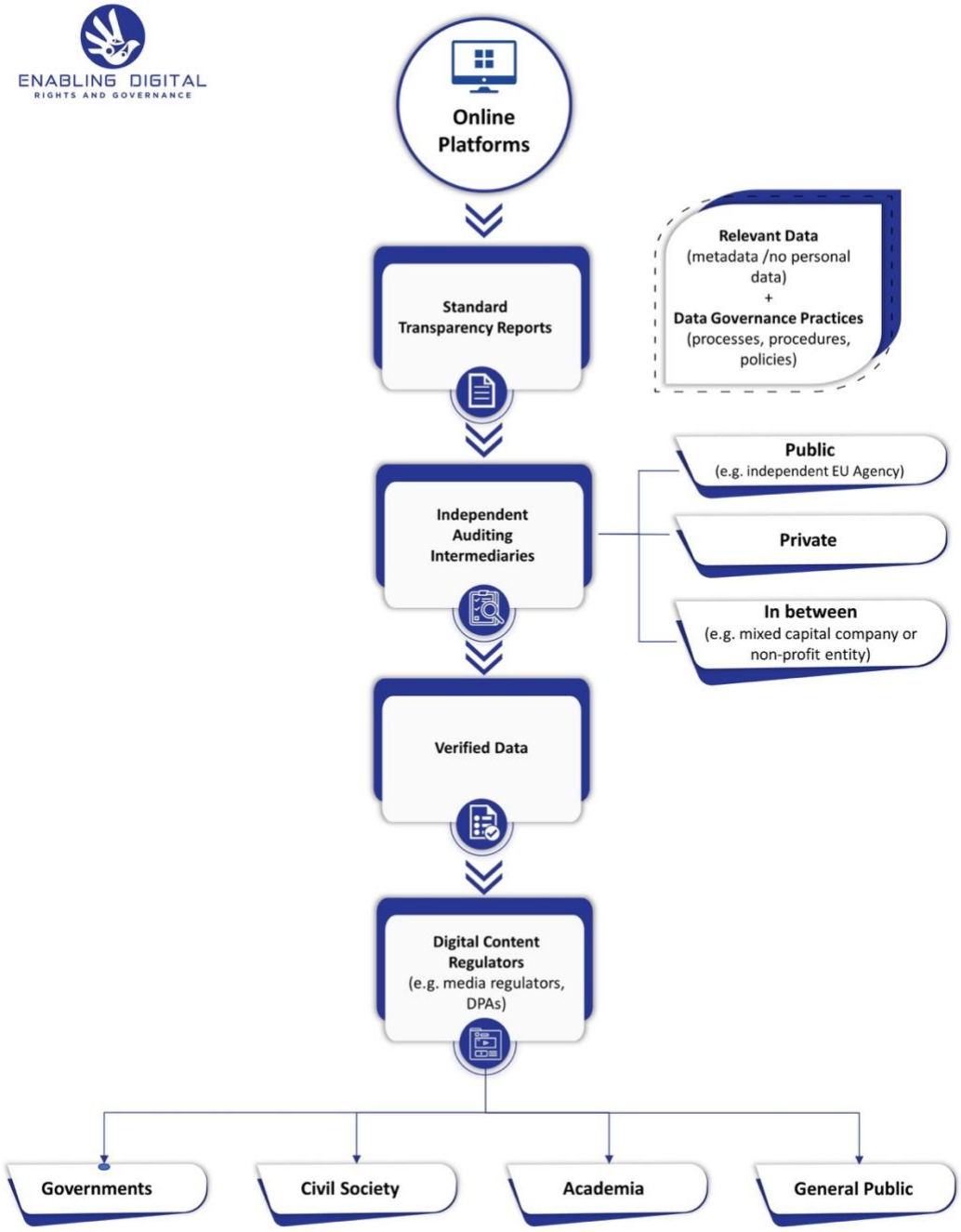
Fourth, having numerous regulators involved in auditing is likely to create countless unnecessary and redundant processes in which similar regulators ask similar questions that need to be answered separately over and over again. This challenge is not dissimilar to regulating government surveillance practices (Korff et al. 2017), where ensuring effective oversight depends heavily on ensuring that online platforms are not able to provide conflicting answers to a set of broadly similar questions. Centralizing the answers provided through a

central point thus avoids redundancy and strengthens the coherence of the overall argument being made.

Fifth, organizing auditing of transparency data through an external auditing intermediary ensures that even regulators without the capacity to organize audits themselves still may have access to such a system through auditing intermediaries. Even existing European regulation like the GPDR is posing considerable challenges in regards to enforcement, with key regulators like the Irish Data Protection Authority seen as lacking the capacity to do so effectively (Scally 2020). This challenge is even more critical in jurisdictions that are less developed and therefore have fewer resources to invest in regulation. However, it is precisely these jurisdictions where regulatory support is most needed. The ability to regulate a large online platform should not be limited to the largest and most powerful regulators.

Sixth, there is an ongoing interchange of staff between media regulators and those being regulated, which brings with it the risk of institutional capture of the regulators (Nielsen, Gorwa, and de Cock Buning 2019; Short 2019). This risk is even more pronounced in regard to auditing intermediaries as a result of their potential access to particularly sensitive material. A staff member of an auditing intermediary could not audit Facebook and then work for Google six months or even several years later. As these kinds of restrictions are particularly onerous and limit the recruitment of staff, they should be limited to a small group of auditors rather than a wider regulatory body, although they are, of course, desirable for regulators as well. As such, the creation of an auditing intermediary brings considerable benefits with it, but what would it look like in practice?

The diagram below offers a visualization of the Independent Auditing Intermediaries model that is being proposed here.


Online Platforms
Standard Transparency Reports
Relevant Data
(metadata /no personal data)
+
Data Governance Practices
(processes, procedures, policies)
Independent Auditing Intermediaries
Public
(e.g. independent EU Agency)
Private
In between
(e.g. mixed capital company or non-profit entity)
Verified Data
Digital Content Regulators
(e.g. media regulators, DPAs)
Governments
Civil Society
Academia
General Public

7. Public auditing intermediaries

Auditing intermediaries can be public, private or somewhere in between. In the case of public intermediaries, the first and most important condition is that any such public intermediary would need to be highly independent.[3] This has been a challenge in previous iterations of public sector platform regulation, which is part of why an independent agency – preferably at a European level – would be of such high importance. For example, the German 'Bundesamt für Justiz,' (BfJ) is entrusted with enforcing the Network Enforcement Act (NetzDG) which is, in turn, one of the key current elements of platform regulation in Europe. However, the BfJ is not an independent regulator. Instead, it is directly attached to the German Ministry of Justice and has to follow the Ministry's instructions and the politically-appointed Minister of Justice (Wagner et al. 2020; Wagner and Ferro 2020). As such, a public agency similar to the BfJ would not be in a position to conduct this kind of verification.

Of course, there are independent agencies that may lend themselves to this task at first glance, such as media and other regulators and data protection authorities. Media regulators, for example, are independent agencies within their respective national contexts. However, the extent of their independence varies to a considerable degree, and even those that can be considered sufficiently independent are usually not equipped with the capacities or competencies for auditing data. Although not inconceivable, it would require substantial restructuring of these institutions in every EU member state to allow for such an activity.

Data protection authorities (DPAs) are independent agencies as well. Through their experience and expertise with data protection impact assessments under the GDPR and their in-house technical skills, they would be well-equipped to conduct these kinds of audits. However, they are already significantly understaffed and underfunded to respond to the GDPR without having additional burdens for additional tasks placed upon them. Most importantly, their role as DPAs in ensuring compliance with data protection rules and regulations is very different from auditing the accuracy of transparency reports.

As a result, a new institution would need to be created that draws on auditing expertise in the private and public sectors to verify the claims made by social media providers. Such an institution could be created within the context of the proposed European Digital Services Act (DSA). It should, however, be a distinct legal entity to safeguard its independence from other institutional actors working in this area. The new entity does not have to be the main new DSA body but could provide an important supporting role. The ability to draw on expertise from the European Court of Auditors, from the European Data Protection Supervisor (EDPS),

[3] Auditing intermediaries can take many forms. Whether they are public, private or somewhere in between, they all are legitimate approaches to the challenge of auditing intermediaries. Due to limited space, this paper will only develop the approach of a public intermediary further here.

as well as from the private sector would be essential to enable the effective functioning of a new auditing intermediary.

This institution would be responsible for collecting verified data and making them available only to authorities endowed with the legal competence to use them, to a legally specified extent for a legally specified purpose. The collection and verification of the data on the one hand, and their use for regulatory purposes on the other, would, therefore, be distinct processes, which would further enhance the independence of the institutions involved, and guarantee the security of the data in question.

8. What challenges does the creation of an auditing intermediary create?

The proposal of auditing intermediaries brings with it its own set of challenges. The following section will briefly provide an overview of these potential difficulties and how some of them might be overcome.

a) How much access do auditing intermediaries need?

One of the key challenges raised by the proposal of auditing intermediaries is how much access to data these intermediaries would actually need. The risk to social media users' privacy is a major concern, given the prospect of a government regulator with access to all the digital content they are creating and sharing. However, public regulators do not need access to all digital content to combat disinformation or respond to problematic online content or hate speech. Nor do they – as some policy proposals have suggested - need to 'break encryption' or mandate unencrypted communications on key platforms in order to be able to conduct it effectively.

Instead, like any other similar auditor from the financial sector, they would need access to relevant data about the platform, the infrastructure behind it, and the existing policies in place. This is similar to how compliance with anti-money laundering rules is monitored in the financial sector. In the United States, banks are required under existing US anti-money laundering (AML) legislation to monitor certain types of transactions and submit suspicious activity reports to the Financial Crimes Enforcement Network (Naheem 2015). When compliance with money laundering legislation is audited, auditors are not looking at each individual transaction or document received by the bank, but rather at the procedures and mechanisms that have been put in place to produce these results (Naheem 2016).

So, in analogy to this practice, auditing the processes and procedures in place on social media to produce reporting is likely to be much more effective than providing access to all pieces of data. Thus, auditing mechanisms do not have to include personal data of any individuals. An

understanding of the procedures around how personal data is processed, managed and governed is likely to be far more important. Being able to reproduce and spot check that the transparency reports are being produced accurately represent the platform's data governance practices is critical to any meaningful audit.

An additional important consideration is extra meta-data that any regulator would need to assess the extent to which the content they are seeing is relevant. This type of meta-data could conceivably include:

(1) Frequency of content being viewed/ globally
(2) Frequency of content being viewed by countries
(3) Distribution of content being viewed across accounts

There may also be a need for additional high-level data that can be used to identify key content that is likely to be relevant from regulators' perspective. Importantly, such meta-data may need to change and/or be updated and adapted over time. Thus, it is essential to allow sufficient leeway within any regulation for the list of relevant data to be updated on a regular basis.

b) Misuse of auditing intermediaries for strategic national interests

Even without any kind of direct access, auditing intermediaries remain an important locus of power. Given their ability to gain some degree of access to the dominant online platforms they are auditing, they will quickly become the focus of struggles for power. While this is evidently already the case within powerful online platforms themselves (Moore and Tambini 2018), auditing intermediaries are likely to be in a similar situation. Thus, they need to be adequately shielded from these power struggles by guaranteeing their institutional independence and ensuring their staff selection and maintenance procedure is beyond reproach. Institutional independence, in this case, means:

> […] that the regulatory body is adequately funded with its own budget, is a part of the public sector, and is able to act with complete independence, which in turn implies a decision-making power independent of any direct or indirect external influence. In this regard, the supervisory authorities must be able to ensure swift and effective impartial public decision making within their relevant regulatory framework. In other words, any such body must have institutional independence, rather than being subject to partisan political influence that affects its ability to make impartial decisions. It should have access to the human and financial resources needed to perform its assigned tasks, as well as to any relevant necessary technology. Only then will the regulators have the enforcement capability required to govern (Wagner and Ferro 2020).

Without meaningful protection, auditing intermediaries would quickly lose their credibility as impartial auditors (Funnell, Wade, and Jupe 2016; Gipper, Leuz, and Maffett 2019). This is why it is imperative to safeguard their independence and ensure effective staff selection and maintenance procedures.

c) Standard setting for online platform transparency reports

Finally, one of the most significant challenges is the current lack of common standards for providing data in transparency reports or indeed for different types of regulatory requests. Each company, be it Facebook, Google or Twitter, publishes its own data and each regulator makes requests in its own format. This absence of standardization and structure in reporting requirements is a major challenge for regulators, the general public, academics and dominant online platforms alike. As each platform has developed an 'organic structure' for responses to regulatory compliance, the meaning of platform responses to these requests is far from clear, let alone comparable.

At the same time, standard-setting for transparency reports takes place primarily through individual legislative acts for specific sectors or policy domains. There is no linkage for the reporting standards for privacy under the GDPR and for German the Network Enforcement Law (NetzDG), nor any attempt to coordinate or structure them in a systematic way. This leads to challenges as the systems of the platforms are not providing comparable data because the infrastructure that they have in place was not designed to collect it in such a manner. This challenge of structuring access to data is similar to government requests for additional passenger data from airlines (Hasbrouk 2020). Typically, the ways in which data is requested from online platforms and airlines alike assume common system and reporting mechanisms that allow for a systematic and standardized response. In doing so, they ignore the considerable time and investment required to ensure reporting is possible in a systematic and standardized manner.

Of course, all of this energy would not have to be expended if large online platforms had already, through an industry group, trade body or similar structure, developed their own joint standards for managing and governing content on their platforms. For the airline industry, three airline associations WCO/IATA/ICAO got together, and in their joint 'API Contact Committee' developed a standardized format and protocol called PNRGOV.[4] As such standards are lacking for social media platforms, there is a need for public sector actors to step in and define these standards themselves. Ideally, this systematic standardization of compliance requirements would create a common regulatory framework that allows regulators to make requests to online platforms without constantly reinventing the wheel.

[4] See https://media.iata.org/iata/passenger-data-toolkit/library.html for further details.

9. Conclusion

Current transparency data provided by large online platforms does not stand up to rigorous scrutiny, either by independent academics, governments, media regulators or civil society. In the same way that the financial services regulator relies on 'auditing intermediaries' to ensure the accuracy and veracity of companies' annual reports, so too should media regulators and election boards be able to rely on auditing intermediaries to ensure that the data they receive is accurate. In which other industry would it be considered reasonable to take a private company's claims about its business's critical financial aspects without independent verification? If we can expect this level of audited scrutiny for financial transactions, why not also for digital content?

This policy paper has highlighted several other examples from other areas, most notably the financial services and aviation, where relevant mechanisms exist. Although there is no need to reinvent the wheel, large online platforms' extraordinarily dominant power requires even higher standards of transparency, accountability and good governance, if auditing intermediaries are to be successful. We believe that this paper has shown that it is possible to develop independent auditing intermediaries and that there are many strong reasons to do so.

Oftentimes, the fight against disinformation and the right to freedom of expression seem at odds. However, a regulator of this kind (auditing intermediaries) can strengthen freedom of expression rather than impeding it. Freedom of expression is the right to seek, receive and impart information, even if it is frequently reduced to being able to say whatever you want without facing any consequences for doing so. Auditing intermediaries can strengthen the right to seek and receive information by ensuring that users are thoroughly and accurately aware of how the content on large platforms is governed. By ensuring greater transparency of online platforms, users will know why some content was removed, other content stayed up or why platform algorithms show certain types of content and not others. This contextual information is crucial to being able to exercise freedom of expression rights. Without it, users have to rely on the large online platform's statements without any verification or validation of the underlying data. By doing so, auditing intermediaries can contribute to stopping the spiral of privatization of the governance of freedom of expression, making it more transparent and accountable towards users and the public at large (Wagner 2018).

What is not possible, at this point, is to continue public debates or regulatory policy about the actions of large online platforms based on unverified data. Only if regulators have an accurate picture of what is actually happening on large online platforms, whether regarding disinformation or numerous other public policy issues, can they make accurate determinations of what steps to take. Neither regulators nor the general public should have to rely on social media's benevolence and other online platforms to know what is going on in their own media environments.

References

Albu, Oana Brindusa, and Mikkel Flyverbom. 2019. 'Organizational Transparency: Conceptualizations, Conditions, and Consequences'. *Business & Society* 58(2):268–97. doi: 10.1177/0007650316659851.

Bankston, Kevin. 2018. 'How We Can "Free" Our Facebook Friends'. *New America*. Retrieved 4 October 2020 (http://newamerica.org/weekly/how-we-can-free-our-facebook-friends/).

BBC. 2018. 'Government Announces Anti-Fake News Unit'. *BBC News*, January 23.

Becker, Lukas. 2013. 'Accountability Gets Personal'. *Risk* 26(7):28–29.

Berzina, Lolita. 2019. 'Together We Are Stronger: Social Media Companies, Civil Society, and the Fight against Disinformation'. *Freedom House*. Retrieved 14 September 2020 (https://freedomhouse.org/article/together-we-are-stronger-social-media-companies-civil-society-and-fight-against).

Bovens, Mark. 2010. 'Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism'. *West European Politics* 33(5):946–67.

Breland, Ali. 2019. 'The Bizarre and Terrifying Case of the "Deepfake" Video That Helped Bring an African Nation to the Brink'. *Mother Jones*. Retrieved 13 October 2020 (https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/).

Brokes, Filip. 2020. 'Czech Civil Society Fights Back against Fake News | DW | 10.06.2020'. *Deutsche Welle*. Retrieved 14 September 2020 (https://www.dw.com/en/czech-civil-society-fights-back-against-fake-news/a-53758412).

Cockburn, Harry. 2016. 'Angela Merkel Fears Fake News Could Ruin Her Re-Election Chances'. *The Independent*. Retrieved 2 October 2020 (https://www.independent.co.uk/news/world/europe/angela-merkel-fake-news-german-elections-donald-trump-a7439641.html).

Constine, Josh. 2020. 'Facebook Will Pay Reuters to Fact-Check Deepfakes and More'. *TechCrunch*. Retrieved 27 August 2020 (https://social.techcrunch.com/2020/02/12/reuters-facebook-fact-checker/).

Dreyfuss, Emily. 2019. 'Facebook Is Changing News Feed (Again) to Stop Fake News | WIRED'. *Wired*, October 4.

European Commission. 2018. 'A Multi-Dimensional Approach to Disinformation - Report of the High Level Expert Group on Fake News and Online Disinformation'. *Shaping Europe's Digital Future - European Commission*. Retrieved 1 October 2020 (https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation).

Facebook. 2020. 'Community Standards Enforcement'. Retrieved 2 September 2020 (https://transparency.facebook.com/community-standards-enforcement#fake-accounts).

Facebook. n.d. 'Community Standards'. Retrieved 2 September 2020 (https://www.facebook.com/communitystandards/false_news).

Flyverbom, Mikkel. 2016. 'Digital Age Transparency: Mediation and the Management of Visibilities'. *International Journal of Communication* 10(0):13.

Flyverbom, Mikkel, Lars Thoger Christensen, and Hans Krause Hansen. 2015. 'The Transparency - Power Nexus: Observational and Regularizing Control'. *Management Communication Quarterly* 29(3):385–410. doi: 10.1177/0893318915593116.

Funke, Daniel. 2018. 'These Were Some of the Top Hoaxes on WhatsApp before the Brazilian Election'. *Poynter*. Retrieved 25 August 2020 (https://www.poynter.org/fact-checking/2018/these-were-some-of-the-top-hoaxes-on-whatsapp-before-the-brazilian-election/).

Funnell, Warwick, Margaret Wade, and Robert Jupe. 2016. 'Stakeholder Perceptions of Performance Audit Credibility'. *Accounting and Business Research* 46(6):601–19.

Gipper, Brandon, Christian Leuz, and Mark Maffett. 2019. 'Public Oversight and Reporting Credibility: Evidence from the PCAOB Audit Inspection Regime'. *The Review of Financial Studies*.

Gleicher, Nathaniel. 2019. 'Removing Coordinated Inauthentic Behavior From Moldova'. *About Facebook*. Retrieved 14 September 2020 (https://about.fb.com/news/2019/02/cib-from-moldova/).

Goldman, Jan. 2006. 'Disinformation'. in *Words of Intelligence: A Dictionary*. Scarecrow Press.

Government Communication Service. 2018. 'Alex Aiken Introduces the Rapid Response Unit'. Retrieved 3 October 2020

(https://webarchive.nationalarchives.gov.uk/20200203104056/https://gcs.civilservice.gov.uk/news/alex-aiken-introduces-the-rapid-response-unit/).

Hasbrouk, Edward. 2020. 'Airline Passenger Data and COVID-19'. Retrieved (https://papersplease.org/wp/2020/04/06/airline-passenger-data-and-covid-19/).

Hodge, Neil. 2015. 'Overstepping Their Authority: As Regulatory Actions Increase, Organizations Are Finding That Regulators Can Be Overzealous in Their Pursuit of Justice'. *Risk Management* 62(1):28–33.

Hong, Yu, and Eric Harwit. 2020. 'China's Globalizing Internet: History, Power, and Governance'. *Chinese Journal of Communication* 13(1):1–7.

Human Rights Watch. 2018. 'Germany: Flawed Social Media Law'. *Human Rights Watch*. Retrieved 1 October 2020 (https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law).

Jiang, Min. 2019. 'Cybersecurity Policies in China'. in *CyberBRICS: Mapping cybersecurity frameworks in the BRICS*, edited by L. Belli. Rio de Janeiro: FGV Direito Rio.

Jiang, Min, and King-Wa Fu. 2018. 'Chinese Social Media and Big Data: Big Data, Big Brother, Big Profit?' *Policy & Internet* 10(4):372–92.

Keller, Daphne. 2018. 'Comments on the Guidelines on Transparency under Regulation 2016/679'. *SSRN Scholarly Paper*.

Kloet, Jeroen de, Thomas Poell, Zeng Guohua, and Chow Yiu Fai. 2019. 'The Platformization of Chinese Society: Infrastructure, Governance, and Practice'. *Chinese Journal of Communication* 12(3):249–56.

Korff, Douwe, Ben Wagner, Julia Powles, Renata Avila, and Ulf Buermeyer. 2017. 'Boundaries of Law: Exploring Transparency, Accountability, and Oversight of Government Surveillance Regimes'. *University of Cambridge Faculty of Law Research Paper No 16/2017*.

Levin, Sam. 2017. 'Facebook Promised to Tackle Fake News. But the Evidence Shows It's Not Working'. *The Guardian*. Retrieved 21 August 2020 (http://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working).

Levin, Sam. 2018. '"They Don't Care": Facebook Factchecking in Disarray as Journalists Push to Cut Ties'. *The Guardian*. Retrieved 14 September 2020

(http://www.theguardian.com/technology/2018/dec/13/they-dont-care-facebook-fact-checking-in-disarray-as-journalists-push-to-cut-ties).

Lomas, Dan. 2018. 'British Government's New "anti-Fake News" Unit Has Been Tried before – and It Got out of Hand'. *The Conversation*, January 26.

Matsa, Katerina Eva. 2018. 'Most Western Europeans Get News from TV as Print Reading Lags'. *Pew Research Center*. Retrieved 27 August 2020 (https://www.pewresearch.org/fact-tank/2018/09/27/most-western-europeans-prefer-tv-news-while-use-of-print-outlets-lags/).

Mccann, Kate, and Ben Farmer. 2018. 'Britain to Launch Counter-Propaganda War against Russia as Theresa May Unveils "Fusion Doctrine" Defence Plan'. *The Telegraph*, March 27.

McGuinness, Alan. 2020. 'Coronavirus: Downing Street Slams Russian "disinformation" over Boris Johnson Ventilator Claim'. *Sky News*. Retrieved 14 September 2020 (https://news.sky.com/story/coronavirus-downing-street-slams-russian-disinformation-over-boris-johnson-ventilator-claim-11969398).

Moore, Martin, and Damian Tambini. 2018. *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. Oxford: Oxford University Press.

Mozur, Paul. 2018. 'A Genocide Incited on Facebook, With Posts From Myanmar's Military'. *The New York Times*, October 15.

Mueller, Robert S. 2019. *Report on the Investigation into Russian Interference in the 2016 Presidential Election*. Washington, D.C.: U.S. Department of Justice.

Murgia, Madhumita, Stephanie Findlay, and Andres Schipani. 2019. 'India: The WhatsApp Election'. May 5.

Murphy, Hannah, and Yuan Yang. 2020. 'Twitter Removes Thousands of China-Backed Disinformation Accounts'. Retrieved 2 September 2020 (https://www.ft.com/content/deeb3704-5924-44e3-9363-dd28ce7f67f1).

Naheem, Mohammed Ahmad. 2015. 'HSBC Swiss Bank Accounts-AML Compliance and Money Laundering Implications'. *Journal of Financial Regulation and Compliance* 23(3):285–97. doi: 10.1108/JFRC-03-2015-0016.

Naheem, Mohammed Ahmad. 2016. 'Money Laundering: A Primer for Banking Staff'. *International Journal of Disclosure and Governance* 13(2):135–56. doi: 10.1057/jdg.2015.10.

Newton, Casey. 2020a. 'WhatsApp Puts New Limits on the Forwarding of Viral Messages'. *The Verge*. Retrieved 2 September 2020 (https://www.theverge.com/2020/4/7/21211371/whatsapp-message-forwarding-limits-misinformation-coronavirus-india).

Newton, Casey. 2020b. 'YouTube Brings Fact-Check Panels to Searches in the United States'. *The Verge*. Retrieved 27 August 2020 (https://www.theverge.com/2020/4/28/21239792/youtube-fact-check-panels-videos-united-states-misinformation-covid-coronavirus).

Nielsen, Rasmus Kleis, Robert Gorwa, and Madeleine de Cock Buning. 2019. *What Can Be Done? Digital Media Policy Options for Europe (and Beyond)*. Oxford: Reuters Institute.

Nyhan, Brendan, and Jason Reifler. 2010. 'When Corrections Fail: The Persistence of Political Misperceptions'. *Political Behavior* 32(2):303–30. doi: 10.1007/s11109-010-9112-2.

Porter, Ethan. 2020. 'Why Is Facebook So Afraid of Checking Facts?' *Wired*, May 14.

Porter, Jon. 2020. 'WhatsApp Adds Search Feature to Help Users Debunk Viral Messages'. *The Verge*. Retrieved 27 August 2020 (https://www.theverge.com/2020/8/4/21353807/whatsapp-magnifying-glass-search-coronavirus-misinformation-fact-check).

Rosen, Guy. 2020. 'An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19'. *About Facebook*. Retrieved 20 August 2020 (https://about.fb.com/news/2020/04/covid-19-misinfo-update/).

Scally, Derek. 2020. 'German Regulator Says Irish Data Protection Commission Is Being "Overwhelmed"'. *The Irish Times*, February 3.

Shapiro, Ari, and Aubri Juhasz. 2020. 'Twitter Vows That As Disinformation Tactics Change, Its Policies Will Keep Pace'. *NPR.Org*. Retrieved 2 September 2020 (https://www.npr.org/2020/03/04/811686225/twitter-vows-that-as-disinformation-tactics-change-its-policies-will-keep-pace).

Short, Jodi L. 2019. 'The Politics of Regulatory Enforcement and Compliance: Theorizing and Operationalizing Political Influences'. *Regulation & Governance*.

Singh, Manish. 2020. 'WhatsApp's New Limit Cuts Virality of "Highly Forwarded" Messages by 70%'. *TechCrunch*. Retrieved 3 September 2020 (https://social.techcrunch.com/2020/04/27/whatsapps-new-limit-cuts-virality-of-highly-forwarded-messages-by-70/).

Sumbaly, Roshan, Mahalia Miller, Hardik Shah, Yang Xie, Sean Chang Culatana, Tim Khathevich, Hervé Jegou, Matthijs Douze, and Zeki Yalniz. 2020. 'Using AI to Detect COVID-19 Misinformation and Exploitative Content'. Retrieved 19 August 2020 (https://ai.facebook.com/blog/using-ai-to-detect-covid-19-misinformation-and-exploitative-content/).

Twitter. n.d. 'The Twitter Rules'. Retrieved 2 September 2020 (https://help.twitter.com/en/rules-and-policies/twitter-rules).

Viscusi, W. Kip. 1996. 'Regulating the Regulators'. *The University of Chicago Law Review* 1423–61.

Wagner, Ben. 2018. 'Free Expression? – Dominant Information Intermediaries as Arbiters of Internet Speech'. in *Digital Dominance: Implications and Risks*, edited by M. Moore and D. Tambini. Oxford: Oxford University Press.

Wagner, Ben, and Carolina Ferro. 2020. *Governance of Digitalization in Europe: A Contribution to the Exploration Shaping Digital Policy - Towards a Fair Digital Society?* Gütersloh, Germany: Bertelsmann Stiftung.

Wagner, Ben, Krisztina Rozgonyi, Marie-Therese Sekwenz, Jatinder Singh, and Jennifer Cobbe. 2020. 'Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act'. Barcelona, Spain.

Wagner, Ben, and Lubos Kuklis. 2021. 'Disinformation, Data Verification and Social Media: Verifying Data through Auditing Intermediaries'. in Dealing with Digital Dominance, edited by M. Moore and D. Tambini. Oxford, UK: Oxford University Press.

Weber, Max. 1980. *Wirtschaft Und Gesellschaft: Grundriss Einer Verstehenden Soziologie*.

Yan, Xingyu. 2018. 'The Jurisdictional Delimitation in the Chinese Anti-Monopoly Law Public Enforcement Regime: The Inevitable Overstepping of Authority and the Implications'. *Journal of Antitrust Enforcement* 6(1):123–49.