Journal of Management and Governance
https://doi.org/10.1007/s10997-023-09694-5

Social media platforms' responses to COVID-19-related mis- and disinformation: the insufficiency of self-governance

Lina Warnke[1] · Anna-Lena Maier[2] · Dirk Ulrich Gilbert[3]

Accepted: 11 December 2023

Abstract
The spread of mis- and disinformation on social media platforms is a significant societal threat. During the COVID-19 pandemic, mis- and disinformation played an important role in counteracting public health efforts. In this article, we explore how the three most relevant social media platforms, Facebook, YouTube, and Twitter, design their (IT) self-governance as a response to COVID-19-related mis- and disinformation, and provide explanations for the limited scope of their responses. Exploring the under-researched connection between the operating principles of social media platforms and their limited measures against mis- and disinformation, we address a relevant research gap in the extant literature on digital platforms and self-governance, particularly the role of IT governance (ITG), providing the ground for our argument against an overreliance on self-governance. In our qualitative study that draws on publicly available documents, we find that the shortcomings of current responses to mis- and disinformation are partly due to the complex nature of mis- and disinformation, as well as the wider political and societal implications of determining online content's factuality. The core problem, however, is grounded in the current overreliance on self-governance. We argue for an enhanced dialogue and collaboration between social media platforms and their relevant stakeholders, especially governments. We contribute to the growing ITG literature and debate about platforms' roles and responsibilities, supporting the intensifying calls for governmental regulation.

Keywords COVID-19 · Disinformation · IT governance · Misinformation · Self-governance · Social media platforms

1 Introduction

Digital platforms assume increasingly powerful roles in society (Lindman et al., 2023). The rise of social media platforms as a subtype of digital platforms has been accompanied by increasingly critical accounts of their destructive potential

Extended author information available on the last page of the article

Published online: 01 February 2024

(Cusumano et al., 2022), as the heated debates in connection with Twitter's acquisition by Tech billionaire Elon Musk illustrate (BBC, 2022). The platform has since been renamed "X", however in this paper we will refer to it as Twitter. The reliance on social media as younger generations' primary information source and the uptake of mis- and disinformation on these platforms increase this threat (Marin, 2021, p. 2; Reuters Institute for the Study of Journalism, 2021a). This was particularly observable during the Covid-19 pandemic, which has also been referred to as an *infodemic*, prompting scholarly calls for measures against both mis- and disinformation and their underlying causes (Marin, 2020). The term infodemic means "a flood of information on the Covid-19 pandemic", which has been fueled by misinformation and disinformation spreading on social media (World Health Organization [WHO], 2021b).

How social media platforms manage and govern mis- and disinformation can be understood as a matter of both ethics and governance. Recent research has advocated for exploring how Information Technology Governance (ITG) may be used to proactively address ethical issues related to different kinds of information technology (IT) (Wilkin & Chenhall, 2020). With this article, we contribute to the growing ITG literature by exploring how the three most relevant social media platforms, i.e., Facebook, YouTube, and Twitter, design their self-governance measures to respond to COVID-19-related mis- and disinformation and provide explanations for the limited scope of these responses. Self-governance in this context is defined as regulations and guidelines that are issued by either a single company or a group of companies (industry self-governance) and applied to themselves to manage and control their businesses (Cusumano et al., 2021). The case of COVID-19-related mis- and disinformation on social media platforms is especially useful for this study as it has gained immense attention from various stakeholders on a global scale. Further, it demonstrated the destructive potential of mis- and disinformation spreading on social media. Roozenbeek et al., (2020, p. 12) find that a higher susceptibility to misinformation is directly linked to people's behaviors during the pandemic, resulting in vaccine hesitancy and less adherence to public health measures. Therefore, the COVID-19-pandemic presents a valuable empirical context for understanding social media platforms' self-governance mechanisms.

Corporate governance as the study of how an organization is governed and how decisions are made is a critical element in analyzing social media platforms' responses to mis- and disinformation. ITG is described as "an integral part of corporate governance and addresses the definition and implementation" of three key aspects: governance structures, processes, and relational mechanisms, which enable "both business and IT people to execute their responsibilities in support of business/IT alignment and the creation of business value from IT-enabled business investments" (Van Grembergen & De Haes, 2009, p. 3). We therefore understand ITG as a form of self-governance. IT *governance structures* refer to "organizational units and roles responsible for making IT decisions and for enabling contacts between business and IT management (decision-making) functions" (Van Grembergen & De Haes, 2009, p. 21). ITG *processes* are designed to ensure the alignment of daily business routines to corporate policies and provide feedback through "the formalization and institutionalization of strategic IT decision-making or IT monitoring procedures" (Van Grembergen & De Haes, 2009, p. 22). Lastly, *relational mechanisms* are announcements, channels, and educational efforts that are designed through participation and collaboration between the different corporate levels consisting of executives as well as business and IT managers. The latter is especially important for business and IT alignment (Van Grembergen & De Haes, 2009, p. 22).

We contribute to extant ITG research by exploring governance challenges linked to social media platforms as particularly relevant and powerful actors in the context of COVID-19-related mis- and disinformation. We specifically focus on social media platforms which "enable people […] to make connections by sharing expressive and communicative content, building professional careers, and enjoying online social lives" (van Dijck, 2013, p. 4). Globally, there are 4.8 billion active social media users (DataReportal et al., 2023) with an average daily use of 2 h and 22 min (Buchholz, 2022). Therefore, social media platforms are a large part of most people's lives and, hence, may significantly impact them and society more generally. A rather new and unintentional function of social media platforms is gathering information and news. A study by the Reuters Institute for the Study of Journalism (2021a) showed that 67% of respondents under the age of 25 use social media as a source of information. At the same time, relying on social media as the main information source results in more susceptibility to misinformation (Allcott & Gentzkow, 2017, p. 17).

Against this background, misinformation and disinformation have become particularly salient and consequential. Vosoughi et al., (2018, p. 1146) define *misinformation* as "information that is inaccurate or misleading". The degree of intention differentiates misinformation from *disinformation*. While misinformation is unintentional, disinformation spreads false stories deliberately (Geeng et al., 2020, p. 1). The COVID-19 pandemic has not been the first event that led to an increasing spread of misinformation and disinformation on social media platforms. The 2016 U.S. elections are a rather prominent example underlining the societal and political relevance of mis- and disinformation on social media platforms (Allcott et al., 2019).

The extent to which false information is spread *deliberately* is hard to assess. The three social media platforms we study correspondingly seem to prefer to speak of misinformation when describing their measures. However, most phenomena they refer to in this context are, in fact, better described as disinformation, for example, anti-vaccination conspiracy theories. Effectively countering disinformation on social media platforms has become particularly relevant in the ongoing COVID-19 pandemic. Disinformation that spreads quickly and widely undermines confidence in public health measurees and thus negatively impacts crisis management (Algan et al., 2022). Lack of trust in governments and limited scientific knowledge contribute to the consumption and spread of misinformation and disinformation on social media (Chowdhury et al., 2021). This is particularly problematic as anti-vaccine content and corresponding disinformation efforts have been found to directly contribute to vaccination refusal (Muric et al., 2021), thus counteracting public health efforts. Our main argument, and at the same time our central contribution, is that the spread of such harmful information is not accidental, but rather part of the fundamental design of social media platforms. It can thus be argued to be grounded in their basic operating principles and practices of value creation. To understand how a platform's value is created, it is essential to understand the basic structure of the underlying business model. Van Dijck et al. (2018, p. 10) point out that digital platforms monetize attention, data, and user valuation, which arguably affects their processes, governance structures and relational mechanisms, as well as measures to moderate and manage content. Zuboff (2015, 2019, 2022) argues that this is problematic because users of social media platforms are hardly aware of their role, thus creating a "behavioral surplus" that, in turn, creates revenue for the social media platforms, incentivizing the latter to manipulate users' online behavior so as to create further revenue.

The kind of information a user sees while using social media is highly determined by the platform's algorithms. Users can become trapped in filter bubbles as search results are personalized and preselected based on personal characteristics such as location or previous searches (Kompetenzzentrum Öffentliche IT, 2016, p. 99). Because algorithms are trained to promote provoking, sensational content that users engage with, and misinformation fits many of these criteria, they are often spread through the platforms' algorithms (Avaaz, 2020; Culliford, 2020; Eisenstat, 2021; Roberts, 2020). Consequently, this sparks a debate concerning the societal and political roles and responsibilities of social media platforms. Recent attempts to regulate social media platforms more strictly such as the Digital Services Act have been unsuccessful (Turillazzi et al., 2023), and, thus, in the absence of sufficient government regulation, an (over)reliance on self-governance by social media platforms themselves is observed. Cusumano et al., (2021, p. 1273f.) argue that self-regulatory responses of many platforms were adopted too late and remain insufficient to address current challenges. Empirically exploring the under-researched connection between governance principles and measures against mis- and disinformation, we thus ask: How do social media platforms design their (IT) self-governance as a response to COVID-19-related misinformation and disinformation?

To answer this question, this article proceeds as follows. We discuss some central operating principles of social media platforms, ITG mechanisms, their interaction and lastly address the current overreliance on self-governance in contrast to government regulation. In the subsequent methods section, we justify our case selection and provide more granular details on the empirical context of our study. We then present our core findings regarding the nature and scope of strategic responses of three social media platforms of interest. We then discuss our findings, highlighting potential pathways towards rebalancing voluntary action in terms of IT self-governance and government regulation. Our conclusion contains the most relevant contributions and addresses some of the limitations of this study.

2 Interdisciplinary literature review

In the following, we describe the most relevant features of social media platforms that lead to governance problems in general and mis- or disinformation in particular, both of which present important and critical issues to be addressed through effective self-governance in general and ITG in particular. It is thus important to first understand the basic operating principles of social media platforms and how they generate governance problems before clarifying the relationship between self-governance and government regulation.

2.1 Operating principles of social media platforms as a facilitator of mis- and disinformation

Digital platforms create value through network effects, which according to Parker et al., (2016, p. 17) "refers to the impact that the number of users of a platform has on the value created for each user". According to Srnicek (2017, p. 256), network effects are possibly the most value-defining feature of a platform that draw in more and more users and eventually lead to a monopoly. Algorithms, algorithmic decision-making, and the corresponding network effects are central to the functioning of digital platforms in general and social media platforms in particular (Zarsky, 2016). In the context of social media platforms, they (co-)determine what content users see and engage with.

Srnicek (2017, p. 254) argues that data are central to social media platforms and the main source of their economic and political power. Through users' information, photos, and activities, social media platforms are provided with more and more data, which teaches their self-learning software to better understand and predict users' actions (Royakkers et al., 2018, p. 139). Therefore, the cost of their power is the privacy of users. However, in contrast to other challenging areas of digitalization, data protection and privacy are receiving the most legal attention and supervision, for example, provided by the European Data Protection Regulation. Still, critics doubt the effectiveness and suitability of these regulations. Although there are some laws in place already, they do not necessarily consider the intertwined relationships between data, users, algorithms, and other factors that contribute to the functioning of platforms (European Commission, 2022a).

Personalization, as one of the main features determining why users are attracted and loyal to a particular platform (van Dijck et al., 2018, p. 42), creates a lock-in effect. The algorithms governing this process are kept secret as accurate predictions generate a competitive advantage. The algorithms on which personalization is based result in search engines being biased. Pariser (2011, p. 9) refers to this as *filter bubbles*, "unique universes of information" created by engineers that affect how ideas and information are perceived. Through the creation of individual online universes, the personalization movement is also threatening democracy (Zuboff, 2022). Democracy relies on taking on different viewpoints and shared facts, which are increasingly undermined by individuals' personalized environments online. Further, it limits one's autonomy and severely impacts meaningful decisions as the possible options presented were preselected by algorithms (Pariser, 2011, p. 16).

Similarly, echo chambers are an increasingly central issue in the spread of misinformation and disinformation on the Internet. Echo chambers refer to communities, especially on social media platforms, that share the same worldview (Colleoni et al., 2014, p. 319). Established views are reinforced within these echo chambers, and members are rarely exposed to alternative views and opinions (Lütjen, 2016, p. 17).

Dis- and misinformation in line with a certain echo chamber's view and ideology are diffused more quickly through the echo chamber (Törnberg, 2018). In contrast to filter bubbles controlled by algorithms, echo chambers develop through personal action. Moreover, social media platforms consciously and by design encourage predominantly negative emotional content, e.g., expressions of fear and anger (Steinert, 2020).

In sum, filter bubbles and echo chambers are best understood as a direct result of a social media platform's basic operating principles and form a central part of their value creation. At the same time, governance is costly, and implementing governance measures often creates tensions between economic value creation and governance costs (Huber et al., 2017). Therefore, implementing and potentially extending governance that would, for example, target filter bubbles and echo chambers on social media platforms, might be seen as a direct threat to the business models of large social media platforms relying on network effects.

2.2 IT governance

Mis- and disinformation, we argue, constitute a societal threat that is facilitated through social media platforms and reinforced through their basic operating, business model-informing principles. This is enabled through several circumstances. According to Marin (2021), sharing is a split-second decision regardless of its truth content. Further, Geeng et al. (2020) find that most users do not investigate the content they are sharing. This contributes to false information spreading faster and farther on social media than the truth (Vosoughi et al., 2018, p. 1149) and shapes users' opinions. However, the spread of misinformation is not just promoted by users but also by bots and algorithms. This is where ITG, as discussed earlier, comes into play due to the large role IT plays in governing and shaping social media platforms.

As briefly explained in the introduction, ITG broadly refers to governance structures, processes, and relational mechanisms within the respective organization (Van Grembergen & De Haes, 2009). The main task of ITG is to effectively and efficiently enable the organization to create business value, to mitigate risks associated with IT, and to facilitate the alignment between corporate vision, management practices, and the IT infrastructure (Bowen et al., 2007). Most ITG research focuses on the role of ITG for business and IT alignment and the resulting performance effects (e.g., De Haes & Van Grembergen, 2017).

How IT Governance mechanisms are used in an organization, among other things, depends on their inherent dependence on IT. Based on Nolan and McFarlan (2005), Héroux and Fortin (2014) categorize an organization's dependence on IT by means of four IT modes. The IT modes range from highly defensive, somewhat offensive, moderately offensive to highly offensive. Attributes determining the IT mode are IT intensity as well as size and decentralization of the IT function. We argue that due to their business model and as born digitals (Monaghan et al., 2020), social media platforms can be categorized as highly offensive with high IT intensity, a large IT function, and a low-moderate decentralization of their IT function. In highly offensive IT modes, ITG processes and relational capabilities are used to a moderate to high degree, whereas ITG structures are only used to a low to moderate degree (Héroux & Fortin, 2014, p. 161). To gain a better understanding of ITG mechanisms, Héroux and Fortin (2014) use several items to measure each construct in their survey. ITG structures are, for example, specific IT committees for security, projects, or architecture, with the board of directors having both expertise of IT risks and management functions responsible for IT security, risks, and compliance. This understanding is in line with Altemimi and Zakaria's (2015) identified drivers of ITG structures, which are authority and membership as well as coordination mechanisms, demonstrating that ITG structures determine responsibility and decision-making-functions. Héroux and Fortin (2014) further analyze ITG processes, expressed through formal processes regarding IT strategy, and work with external agencies to conduct IT security audits. Drivers of ITG processes are performance monitoring and the alignment of IT decisions with key business, thus addressing corporate strategy (Altemimi & Zakaria, 2015). Lastly, ITG relational capabilities entail senior executives being involved in shaping a vision and IT's role in the organization as well as the implementation of the vision throughout the organization (Héroux & Fortin, 2014). This is related to a number of drivers such as leadership, skills, collaborative relationships, as well as role, responsibility, and commitment and thus leads to establishing commitment and support both at the top management levels and throughout the whole organization (Altemimi & Zakaria, 2015).

2.3 ITG and social media platforms' responses to mis- and disinformation

There are different approaches to principles guiding self-governance efforts of social media platforms. Marin (2021), for example, established a hierarchy of norms of relevance to social media platforms that also affect the spread of (mis-)information. The first layer consists of legal norms. These are kept to a minimum by the platforms and communicated through terms and conditions, often concerning illegal activities, including hate speech and personal attacks. The second layer concerns meta-norms of sociality that aim to promote the further growth of the network within the law. Lastly, the third layer is the core layer concerning local and unpredictable norms. These depend on the group level and can apply to a community within a particular social media platform and, therefore, differ across different communities using the same platform. Social media platforms usually do not intervene or establish any restrictions to this level as long as these community norms follow the law and contribute to expanding the network (Marin, 2021). This hierarchy of norms addresses the requirement of IT decisions being aligned with the key business objectives and principles constituting ITG processes (Altemimi & Zakaria, 2015).

According to Marin (2021), measures to fight the spread of mis- or disinformation entail both elements of human supervision and algorithms. There are numerous approaches to limiting mis- or disinformation on social media platforms. Lazer et al., (2018, p. 1095) distinguish between two types of interventions. First, approaches that empower users to evaluate the encountered information and make informed decisions about their truth content and whether or not to share it. Second, intervention approaches that involve structural changes to prevent users from coming into contact with disinformation in the first place. Intervention approaches are mostly implemented voluntarily by social media platforms as a key self-governing mechanism in their fight against misinformation and disinformation, or driven by government regulation. Since disinformation can also be considered a central part of a social media platform's business model, Trittin-Ulbrich et al., (2021, p. 15) point out that platforms' priority would be to avoid or circumvent governmental regulations. For example, collaboration with fact-checkers is promoted.

Lazer et al., (2018, p. 1096) suggest altering the algorithms to emphasize high-quality information and provide information regarding the source's quality. Furthermore, the personalization of political information should be reduced. These self-governance mechanisms could thus potentially majorly interfere with the platforms' operating principles. Geeng et al., (2020, p. 2) analyze Facebook and Twitter and report that both platforms remove inauthentic and manipulative accounts. Moreover, users can manually flag posts, or these may be automatically detected and then demoted. Lastly, Facebook provides more information to users about an article's source. Before sharing information that is known to be false, the platform warns the user and offers related fact-checked articles (Geeng et al., 2020). Pennycook et al. (2020b) found that nudges, such as accuracy reminders, promote more thoughtful sharing behavior, which would benefit the fight against misinformation that is often shared unintentionally. Especially when it comes to changing platform algorithms to adequately address misinformation and disinformation, ITG processes are needed that are performance monitoring oriented, take into account current developments, and can be adapted to strategic changes at short notice. (Altemimi & Zakaria, 2015).

A mechanism available on almost every social media platform is flagging, sometimes referred to as reporting. This practice is defined by Crawford and Gillespie (2016, p. 411) as "reporting offensive content to a social media platform". In this sense, the users are engaged in shaping the platform's content and, in a way, community values. As flagging is often just the first step in potentially removing content from the platform, this practice adds legitimization to the platform's final decision (Crawford & Gillespie, 2016, p. 412). However, there are also downsides to the practice of flagging. One of the greatest challenges of this mechanism is that users may abuse it. For example, users may flag content as a joke because of an existing feud or competition with other creators, or may even bully and harass creators. Due to the limited communication and interaction, the difference is almost impossible to detect for the platforms, which may leave the mechanism invaluable (Crawford & Gillespie, 2016, p. 420). Moreover, Pennycook et al. (2020a) highlight that the approach may lead to an *Implied Truth Effect*. This states that content that is not labeled is granted higher credibility and is assumed to be accurate, thus creating a false sense of security.

Another option is the voluntary collaboration between governments and social media platforms. For example, in the UK, the national health service (NHS) worked together with social media platforms to promote accurate COVID-19-related information and, at the same time, fight the spread of misinformation. Measures included, among others, the verification of governmental accounts to establish trusted sources, direct links, or easy access to accurate information provided by the NHS for COVID-19-related searches and the exclusion or removal of identified false information and their creators (Lovell, 2020). This self-governance mechanism addresses both the "skills" as well as "collaborative relationship" aspects of relational mechanisms of ITG as indicated by Altemimi and Zakaria's (2015) framework. Whilst some progress has been made with these mechanisms, mis- and disinformation continue to circulate on social media platforms, hinting at the relevance of exploring the underlying causes more thoroughly. In addition to the presented self-governance mechanisms, the issue of mis- and disinformation has also sparked public debate resulting in governmental regulation that will be discussed and contrasted to social media platforms' self-governance approaches in the next section.

2.4 Self-governance versus government regulation: emerging challenges for itg of social media platforms

The potential and challenges of self-governance have been explored by different literatures, for example, concerning labor conditions in global value chains (Bartley, 2007), proactive Corporate Social Responsibility (CSR) initiatives, codes of ethics, or informal agreements (Cohen & Sundararajan, 2015, p. 124–125; Cusumano et al., 2021). In contrast to government regulation, i.e., regulation issued by state authorities in the form of hard law, self-governance mostly refers to soft law, for example, in the form of voluntary standards or corporate codes of conduct.

Rather than understanding government regulation and self-governance as opposites, Gunningham and Rees (1997, p. 366) argue that the latter should be viewed as a continuum. Thus, according to Cusumano et al. (2021), self-governance includes all regulations and guidelines that are issued by either a single company or a group of companies (industry self-governance) and applied to themselves. Often, self-governance includes cooperation with third parties and may also involve governments. Thus, self-governance instruments can include a diverse range of activities adopted by social media platforms to address misbehavior on their platform preempting potential governmental interference.

The rise and spread of mis- and disinformation is of increasing concern for lawmakers and civil society actors (Eisenstat, 2021; European Parliament, 2021). Although self-governance is already in place in many areas, Ghosh (2021) implies that these may be best understood as mere acts of self-preservation. The European Parliament (2021) equally argues that platform guidelines do not adequately respond to current challenges. As a reform of the current system is required, cooperation with governmental actors is essential to evoke change (Ghosh, 2021). However, a central problem of governance is the international nature of platforms which potentially impedes accountability. Van Dijck et al., (2018, p. 138) summarize the difficulty of regulating international platforms: "A key issue is how public values can be forced upon the ecosystem's architecture—an architecture whose core is overwhelmingly controlled by (US) tech giants pushing economic values and corporate interests, often at the expense of a (European) focus on social values and collective interests". Mis- and disinformation are increasingly recognized as a considerable threat to democracy as a whole, but also to citizens' physical security and health. However, initiatives like the European Union's "Code of

Practice on Disinformation" that address mis- and disinformation are only binding to its voluntary signatories and thus limited in scope (European Commission, 2022b). Chase (2019, p. 1) accordingly notes that the code "is unlikely to achieve its goal of curtailing 'disinformation'", highlighting the ambiguity of misinformation as the content is harmful but mostly not illegal. Assessing the code's impact one year after implementation, Briggs (2020) finds that its self-regulatory nature failed to stop the spread of misinformation. She promotes a shift away from the self-regulatory framework to mandatory regulations to create a European-wide legal environment.

Another form of self-governance is the disclosure of information related to the procedure of fighting mis- and disinformation in (voluntary) CSR reporting, such as within the framework of the Global Reporting Initiative (GRI). Within the GRI standards, the general tension between moderating content and not inhibiting the fundamental right of freedom of expression is reflected (GRI, 2014). However, the GRI standard that specifically applies to social media platforms does not address situations in which a restriction may be required.

Another widely used reporting standard is the Sustainability Accounting Standard provided by the Sustainability Accounting Standards Board (SASB). Regarding freedom of expression, two accounting metrics are related to misinformation. Firstly, platforms should report countries where they may be subject to "government-required monitoring, blocking, content filtering, or censoring" (SASB, 2018, p. 17). Secondly, it is suggested that companies disclose requests by governments to remove content and their compliance behavior (SASB, 2018, p. 18). It is noticeable that the SASB standard only proposes to disclose activities related to and sparked by governmental regulation. No further disclosure or voluntary actions are recommended.

Other reporting standards that will be relevant in the future are the European Sustainability Reporting Standards (ESRS). These standards are being developed in connection with the EU's Corporate Sustainability Reporting Directive (CSRD) (Directive (EU) 2022/2464). Under the CSRD, international social media platforms will be required to provide a report in accordance with the ESRS. In addition, sector-specific standards are planned to be released. Under the segment "internet media and services", social media platforms will be subject to sector-specific disclosures (EFRAG, 2022, p. 38). Further, governance is a central part of the ESRS' understanding of sustainability. A greater focus will be placed on firms' governance structures regarding sustainability issues which will likely also include mis- and disinformation due to their societal impact.

Nevertheless, CSR reporting is not yet established and advanced enough to cover the "governance gap" identified by Jørgensen and Zuleta (2020, p. 63). They highlight that a legal framework to protect society and human rights is required. Other than hate speech or violent incitements, misinformation is not illegal but protected by the freedom of expression. Accordingly, regulation faces the challenge of balancing the claim to truth and protecting the freedom of expression (Horn, 2021, p. 9). These overall challenges regarding a balance of self-governance and governmental regulation notwithstanding, finding effective and balanced measures countering mis- and particularly disinformation is an urgent societal task.

3 Methods

We conducted a qualitative case study to analyze how social media platforms respond to COVID-19-related mis- and disinformation, and to thus obtain a better understanding of the potential and the limitations of self-governance in this context. In the following, we describe the empirical context of our study and explain how we collected and analyzed our data.

3.1 Empirical context and case selection

COVID-19 was first reported to the WHO on 31 December 2019 (WHO, 2021a). As of September 2022, the WHO reports more than 6.5 million confirmed COVID-19-related deaths (WHO, 2022). To contain the spread of the virus, various mechanisms were implemented, ranging from testing and tracing approaches to mandatory vaccinations. Early on, containment strategies were complicated by false information spreading online (Gooch, 2020). Such information ranges from deliberate false information to unintended inaccuracies and covers issues such as the prevention or treatment of COVID-19, governmental intents behind containment measures, the origin of the virus, and vaccinations (WHO, 2020).

We focus on the social media platforms Facebook, YouTube, and Twitter as they seem to be particularly important for users to obtain information. Regarding the use of social media platforms for news, Facebook (46%), YouTube (27%), and Twitter (11%) rank in the top three (Reuters Institute for the Study of Journalism, 2021a). According to the number of active users, Facebook and YouTube are the two largest social media platforms, with 2.9 billion and 2.5 billion users, respectively. Although it only ranks 14th with 556 million active users (We Are Social et al., 2023), with 22% global active usage penetration Twitter is right behind Facebook, YouTube, and Meta's other social media platforms (Reuters Institute for the Study of Journalism, 2021b). Table 1 provides an overview of the three social media platforms**.**

3.2 Data collection

We focused our data collection efforts on archival data, i.e., publicly available documents such as corporate websites or press releases. We searched the social media platforms' websites for press releases and articles that addressed the platform's recent mechanisms and initiatives against COVID-19-related mis- and disinformation. All corporate websites also offered a specific section on their website concerning COVID-19 that collected information and resources regarding the pandemic and in particular addressed the public's concern for mis- and disinformation spreading on the platform. Moreover, information available on the platform itself was taken into consideration, for example, Facebook's Central Information Center. All three platforms also have help pages that provide information and explanations of the platform's functions and structure. All documents, which concern recent initiatives, current guidelines, or how users can contribute to fighting mis- or disinformation, were

Table 1 Overview of social media platforms

	Facebook	YouTube	Twitter
Monthly active users	2.958 billion[a]	2.514 billion[a]	556 million[a]
Revenue in 2021	$118 billion[b]	$28.85 billion (Ads)[c]	$5.1 billion[d]
Parent company	Meta platforms Inc	Alphabet Inc	Twitter Inc
Purpose	Connect with family, friends, and businesses	Entertainment and education	Information sharing
Founded in	February 2004[e]	February 2005[f]	March 2006[g]
Other associated services	WhatsApp, Messenger, Instagram, Oculus, Workspace[h]	Diverse Google services, Android, Google Cloud[i]	Vine Archive, MoPu, Twitpic Archive, Revue, Scroll[j]

[a] We Are Social et al. (2023)

[b] Meta (2022, p. 64)

[c] Alphabet (2022, p. 33)

[d] Twitter (2022, p. 45)

[e] Facebook (2021a)

[f] Google (2006)

[g] van Dijck (2013, p. 68)

[h] Facebook (2021d)

[i] Alphabet (2022, p. 5)

[j] Twitter (2021d)

collected in 2020 and 2021. Table 2 below provides an overview of the type of data on the platforms' initiatives against COVID-19-related mis- and disinformation. As the goal was to analyze the platforms' responses during the COVID-19 pandemic, we focused on collecting data published between 2020 and 2021. However, we also included two blog posts from before the pandemic as they provide general information regarding the platforms' strategies against mis- and disinformation that continue to be relevant.

3.3 Data analysis

We took an abductive approach to data analysis, which we understand as iteratively going back and forth between theory and data (Van Maanen et al., 2007). In our content analysis, this translated into a mix of open coding and applying pre-defined codes derived from the existing literature we reviewed (Ridder, 2020, p. 191). We thus used central concepts from the ITG literature, particularly structures, processes, and relational mechanisms, to systematically code and interpret the data. Following our interdisciplinary literature review, we expected to find some first order codes and second order themes related to empowering user intervention. At the same time, we paid particular attention to remaining open to potentially surprising findings to remain consistent with our interpretative methodology.

Table 2 Overview of collected data

Data type	Source	Published	Number of documents analyzed
Facebook			
Press release	Newsroom	2020, 2021	4
Article	Help center, transparency center, corporate website	2021	6
FAQ section	Investor relations website	2021	1
Blog post	Facebook for media	2017	1
Central information center	Platform	2021	1
Image	Corporate website	2020	1
Report	Corporate website, European commission	2021	2
YouTube			
Article	Corporate website	2021	5
Blog post	YouTube official blog: inside YouTube	2019, 2020	3
Terms of use	Corporate website	2021	1
Report	Corporate website, European commission	2021	2
Twitter			
Blog post	Twitter blog: company	2021	4
Policy	Help center	2021	3
Report	Corporate website, European commission	2021	2

The data analysis process was aided by the software MAXQDA 2020, as it facilitates the organization and visualization of codes. After the first author coded the first documents, the code list was revised and consolidated together with the second author to allow for a more coherent coding process to establish reliability. We wrote memos along the analysis process to document upcoming thoughts about connections, patterns, and atypical findings. After the coding process, we discussed these extensively and constructed our second order themes and the resulting aggregate dimensions. In doing so, we broadly followed the approach to coding proposed by Gioia et al. (2013). Table 3 contains our data structure and Table 4 shows illustrative examples of quotes from the analyzed documents and their respective codes.

4 Findings

Our findings suggest that certain screening methods and intervention approaches as well as partnerships are appropriate modes of self-governance of social media firms to address misbehavior on their platforms. Moreover, community guidelines and precisely defined rules outlining consequences for violations play a major role. The identified mechanisms for addressing mis- and disinformation on social media platforms are presented in the following findings sections, categorized by the identified data structure.

Table 3 Data structure

First order codes	Second order themes	Aggregate dimensions
Artificial intelligence systems	Algorithmic screening	Screening methods
Machine learning / automated flagging / automated tools		
Flagging	Manual screening	
Manual review		
Trusted flagger program		
User reporting		
Human expertise		
Mixed screening		
Improving search results	Empowering user intervention	Intervention approaches
Labeling		
Promoting trusted sources and information		
Make informed decisions		
Alerts		
Messages to users who have interacted with misinformation		
Central information centers		
Promote news / media and eHealth literacy		
Reduce exposure / reduce visibility	Structural intervention	
Higher exposure to trusted content / improve recommendation and ranking		
Delete fake accounts		
Restrict engagement options		
Remove content		

Table 3 (continued)

First order codes	Second order themes	Aggregate dimensions
Cross-sector collaboration		Partnerships
(Health) Experts		
Media industry		
Industry peers		
Global and national health and governmental organizations		
Fact-checkers		
Organizations		
National Adaptation		
Community guidelines		Community Guidelines
Approval in groups by admins		Consequences for violating users
Lock / remove account		
Strike system		

Table 4 Illustrative examples of the data analysis

Source	Quote	First order code	Second order theme
Rosen (2020) [Facebook]	"During the month of April, we put warning labels on about 50 million pieces of content related to COVID-19 on Facebook, based on around 7500 articles by our independent fact-checking partners."	Labeling Fact-checkers	Empowering user intervention Partnerships
Twitter (2021a)	"Our systems learn from past decisions by our review teams, so over time, the technology is able to help us rank content or challenge accounts automatically. For content that requires additional context, such as misleading information around COVID-19, our teams will continue to review those reports manually."	Automated flagging Manual review	Algorithmic screening Manual screening
Twitter Safety (2020)	"Starting in early 2021, we may label or place a warning on Tweets that advance unsubstantiated rumors, disputed claims, as well as incomplete or out-of-context information about vaccines. Tweets that are labeled under this expanded guidance may link to authoritative public health information or the Twitter Rules to provide people with additional context and authoritative information about COVID-19."	Labeling	Empowering user intervention
Jin (2020) [Facebook]	"To further limit the spread of misinformation, this week we are launching a dedicated section of the COVID-19 Information Center called Facts about COVID-19. It will de-bunk common myths that have been identified by the World Health Organization such as drinking bleach will prevent the coronavirus or that taking hydroxychloroquine can prevent COVID-19."	Central information centers	Empowering user intervention
YouTube (2021d)	"For content where accuracy and authoritativeness are key, including news, politics, medical, and scientific information, we use machine learning systems that prioritize information from authoritative sources in search results and recommendations."	Machine learning Promoting trusted sources and information Improving search results Higher exposure to trusted content / improve recommendation and ranking	Algorithmic Screening Empowering user intervention Structural intervention

Table 4 (continued)

Source	Quote	First order code	Second order theme
Twitter Safety (2020)	"Starting next week, we will prioritize the removal of the most harmful misleading information, and during the coming weeks, begin to label Tweets that contain potentially misleading information about the vaccines."	Label Remove content	Empowering user intervention Structural intervention
Twitter Safety (2021)	"The Q&A featured Dr. Anthony Fauci, US President Biden's chief medical advisor, and other members of the White House COVID-19 response team. In India, we worked with the Ministry of Health to organize Vaccine Vartha, a weekly expert talk hosted on Twitter that enables vaccine experts to answer citizen questions."	Promoting trusted sources and information (Health) experts National adaptation	Empowering user intervention Partnerships
YouTube (2021a)	"Note: YouTube's policies on COVID-19 are subject to change in response to changes to global or local health authorities' guidance on the virus."	Community guidelines	
Rosen (2020) [Facebook]	"We are also requiring some admins for groups with admins or members who have violated our COVID-19 policies to temporarily approve all posts within their group."	Approval in groups by admins	Consequences for users

Approaches and initiatives were mixed, and the data analysis identified common themes across all three platforms. Most analyzed documents specifically address COVID-19 responses. In the following findings section, the identified mechanisms addressing mis- and disinformation on social media platforms are presented, categorized according to the identified data structure. Understanding and categorizing these responses, although triggered by the specific event of the COVID-19 pandemic, promises to provide the empirical basis for being able to assess future ITG responses to similar large-scale events with considerable societal impact.

4.1 Screening methods

Before content can be removed or labeled, it needs to be detected and reviewed. For that purpose, social media platforms apply several different screening methods to identify and evaluate content. Three screening methods can be distinguished, i.e., algorithmic screening, manual screening, and a mixed screening approach. Algorithmic screening includes the use of algorithms to detect false or misleading content. All three platforms apply this method. The platforms claim a high success rate for algorithms detecting content automatically rather than it being reported by users or through manual screening (Facebook, 2021g).

The platforms also highlight the advantages of algorithmic screening and the areas in which it is particularly useful. In a report, YouTube (2021g) points out that automatic detection allows for faster and more precise action when enforcing its policies. Further, their machine learning tools are improving in different languages, and fact-checking agencies work in more than 60 languages (Twitter Safety, 2021).

The screening process can be divided into two steps. Firstly, content that may be violating the platform's policies needs to be detected. Secondly, it is reviewed and evaluated before deciding whether it violates policies and which intervention approach should be applied. These two steps may be carried out by either algorithmic or manual screening or a combination thereof. For example, Facebook applies artificial intelligence (AI) to remove COVID-19-related misinformation after the questionable content has been flagged through manual screening (Rosen, 2021).

A key characteristic of machine learning is that it needs to be trained by manual inputs. Manual screening includes the screening by platform employees as well as user reporting. This mechanism is still heavily applied by all platforms. Besides relying on their own personnel that manually screens content, during the COVID-19 pandemic, the social media platforms extensively collaborated with partners such as health experts and (governmental) health organizations. The mechanism is not just used in the context of misinformation, but in all violations of the platforms' community guidelines and policies (YouTube, 2019). Interestingly, Twitter is the only platform that does not communicate about user reporting, but rather communicates that it only relies on screening through trusted partners like public health authorities, NGOs, or governments (Twitter, 2021b). Section 4.3 contains more findings regarding partnerships that were implemented as a response to COVID-19-related mis- and disinformation.

According to YouTube, algorithmic screening is the most important and successful method. In second and third place are user reporting and detection through the *Trusted Flagger Program*. Only a small fraction of the removed videos is detected by NGOs and government agencies (YouTube, 2021e). After questionable content is flagged, it may be reviewed through manual screening as well (YouTube, 2021f). Facebook also cooperates with fact-checkers to make qualified decisions on the accurateness of COVID-19-related content (Rosen, 2021). Although machine learning processes are applied widely across the platform, YouTube (2019) acknowledges that "human expertise is still a critical component of [their] enforcement efforts".

Lastly, a mixed approach combining both algorithmic and manual screening can be found. This may refer to circumstances where algorithmic screening is not yet advanced enough so that, in consequence, manual involvement is still required. Further, a combined approach may be applied after content is flagged automatically. When a decision on further actions cannot be reached by the algorithm, manual screening is required. At Facebook, a mixed approach allows machine learning tools to evolve and be trained to be more effective and efficient (Rosen, 2021). Twitter equally relies on a mixed methods approach and holds that accounts will not permanently be suspended solely based on automated enforcement systems, but only after human review (Twitter, 2021a). After screening content and potentially identifying misleading information, intervention approaches are applied, which will be addressed in the following section.

4.2 Intervention approaches

Regarding the question how mis- and disinformation is dealt with, two types of intervention approaches of self-governance can be distinguished, which have also been introduced by Lazer et al., (2018, p. 1095). Platforms choose either an empowering user intervention approach or a structural intervention approach. Empowering users refers to providing tools to support users in making informed decisions. For example, a tool often used by social media platforms is *labeling*. Thus, content that is known to include false information is not removed, instead a disclaimer is added. Labels are applied either as an explicit warning or by providing links to additional information to offer context to the questionable content. By providing reliable sources, the users are supported in informing themselves about COVID-19 (Twitter Safety, 2020).

This approach is applied by all three platforms and often includes links to authoritative sources and third-party sites since a central part of the empowering user intervention is to promote trusted sources and information. In this regard, platforms are also working on improving search results for users who use social media platforms to find credible information about COVID-19 (Twitter, 2021a). In this context, Facebook even created a specialized information center concerning COVID-19, which features real-time updates from organizations such as the WHO (Clegg, 2020). The information center thus addresses various aspects around COVID-19 and collects all relevant information in one place for Facebook's users. It also aims to educate users of the globally practiced physical and social distancing approaches, guide people with a potential infection, and, lastly, provide links to relevant health authorities and organizations (Facebook, 2021e). The information center is continuously updated and extended, now also including a section addressing misinformation. Twitter created a comparable central place to collect information regarding COVID-19. On the platform, it is referred to as the COVID-19 Events page and "is available at the top of the Home timeline for everyone in 30 + countries" (Twitter, 2021a).

Overall, all governance mechanisms pursue the aim to empower users to make their own decisions based on reliable information. A common concept for all three platforms was that empowering user intervention approaches aim to promote informed decision-making (Facebook, 2021g). Facebook is also promoting the development and improvement of users' news literacy, which is an important skill in the fight against mis- and disinformation in the long term.

Lastly, the platforms may send out messages and alerts to users. In particular, Facebook sends messages to users who have interacted with content in the past, which since has been declared to include mis- and disinformation (Rosen, 2020). Regarding structural intervention, two approaches are mostly applied, namely removing content or reducing the visibility of content. Facebook clearly states what the conditions are for either approach to be applied. Facebook removes misinformation that is a potential threat to physical integrity. With this step, the platform relies on external health experts such as the WHO. Only content which promotes debunked information is removed (Facebook, 2021b). On YouTube, content that is not an imminent threat is not removed, but rather the visibility is decreased to limit its spread (YouTube, 2021b).

This approach is often combined with labeling, an empowering user intervention tool. Content that does not clearly contradict the platform's guidelines but may be misleading or false is labeled accordingly, or additional contextual information is provided (Twitter, 2021a). In combination with the provision of context and authoritative sources on questionable content, the platforms also pursue an approach where trusted content is given higher exposure and is promoted more through the platform's recommendation or ranking systems (YouTube, 2021d).

By downranking misinformation and highlighting trusted content, the platforms try to foster an environment where users encounter less mis- and disinformation and are presented with reliable information regarding COVID-19. In addition, fake accounts that only pursue spreading disinformation are being targeted and removed (Mosseri, 2017). This is also particularly pursued by restricting users' engagement options with content that is misleading but not removable. For example, Twitter (2021b) disabled engagement functions while content is being reviewed.

The platforms actively pursue a combination of these two intervention approaches and their different tools. This shows that, on one hand, machine learning mechanisms can be applied to stop misinformation from spreading and to reduce the exposure of users to misinformation. Nevertheless, on the other hand, the social media platforms acknowledge that it is also important to educate users and improve their news literacy skills in order for them to make informed decisions. In the long run, this might also improve the platform's problem of being polluted by mis-and disinformation. Although both approaches are actively pursued and combined by the platforms, our analysis showed that structural intervention is pursued much less frequently than empowering user intervention.

4.3 Partnerships

In many of their intervention approaches and several screening methods, the platforms cooperate with external partners. The most mentioned form of partnering is with health organizations or health experts. This often occurs in combination with "promoting trusted sources and information" and "labeling", which are both empowering user interventions.

All platforms pursue the aim to provide users with authoritative information. For that purpose, the platforms rely on public health experts, public health organizations, as well as on governments. While the most mentioned global health organization is the WHO, a focus is set on national adaptation as well. The platforms also provide links to national health organizations, often alongside global information from the WHO, and adapt their mechanisms to local circumstances (Twitter, 2021a).

The contribution of the WHO includes the provision of links to their website to provide a trusted and reliable source for users where they could find further information about the virus without the threat of encountering mis- or disinformation. Further, the WHO also publishes common COVID-19-related myths and "debunked" these on their website, which provides the social media platforms with a baseline and reference in regard to which content is inaccurate and serves as a guide for decision-making (Rosen, 2020).

In addition, Facebook tasks fact-checkers with reviewing content. Over time, this practice has expanded so that the platform now works with over 80 independent fact-checkers, allowing content to be reviewed in over 60 languages (Rosen, 2021). To ensure the independence and quality of the fact-checking organizations, they "are certified through the non-partisan *International Fact-Checking Network*, which is a subsidiary of the journalism research organization "The Poynter Institute" (Facebook, 2021f).

Another large group of partners includes experts. For example, the platforms show that they frequently consult global health experts when developing new strategies and policies. Twitter even helps experts to be heard and found by verifying their accounts, which might increase their reach and credibility (Twitter Safety, 2020). In this area, Twitter also organizes events where health experts can interact with users and answer questions concerning the virus (Twitter Safety, 2021).

Less common partnerships include the cooperation with industry peers, organizations, and the media industry. The cooperation with organizations mostly aims to support users' news literacy (Facebook, 2021f). The cooperation with the media industry aimed to support and protect journalists to ensure the availability of qualitative and reliable information and was mostly accomplished through donations (Facebook, 2021h; Twitter, 2021a). The platforms mention "[w]orking together with industry peers to keep people safe" (Twitter, 2021a), but concrete actions are not communicated.

Lastly, the platforms also promote cross-sector collaboration working together with the aforementioned groups as well as with users, governments, and NGOs. As an example, we can cite YouTube's Trusted Flagger Program that we described earlier or a similar approach adopted by Twitter (2021c).

4.4 Community guidelines and consequences for violations

Community guidelines build the basis of the platform governance and are established based on core values such as freedom of expression (Facebook, 2021c). They are also a key element in the platform's fight against mis- and disinformation. The guidelines contain policies and rules as to what is allowed and what is prohibited from being posted on social media platforms. Although the exact content of the community guidelines and the type of content that is prohibited go beyond the scope of this article, it is interesting to note that the guidelines specifically identify COVID-19-related topics that are subject to consequences. Thus, rather than formulating vague guidelines which may provide more leeway for both users and platforms, the platforms have decided to implement very specific policies. As new conspiracy theories or myths regarding COVID-19 develop and are debunked by official sources, these need to be added to the guidelines so that the platforms are able to limit the spread of such content (YouTube, 2021a). YouTube (2019) also provides more insights into the development process of community guidelines and their adaptations, showing that the platforms are eager to involve various stakeholders to improve their service. Further, they have noted that their COVID-19 policies "are subject to change in response to changes to global or local health authorities' guidance on the virus" (YouTube, 2021a).

Of equal concern are the consequences for users or content that violates the platform's community guidelines. This is in part a complement to the intervention approaches already mentioned. There are different consequences or corrective measures for content and users who violate the platforms' rules. For example, Twitter (2021b) established a strike system where different types of violations lead to the user accumulating strike points. The more strikes a user accumulates, the more severe the consequences are. After a 12-h account suspension, a 7-day suspension is imposed. If a user accumulates more than five strikes, the account is locked permanently. By this measure, the platform intends to foster a learning effect by increasing users' awareness of policies (Twitter Safety, 2021). YouTube and Facebook use similar strike systems. Another mechanism introduced by Facebook requires group admins to temporarily review and approve group content for groups whose members have previously violated COVID-19 policies (Rosen, 2020). Group admins are also responsible when the content they approved contains a violation, and may receive a strike (Facebook, 2021b).

5 Discussion

Social media platforms pursue various approaches of self-governance to counter mis- and disinformation related to COVID-19. In the following, we discuss our findings and use the mechanisms described in the ITG literature to answer our research question of how social media platforms design their IT self-governance as a response to Covid-19-related mis- and disinformation.

The platforms' responses identified in our analysis can be related to the current ITG framework consisting of structures, processes, and relational mechanisms as summarized in Table 5. In this sense, *screening methods* and *community guidelines* can be understood in terms of ITG processes. Whether content violates the platform's rules is decided through formal procedures for IT-related decision-making that are executed through the screening methods and community guidelines. Further, they facilitate and enable the interaction between management and business operations. *Structural intervention* can also be understood in terms of ITG processes. The decision on limiting content's exposure through technical interference constitutes both the decision-making aspect and responsibility of ITG structures as well as the establishment of an IT strategy and policy as ITG processes entail. *Empowering user intervention* can be interpreted in light of ITG processes as an implementation of the platforms' IT strategies and policies to educate users. Further, the identified *consequences for users who violate the platforms'* are part of the formalized ITG processes, which are based on *community guidelines* and thus align business routines to corporate policies. Lastly, the identified *partnerships* can be interpreted in the light of ITG relational mechanisms focusing on the partnerships' collaborative characteristic. At the same time, external partnerships present an addition to the current ITG framework as it is not solely limited to collaboration within the organization but also includes external partners. We propose that in other ITG-related situations, involving external partners in the ITG mechanisms would also benefit the organization. As indicated earlier, ITG structures do not emerge from our extensive body of data. This absence already suggests that creating new ITG structures by, for example, setting up IT committees with specific expertise, would contribute to the effectiveness of social media platforms' responses to mis- and disinformation.

The data show that, although both empowering user intervention and structural intervention approaches are widely applied across all three platforms, structural

Table 5 Social media platforms' responses to mis- and disinformation related to ITG mechanisms

ITG mechanisms	Social media platforms' responses to mis- and disinformation
Structures	No findings in our data
Processes	Screening methods
	Algorithmic screening
	Manual screening
	Intervention approaches
	Structural intervention
	Empowering user intervention
	Consequences for violations
	Community guidelines
Relational mechanisms	Partnerships

intervention approaches were less present in the data. Since structural intervention, e.g. altering algorithms to prioritize verified content when adapting rating and recommendation systems, has a greater impact on the platforms' business model, this in turn potentially increases the costs of governance activities. It appears that platforms are more reluctant to apply and develop those intervention approaches. Our study thus empirically contributes to the growing literature on the interlinkages of platforms' operating principles and their ITG efforts. The broader ITG literature supports this interpretation, as it shows that governance activities are generally associated with considerable costs, which, in turn, impacts the design and extent of such activities (Huber et al., 2017). Going beyond the general role of (self-)governance costs, other studies have shown that content containing mis- or disinformation is attracting more attention to social media platforms than other content (Vosoughi et al., 2018, p. 1149) and is thus ranked higher by the algorithms (Avaaz, 2020; Culliford, 2020; Eisenstat, 2021; Roberts, 2020). This is supported by our observation of empowering user intervention approaches being pursued over structural intervention approaches which would actively interfere with the platform's algorithms and might alienate some users. These are further indicators that the platforms' economic interests in growing the platform might dominate and affect the scope of their interventions, which would require active alterations to their systems such as those presented in our findings.

Going back to Marin's (2021) conceptualization of a hierarchy of norms determining the spread of unintentional and deliberate false information, we see that through the core layer, platforms enable echo chamber forming and problems such as misinformation spreading. By keeping legal norms in the first layer to a minimum, there is greater room for platforms to evade their responsibility without having a compliance issue. Thus, we argue that the norms should be revised and that platforms should place more emphasis on the first layer. By increasing the legal norms, a safe environment can be established and dominating problems such as mis- and disinformation can be reduced. Focusing their responses on empowering user intervention approaches allows platforms to defer their responsibility to users without actively decreasing the amount of mis- and disinformation on their platform. This is further supported by van Dijck et al., (2018, p. 147), who state that platforms and their stakeholders "need to put long-term public value creation over short-term economic gain". Furthermore, such measures would indicate a shift from voluntary self-governance to compliance with legal norms, i.e., government regulation, thus potentially reducing complexity. When governance challenges related to managing content on social media platforms become a compliance issue, we argue, this may result in more responsible processes (e.g., daily business routines), governance structures (e.g., roles and responsibilities), and relational mechanisms (e.g., collaboration between different corporate levels) given a reduced level of ambiguity. At the same time, initial investments into corresponding governance measures might eventually lead to a reduction of governance costs (Huber et al., 2017). ITG research should establish whether and to what extent measures on the spectrum of government regulation (compliance) and self-governance (voluntary action) result in de- or increased complexity or costs.

Our findings also contribute to the emerging discussion on platforms' responsibilities by highlighting the complexity of factors underlying the appropriate balance of self-governance and government regulation. Focusing on transparency, for example, has limited effects: As an investigation of the Oversight Board (2021) found, Facebook has inconsistencies in its community standards as changes communicated on the website were not sufficiently transparent. Since it is externally unclear which rules the platform follows when making content decisions, an unfair process occurs that could lead to differential treatment in similar circumstances. While the Oversight Board's (2021) recommendation suggests more transparent reporting of the processes, it is questionable whether this self-governance measure will be sufficient.

The alternative approach of binding regulation might help but, in reality, is also hard to implement. The self-governance approach, on one hand, does seem to be effective since platforms release reports, for example, requested by the EU "Code of Practice on Disinformation", contributing to the aim of increasing transparency. On the other hand, self-governance as expected by the EU mainly serves the objective of transparency and does not provide further binding guidelines regarding the platform's behaviors and responses towards mis- and disinformation. National or international regulators could develop rules and guidelines to establish how mis- and disinformation is assessed, and which tools should be applied. They might also mandate greater investments in effective ITG structures. However, stricter regulation of e.g. algorithms and content management might lead to a decline in user satisfaction and, thus, contradicts the platform's economic interest. Collaborative governance initiatives involving a multitude of stakeholders seem particularly promising to address this challenge, as recent research finds that government regulation and self-governance can in fact beneficially reinforce each other (Schrempf-Stirling & Wettstein, 2023).

Since several intervention and screening tools are carried out through the platforms' algorithms, the general problem of opaqueness regarding algorithms remains. Examples for these are AI used in screening processes, automated flagging, as well as rating and recommendation systems influencing the exposure and visibility of content. Van Dijck et al., (2018, p. 70) demand that platforms provide more transparency regarding their algorithms and the enforcement of their guidelines. The findings show that algorithms are used extensively and, despite often being associated with problematic phenomena such as discrimination or filter bubbles, also show a great potential for limiting the spread of mis- and disinformation. We contribute to existing research by highlighting that transparency measures rely on consistency in their application.

The analyzed data provide little information on how the algorithms operate since they are also the platform's main asset, which, again, hints at the role of economic interests in hampering effective self-governance. Rather, the platforms focus on intervention approaches aimed at empowering users – one could also argue, by shifting the responsibility to the individual level. However, one positive indication is that the algorithmic screening considers decisions originally made by humans. Nevertheless, a myriad of datapoints on the platforms influence algorithmic decision-making. This highlights the foundational problem of social media platform's opaqueness regarding the mechanisms governing the platform (Zarsky, 2016). This also strengthens our argument in favor of increasing transparency regarding the algorithms governing social media platforms to fight mis- and disinformation. In the future, ITG research should shift its focus from business/IT alignment to governing emerging phenomena such as algorithmic decision-making more effectively and responsibly.

Along with algorithmic approaches, platforms referred to implementing intervention approaches by providing users with reliable information, e.g., through labels. Especially considering the high number of users consulting social media for information (Reuters Institute for the Study of Journalism, 2021a), this approach seems justified and may have a significant impact on improving the quality of information encountered by users on the platforms and, hence, may lead to less sharing of misinformation. For the platforms this approach has many advantages such as a greater alignment with its business strategy, protecting network effects, reducing risks associated with changes to its algorithms, and comparably low (governance) costs. However, its effectiveness regarding governance remains questionable due to the lack of taking responsibility and actively shaping the platform. So far, little governance attention has focused on this area.

CSR reports could be a potential means to include information regarding algorithms and measures countering mis- and disinformation, including empowering users' intervention approaches. The CSR standards introduced earlier encourage sections to address areas related to mis- and disinformation, such as freedom of expression or media literacy. As a result, including such sections in CSR reports might increase transparency, awareness, as well as accountability. However, the previous introduction of CSR standards has shown shortcomings that need to be addressed. Hence, the newly developed ESRS and its planned extension might be a useful tool fighting mis- and disinformation in the future, especially since the standards do not only require reporting but also imply a tighter engagement with sustainability risks, opportunities, and impacts in governance and management systems and structures. Here, the platforms should address additional areas of concern. Since algorithmic decision-making is still limited, the platforms heavily rely on manual screening and decision-making, as our findings regarding manual screening methods show, which, however, limit the ability to quickly react to mis- and disinformation. Because of the platforms' operating principles and algorithms, content spreads fast on the platform and is duplicated numerously to engage and attract new users to grow the network. The longer the detection of and decision-making on mis- and disinformation takes, the higher the potentially harmful impact the content may have. Further, the decision whether content is removed from the platform is also highly influenced by the platform's business model. This is partly seen in the approach how the platforms update their community guidelines with only specific debunked mis- and disinformation. Thus, the platforms should disclose information on the duration and underlying principles of the decision-making process.

While platforms comply with the currently weak regulation to prevent stricter laws (Ghosh, 2021; Trittin-Ulbrich et al., 2021), demands for tighter control are increasing. It is then likely that social media platforms will have to redesign their responses to mis- and disinformation by, for example, introducing more structural measures (e.g., specific IT committees possibly focusing on problem areas such as mis- and disinformation), thus also potentially increasing their governance costs. Eisenstat (2021) demands clear definitions of platforms' responsibilities and opportunities to hold them accountable when their actions lead to criminal activity. She argues that regulation should address the platforms' tools, such as recommendation and ranking systems. By making these tools governed by algorithms more transparent and enabling accountability, platforms may do more than the minimum of self-governance and positively contribute to fighting the problem of mis- and disinformation (Eisenstat, 2021). Similarly, Ghosh (2021) stresses that stricter governmental regulation regarding responses to mis- and disinformation is required. Our study supports this argument whilst highlighting that those self-regulatory measures that are quite effective should not be discarded completely. Considering the technical difficulty of regulating social media platforms and especially their algorithms (Fukuyama & Grotto, 2020, p. 200), alternative approaches that combine regulatory and self-governance approaches should be explored further. Cusumano et al., (2021, p. 1274) also argue that self-governance approaches would likely be more efficient if social media platforms joined forces, for example, by developing a joint code of conduct. This collaborative approach to governance appears even more promising, considering that the responses to COVID-19-related mis- and disinformation identified in this study already show great similarities among the different platforms.

A remaining conflict is the difficulty of regulating social media platforms in the international sphere, paired with the challenge of a myriad of cultural values and national laws. Some authors critically argue that U.S. American platforms, in particular, primarily pursue economic interests, which (arguably) contradict European social values and the pursuit of collective interests (van Dijck et al., 2018, p. 138). Rather than *solely* relying on direct government regulation of how social media platforms should counter mis- and disinformation, government regulation could also indirectly influence such approaches. For example, governments could impose regulation geared towards increasing competitiveness and countering the current oligopolistic structures of the social media industry, which might eventually reduce these platforms' societal and political power (Fukuyama & Grotto, 2020). This is also reflected in recent regulation such as the EU's Digital Services Act (European Commission, 2022a) or Digital Markets Act, which indicate a gradual shift towards more encompassing regulation of the digital industry.

Another aspect that is rarely considered in the literature is what consequences should exist for the users who create and spread mis- and disinformation. The consequences identified in the analysis have several shortcomings in this regard. The mechanism of removing users is likely to have little impact since users can easily create a new profile and continue using the platform (Daniel, 2021). Further, the echo chamber effect might make group administrators' approval counter-effective since group members and admins often share the same worldview and may, thus, not see certain content as problematic. Consequently, the anonymity of users on the internet impedes taking real action against persons or groups who create and share disinformation. However, since there are also reasons for protecting users' identity on the internet, e.g. to ensure the safety of whistleblowers, minorities, and users in countries that restrict freedom of speech (BBC, 2021), and anonymity cannot be granted situationally on a platform, this issue adds another layer of complexity.

In this context, greater collaboration between governments and platforms could aid the fight against mis- and disinformation, although such approaches are dependent on further conditions (Lahat & Sher-Hadar, 2020). Our findings show that partnerships, when responding to mis- and disinformation, allow for the diffusion of verified information provided by official (government) sources. This improves the trustworthiness, security, and timeliness of the provided information while decreasing the threat of mis- and disinformation. While, during the COVID-19 pandemic, governmental involvement was already very high, governments could benefit from this collaboration further by, for example, using local alerts to inform citizens about current events in crises as the platforms are striving to achieve a balance between global and national information for their users. Thus, whilst social media platforms mostly pursue a global approach in their strategy, they need to adapt nationally, especially with regard to COVID-19. As our findings show, platforms are seeking support from experts in developing their guidelines to ensure that definitions are correct. Additional cooperation with governments and experts could improve platforms' policies. Further, this could provide consistent definitions and approaches across platforms. Several scholars propose a cross-sector collaboration approach involving all stakeholders as a valuable source for defining new standards and guidelines to fight mis- and disinformation effectively and be well-prepared for future crises. Therefore, all involved parties need to reevaluate the role of self-governance and take into consideration the lessons learned from this infodemic, as it will certainly not be the last of its kind.

6 Conclusion

Our study highlights that mis- and disinformation, especially when being linked to public health emergencies, pose significant societal challenges. Our findings indicate that current responses of social media platforms to COVID-19-related mis- and disinformation are rather limited. Among other challenges, the processes of developing policies (i. e. community guidelines) to regulate content seem inflexible and insufficient, consequences against users are ineffective due to the internet's anonymity, and focusing on empowering users intervention approaches leads to an abdication from the platforms' responsibility and an individualization of a structural problem. These limitations might partly be explained by the complex nature of mis- and disinformation, as well as the wider political and societal implications of determining online content's factuality, for example, regarding freedom of speech. We contribute to the intensifying debate on social media platforms' responsibility considering mis- and disinformation by (a) showing that the currently observable over-reliance on self-governance is problematic as it is inherently limited by the platforms' operating principles and economic interests, and by (b) empirically identifying the shortcomings of ITG measures, e.g. the absence of ITG structures. As self-governance is insufficient, policy-makers should issue government regulation in a democratic process that ensures that social media platforms do not facilitate the spread of mis- and disinformation. Table 6 summarizes the key findings of our study and the associated challenges for effective ITG accordingly.

We contribute to extant research on ITG by initially highlighting the need to intensify scholarly attention on societal challenges such as mis- and disinformation and the still underestimated role of social media platforms in that regard. We go beyond the existing ITG literature by, first, applying the relatively under-explored case of social media platforms. The (self-)governance of their operating principles thus far receive limited attention in the ITG literature. Researchers have only just begun to explore the potential and limitations of self-governance of digital platforms (e.g., Cusumano et al., 2021, 2022). Thus, we conclude that future ITG research investigating effective governance mechanisms that help combating increasingly relevant and problematic issues such as mis- and disinformation seems particularly promising. Second, we extend the literature by looking beyond governance structures, processes, and relational mechanisms and how they affect Business/IT alignment and connect the ITG literature to research on self-governance of business in the context of the societal challenge of mis- and disinformation. We argue that the ITG literature would benefit from further exploring broader governance challenges of increasingly powerful actors such as social media platforms, and, in particular, the measures they employ to manage and govern the grand challenge of mis- and disinformation. Overall, we find that further incorporating insights from the multidisciplinary literature on multi-stakeholder collaboration and the role of governments in the governance of issues of societal concern could enrich the ITG literature further and contribute to identifying more effective governance mechanisms. The dynamic and complex nature of mis- and disinformation indicates that a more comprehensive and holistic approach to ITG might be needed. Our study presents a first step to fill the corresponding research gap.

Table 6 Overview of key findings and challenges for effective ITG

Key findings	Key challenge for effective ITG
Social media platforms currently shift most of the responsibility to users, e.g. through *empowering user intervention* and shortcomings of *consequences for violations*	Balance of corporate and individual responsibility
The operating principles of social media platforms are linked to their inherently limited responses to mis- and disinformation, e.g. their limited application of *algorithmic screening methods* and *structural intervention approaches*	Balance of business models and responsibility
The current reliance on self-governance is insufficient, which is why more government regulation is needed to overcome the limitations caused by the platforms' operating principles, e.g. more proactive *manual screening* and stricter *consequences for violations*	Shift from voluntary action to accountability
There remains room for self-governance; for example, social media platforms could voluntarily set up ITG structures, e.g. IT committees involved in development of stricter, more proactive *community guidelines*	Balance of self-governance and government regulation

As a practical consequence of our research, it turns out that social media platforms should proactively invest in ITG structures such as setting up IT committees with sufficient expertise regarding the subject of mis- and disinformation. However, since mis- and disinformation related to COVID-19 (and subsequent pandemics) constitute a matter of public health, governance should not be completely privatized. As observable in other sectors, governments increasingly reassert their authority regarding CSR, thus diminishing the room for purely voluntary initiatives and, correspondingly, shifting the balance towards compliance with regulation (Kourula et al., 2019). To face this challenge, social media platforms would be well advised to reflect upon their business models in light of societal challenges, and to proactively engage with governmental counterparts. One of the reasons for this conclusion is that self-governance of social media platforms is *inherently* limited through their basic operating principles and will necessarily fall short given the platforms' business models, as the latter limit their motivation to counter filter bubbles and echo chambers effectively. This is shown through the platforms' focus on empowering user intervention approaches in the fight against mis- and disinformation, rather than expanding structural intervention approaches which would be more effective in limiting the engagement with mis- and disinformation but at the same time compromise their business model.

We propose this empirically-grounded relationship as our central contribution to the growing literature on the roles and responsibilities of digital platforms in society, as well as to the intensifying scholarly debate on platform governance. Furthermore, our findings show that whilst platforms' self-governance is insufficient in combatting the spread of COVID-19-related mis- and disinformation, government regulation also has its limitations. This implies that there should be an enhanced dialogue and collaboration between social media platforms and their relevant stakeholders, especially governments. We argue that this could contribute to more nuanced responses to the COVID-19-related infodemic, e.g., with regard to the challenge of combatting mis- and disinformation whilst protecting the fundamental right of freedom of expression.

Our study is not without limitations. For example, we exclusively relied on archival data and were not able to assess the actual effectiveness of the platform's responses. Whilst offering an extension of the existing ITG literature by discussing governance challenges of social media platforms in connection to COVID-19, we do lack appropriate data on the social media platforms' governance structures as summarized in the ITG literature. As ITG mainly addresses internal organizational structures and processes, these were hard to assess from our external view. Future research could therefore draw on more immersive data collection methods like participatory observation and interviews to explore the effect of the platforms' operating principles on their responses to mis- and disinformation (as well as hate speech) and to gain insights on the internal ITG processes, structures, and relational mechanisms. A deeper inquiry of ITG structures related to platforms' responses to mis- and disinformation thus seems to be a particularly promising avenue for future research. Furthermore, we do not make any explicitly causal claims given the interpretative epistemology underlying our research design. Provided sufficient access to data can be obtained, future research could establish the relationship between business model and governance measures beyond conceptual theorization.

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was received for conducting this study.

Data availability The datasets generated during and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Algan, Y., Cohen, D., & Péron, M. (2022). *Why is trust key to managing crises?* World Economic Forum. Retrieved March 12, 2022, from https://www.weforum.org/agenda/2022/02/trust-factors-covid19-crisis/

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics, 6*(2), 1–8.

Allcott, H., & Gentzkow, M. (2017). *Social media and fake news in the 2016 election* (NBER Working Paper Series). Cambridge, Massachusetts.

Alphabet. (2022). *Form 10-K: annual report pursuant to section 13 or 15(d) of the securities exchange act of 1934 for the fiscal year ended December 31, 2021*. Retrieved September 15, 2022, from https://abc.xyz/investor/static/pdf/20220202_alphabet_10K.pdf?cache=fc81690

Altemimi, M., & Zakaria, M. (2015). Developing factors for effective IT governance mechanism. In *2015 9th Malaysian Software Engineering Conference (MySEC)*, (pp. 245–251).

Avaaz (2020). *Facebook's algorithm: A major threat to public health*. Avaaz. Retrieved September 10, 2022, from https://secure.avaaz.org/campaign/en/facebook_threat_health/

Bartley, T. (2007). Institutional emergence in an era of globalization: The rise of transnational private regulation of labor and environmental conditions. *American Journal of Sociology, 113*(2), 297–351. https://doi.org/10.1086/518871

BBC. (2021). *Social media: should people be allowed to be anonymous online?* Retrieved September 10, 2022, from https://www.bbc.co.uk/newsround/56114122

BBC. (2022). *Elon Musk warned he must protect Twitter users*. Retrieved September 10, 2022, from https://www.bbc.com/news/business-61225355

Bowen, P. L., Cheung, M. Y. D., & Rohde, F. H. (2007). Enhancing IT governance practices: A model and case study of an organization's efforts. *International Journal of Accounting Information Systems, 8*, 191–221. https://doi.org/10.1016/j.accinf.2007.07.002

Briggs, M. (2020). *Assessment of the code of practice on disinformation*. MediaWrites. Retrieved May 22, 2022, from https://mediawrites.law/assessment-of-the-code-of-practice-on-disinformation/

Buchholz, K. (2022). *Where people spend the most & least time on social media*. In Statista. Retrieved May 22, 2022, from https://www.statista.com/chart/18983/time-spent-on-social-media/

Chase, P. H. (2019). *The EU code of practice on disinformation: the difficulty of regulating a nebulous problem.* Transatlantic Working Group Working Paper.

Chowdhury, N., Khalid, A., & Turin, T. C. (2021). Understanding misinformation infodemic during public health emergencies due to large-scale disease outbreaks: A rapid review. *Journal of Public Health*. https://doi.org/10.1007/s10389-021-01565-3

Clegg, N. (2020). *Combating COVID-19 misinformation across our apps*. Meta. Retrieved May 22, 2022, from https://about.fb.com/news/2020/03/combating-COVID-19-misinformation/

Cohen, M., & Sundararajan, A. (2015). Self-regulation and innovation in the peer-to-peer sharing economy. *University of Chicago Law Review Online, 82*(1), 116–133.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in twitter using big data. *Journal of Communication, 64*(2), 317–332. https://doi.org/10.1111/jcom.12084

Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society, 18*(3), 410–428. https://doi.org/10.1177/1461444814543163

Culliford, E. (2020). On Facebook, health-misinformation 'superspreaders' rack up billions of views: report. *Reuters*. Retrieved May 11, 2022, from https://www.reuters.com/article/us-health-coronavirus-facebook-idUSKCN25F1M4

Cusumano, M. A., Gawer, A., & Yoffie, D. B. (2021). Can self-regulation save digital platforms? *Industrial and Corporate Change, 30*(5), 1259–1285. https://doi.org/10.1093/icc/dtab052

Cusumano, M. A., Yoffie, D. B. & Gawer, A. (2022). Pushing social media platforms to self-regulate. *The Regulatory Review*. University of Pennsylvania Law School. Retrieved February 14, 2023, from https://www.theregreview.org/2022/01/03/cusumano-yoffie-gawer-pushing-social-media-self-regulate/

Daniel, E. (2021). Twitter introduces "strike system" for vaccine misinformation. *Verdict*. Retrieved April 11, 2022, from https://www.verdict.co.uk/twitter-strike-system/

DataReportal, Meltwater & We Are Social. (2023). Number of internet and social media users worldwide as of April 2023 (in billions). In Statista. Retrieved June 20, 2023, from https://www.statista.com/statistics/617136/digital-population-worldwide/

De Haes, S., & Van Grembergen, W. (2017). An exploratory study into IT governance implementations and its impact on business/IT alignment. *Information Systems Management, 26*(2), 123–137. https://doi.org/10.1080/10580530902794786

EFRAG. (2022). *European sustainability reporting standard SEC1 sector classification standard: Working paper*. Retrieved May 11, 2022, from https://www.efrag.org/News/Project-572/EFRAG-publishes-today-the-next-set-of-PTF-ESRS-Cluster-Working-Papers

Eisenstat, Y. (2021). How to hold social media accountable for undermining democracy. *Harvard Business Review*. Retrieved June 11, 2022, from https://hbr.org/2021/01/how-to-hold-social-media-accountable-for-undermining-democracy

European Commission. (2022a). *Questions and answers Digital Markets Act*. Retrieved October 9, 2023, from https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_20_2349/QANDA_20_2349_EN.pdf

European Commission. (2022b). *Tackling online disinformation*. Retrieved January 18, 2024, from https://digital-strategy.ec.europa.eu/en/policies/online-disinformation

European Parliament. (2021). *Social media and democracy: We need laws, not platform guidelines*. Retrieved May 10, 2022, from https://www.europarl.europa.eu/news/en/headlines/society/20210204STO97129/social-media-and-democracy-we-need-laws-not-platform-guidelines

Facebook. (2021a). *Company info*. Retrieved June 20, 2022, from https://about.fb.com/company-info/

Facebook. (2021b). *COVID-19 and vaccine policy updates & protections*. Retrieved August 11, 2022, from https://www.facebook.com/help/230764881494641

Facebook. (2021c). *Environmental, social and governance FAQs*. Retrieved August 11, 2022, from https://investor.fb.com/esg-resources/frequently-asked-questions-esg/default.aspx

Facebook. (2021d). *Facebook brand resource center*. Retrieved July 7, 2021, from https://en.facebookbrand.com/

Facebook. (2021e). *How can I use Facebook to stay updated about the coronavirus (COVID-19)?* Retrieved August 11, 2022, from https://www.facebook.com/help/231416334748066/?helpref=search&query=COVID-19&search_session_id=8d2268c131b66b9478d88eebd06d648a&sr=2

Facebook. (2021f). *Our approach to misinformation*. Retrieved May 11, 2022, from https://transparency.fb.com/features/approach-to-misinformation/

Facebook. (2021g). *Promoting safety and expression*. Retrieved August 12, 2022, from https://about.facebook.com/actions/promoting-safety-and-expression/

Facebook. (2021h). *Timeline: Action against COVID-19*. Retrieved August 12, 2022, from https://about.facebook.com/actions/responding-to-COVID-19/

Fukuyama, F., & Grotto, A. (2020). Comparative media regulation in the United States and Europe. In N. Persily & J. A. Tucker (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform* (pp. 199–219). Cambridge University Press. https://doi.org/10.1017/9781108890960

Geeng, C., Yee, S., & Roesner, F. (2020). Fake news on Facebook and Twitter: Investigating how people (don't) investigate. In R. Bernhaupt, F. Mueller, D. Verweij, & J. Andres (Eds.), *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). New York, NY, USA: ACM. https://doi.org/10.1145/3313831.3376784

Ghosh, D. (2021). Are we entering a new era of social media regulation? *Harvard Business Review*. Retrieved August 11, 2022, from https://hbr.org/2021/01/are-we-entering-a-new-era-of-social-media-regulation

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research. *Organizational Research Methods, 16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Gooch, A. (2020). *Fighting disinformation: A key pillar of the COVID-19 recovery*. OECD Forum. Retrieved May 8, 2022, from http://www.oecd-forum.org/posts/fighting-disinformation-a-key-pillar-of-the-COVID-19-recovery

Google. (2006). *Google to acquire YouTube for $1.65 billion in stock*. Retrieved May 8, 2022, from http://googlepress.blogspot.com/2006/10/google-to-acquire-youtube-for-165_09.html

GRI. (2014). *GRI G4 media sector disclosures*. Retrieved May 8, 2022, from https://www.globalreporting.org/search/?query=G4+Media+Sector

Gunningham, N., & Rees, J. (1997). Industry self-regulation: An institutional perspective. *Law & Policy, 19*(4), 363–414. https://doi.org/10.1111/1467-9930.t01-1-00033

Héroux, S., & Fortin, A. (2014). Exploring IT dependence and IT governance. *Information Systems Management, 31*(2), 143–166. https://doi.org/10.1080/10580530.2014.890440

Horn, N. (2021). *Grundlagen der digitalen Ethik – eine normative Orientierung in der vernetzten Welt*. Stiftung Datenschutz. Leipzig. Retrieved August 10, 2022, from https://stiftungdatenschutz.org/fileadmin/Redaktion/Dokumente/Digitale_Ethik/SDS_Broschuere_Digitale_Ethik_Download.pdf

Huber, T. L., Kude, T., & Dibbern, J. (2017). Governance practices in platform ecosystems: Navigating tensions between cocreated value and governance costs. *Information Systems Research, 28*(3), 563–584. https://doi.org/10.1287/isre.2017.0701

Jin, K. X. (2020). *Keeping People Safe and Informed About the Coronavirus*. Facebook. Retrieved July 14, 2021, from https://about.fb.com/news/2020/12/coronavirus/

Jørgensen, R. F., & Zuleta, L. (2020). Private governance of freedom of expression on social media platforms: EU content regulation through the lens of human rights standards. *Nordicom Review, 41*(1), 51–67. https://doi.org/10.2478/nor-2020-0003

Kompetenzzentrum Öffentliche IT (2016). *Digitalisierung des Öffentlichen*. Berlin. Retrieved May 11, 2022, from https://www.oeffentliche-it.de/documents/10181/14412/Digitalisierung+des+%C3%96ffentlichen

Kourula, A., Moon, J., Salles-Djelic, M. L., & Wickert, C. (2019). New roles of government in the governance of business conduct: Implications for management and organizational research. *Organization Studies, 40*(8), 1101–1123. https://doi.org/10.1177/0170840619852142

Lahat, L., & Sher-Hadar, N. (2020). A threefold perspective: Conditions for collaborative governance. *Journal of Management and Governance, 24*(1), 117–134. https://doi.org/10.1007/s10997-019-09465-1

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science, 359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lindman, J., Makinen, J., & Kasanen, E. (2023). Big Tech's power, political corporate social responsibility and regulation. *Journal of Information Technology, 38*(2), 144–159. https://doi.org/10.1177/02683962221113596

Lovell, T. (2020). NHS joins forces with tech firms to stop the spread of COVID-19 misinformation. *Healthcare IT News*. Retrieved August 10, 2022, from https://www.healthcareitnews.com/news/emea/nhs-joins-forces-tech-firms-stop-spread-covid-19-misinformation

Lütjen, T. (2016). *Die Politik der Echokammer: Wisconsin und die ideologische Polarisierung der USA*. Transcript Verlag. https://doi.org/10.14361/9783839436073

Marin, L. (2020). Three contextual dimensions of information on social media: Lessons learned from the COVID-19 infodemic. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-020-09550-2

Marin, L. (2021). Sharing (mis) information on social networking sites. An exploration of the norms for distributing content authored by others. *Ethics and Information Technology, 23*(3), 363–372. https://doi.org/10.1007/s10676-021-09578-y

Meta. (2022). *Form 10-K: Annual report pursuant to section 13 or 15(d) of the securities exchange act of 1934 for the fiscal year ended December 31, 2021*. Retrieved September 27, 2022, from https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/14039b47-2e2f-4054-9dc5-71bcc7cf01ce.pdf

Monaghan, S., Tippmann, E., & Coviello, N. (2020). Born digitals: Thoughts on their internationalization and a research agenda. *Journal of International Business Studies, 51*, 11–22. https://doi.org/10.1057/s41267-019-00290-0

Mosseri, A. (2017). *Working to stop misinformation and false news*. Meta. Retrieved September 27, 2022, from https://www.facebook.com/formedia/blog/working-to-stop-misinformation-and-false-news

Muric, G., Wu, Y., & Ferrara, E. (2021). COVID-19 vaccine hesitancy on social media: Building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. *JMIR Public Health and Surveillance, 7*(11), e30642. https://doi.org/10.2196/30642

Nolan, R., & McFarlan, F. W. (2005). Information technology and the boards of directors. *Harvard Business Review, 83*(10), 96–106.

Oversight Board. (2021). *Case decision 2020–006-FB-FBR*. Retrieved May 11, 2022, from https://oversightboard.com/decision/FB-XWJQBU9A/

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.

Parker, G. G., van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. W. W. Norton & Company.

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020a). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science, 66*(11), 4921–5484. https://doi.org/10.1287/mnsc.2019.3478

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020b). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science, 31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Reuters Institute for the Study of Journalism. (2021a). *Digital news report—interactive—source of news: Social media*. Retrieved May 3, 2021, from https://www.digitalnewsreport.org/interactive/.

Reuters Institute for the Study of Journalism. (2021b) *Global active usage penetration of leading social networks as of February 2021*. In Statista. Retrieved May 11, 2022, from https://www.statista.com/statistics/274773/global-penetration-of-selected-social-media-sites/

Ridder, H.-G. (2020). *Case study research: Approaches, methods, contribution to theory* (2nd ed.). Rainer Hampp Verlag.

Roberts, J. J. (2020). Facebook's new tool to stop fake news is a game changer—if the company would only use it. *Fortune*. Retrieved September 19, 2022, from https://fortune.com/2020/10/18/facebook-tool-stop-fake-news-viral-content-review-system-fb-business-model/

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science, 7*(10), 201199. https://doi.org/10.1098/rsos.201199

Rosen, G. (2020). *An update on our work to keep people informed and limit misinformation about COVID-19*. Meta. Retrieved September 10, 2021, from https://about.fb.com/news/2020/04/COVID-19-misinfo-update/

Rosen, G. (2021). *How we're tackling misinformation across our apps*. Meta. Retrieved September 10, 2021, from https://about.fb.com/news/2021/03/how-were-tackling-misinformation-across-our-apps/

Royakkers, L., Timmer, J., Kool, L., & van Est, R. (2018). Societal and ethical issues of digitization. *Ethics and Information Technology, 20*(2), 127–142. https://doi.org/10.1007/s10676-018-9452-x

SASB. (2018). *Internet media & services—Sustainability accounting standard*. Retrieved June 17, 2021, from https://www.sasb.org/wp-content/uploads/2018/11/Internet_Media_Services_Standard_2018.pdf

Schrempf-Stirling, J., & Wettstein, F. (2023). The mutual reinforcement of hard and soft regulation. *Academy of Management Perspectives, 37*(1), 72–90. https://doi.org/10.5465/amp.2022.0029

Srnicek, N. (2017). The challenges of platform capitalism: Understanding the logic of a new business model. *Juncture, 23*(4), 254–257. https://doi.org/10.1111/newe.12023

Steinert, S. (2020). Corona and value change. The role of social media and emotional contagion. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-020-09545-z

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS ONE, 13*(9), e0203958. https://doi.org/10.1371/journal.pone.0203958

Trittin-Ulbrich, H., Scherer, A. G., Munro, I., & Whelan, G. (2021). Exploring the dark and unexpected sides of digitalization: Toward a critical agenda. *Organization, 28*(1), 8–25. https://doi.org/10.1177/1350508420968184

Turillazzi, A., Taddeo, M., Floridi, L., & Casolari, F. (2023). The digital services act: An analysis of its ethical, legal, and social implications. *Law, Innovation and Technology, 15*(1), 83–106. https://doi.org/10.1080/17579961.2023.2184136

Twitter. (2021a). *Coronavirus: Staying safe and informed on Twitter*. Retrieved May 11, 2022, from https://blog.twitter.com/en_us/topics/company/2020/covid-19

Twitter. (2021b). *COVID-19 misleading information policy*. Retrieved May 11, 2022, from https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy

Twitter. (2021c). *Glossary*. Retrieved May 12, 2022, from https://help.twitter.com/en/resources/glossary

Twitter. (2021d). *Twitter's services, corporate affiliates, and your privacy*. Retrieved May 12, 2022, from https://help.twitter.com/en/rules-and-policies/twitter-services-and-corporate-affiliates

Twitter. (2022). *Form 10-K: annual report pursuant to section 13 or 15(d) of the securities exchange act of 1934 for the fiscal year ended December 31, 2021*. Retrieved September 27, 2022, from https://s22.q4cdn.com/826641620/files/doc_financials/2021/ar/FiscalYR2021_Twitter_Annual_-Report.pdf on

Twitter Safety. (2020). *COVID-19: Our approach to misleading vaccine information*. Retrieved September 27, 2022, from https://blog.twitter.com/en_us/topics/company/2020/covid19-vaccine

Twitter Safety. (2021). *Updates to our work on COVID-19 vaccine misinformation*. Retrieved May 11, 2022, from https://blog.twitter.com/en_us/topics/company/2021/updates-to-our-work-on-COVID-19-vaccine-misinformation

van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.

van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Van Grembergen, W., & De Haes, S. (2009). *Enterprise governance of information technology*. Springer. https://doi.org/10.1007/978-0-387-84882-2

Van Maanen, J., Sørensen, J. B., & Mitchell, T. R. (2007). The interplay between theory and method. *Academy of Management Review, 32*(4), 1145–1154. https://doi.org/10.5465/amr.2007.26586080

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

We Are Social, DataReportal & Meltwater. (2023). Most popular social networks worldwide as of January 2023, ranked by number of monthly active users (in millions). In Statista. Retrieved June 20, 2023, from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Wilkin, C. L., & Chenhall, R. H. (2020). Information technology governance: Reflections on the past and future directions. *Journal of Information Systems, 34*(2), 257–292. https://doi.org/10.2308/isys-52632

World Health Organization. (2020). *COVID-19 Mythbusters*. Retrieved August 11, 2022, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters

World Health Organization. (2021a). *Coronavirus disease (COVID-19) - Q&A*. Retrieved May 11, 2022, from https://www.who.int/news-room/q-a-detail/coronavirus-disease-COVID-19

World Health Organization. (2021b). *Let's flatten the infodemic curve*. Retrieved May 4, 2023, from https://www.who.int/news-room/spotlight/let-s-flatten-the-infodemic-curve

World Health Organization. (2022). *WHO Coronavirus (COVID-19) dashboard*. Retrieved September 11, 2022, from https://covid19.who.int

YouTube. (2019). *The four Rs of responsibility, part 2: Raising authoritative content and reducing borderline content and harmful misinformation*. YouTube Official Blog. Retrieved September 10, 2021, from https://blog.youtube/inside-youtube/the-four-rs-of-responsibility-raise-and-reduce/

YouTube. (2021a). *COVID-19 medical misinformation policy*. YouTube Help. Retrieved April 11, 2022, from https://support.google.com/youtube/answer/9891785?hl=de&hl=en&ref_topic=9282436

YouTube. (2021b). *How does YouTube combat misinformation?—borderline content*. Retrieved April 5, 2022, from https://www.youtube.com/intl/en_us/howyoutubeworks/our-commitments/fighting-misinformation/#borderline-content

YouTube. (2021c). *How does YouTube combat misinformation?—determining misinfo*. Retrieved April 5, 2022, from https://www.youtube.com/intl/en_us/howyoutubeworks/our-commitments/fighting-misinformation/#determining-misinfo

YouTube. (2021d). *How does YouTube combat misinformation?—raising quality info*. Retrieved April 5, 2022, from https://www.youtube.com/intl/en_us/howyoutubeworks/our-commitments/fighting-misinformation/#raising-quality-info

YouTube. (2021e). *Progress on managing harmful content—Detection Source*. Retrieved April 5, 2022, from https://www.youtube.com/intl/en_us/howyoutubeworks/progress-impact/responsibility/#detection-source

YouTube. (2021f). *Progress on managing harmful content—Removal by views*. Retrieved April 5, 2022, from https://www.youtube.com/intl/en_us/howyoutubeworks/progress-impact/responsibility/#removal-by-views

YouTube. (2021g). *Unsere Fortschritte beim Umgang mit schädlichen Inhalten—Entfernte Videos nach Aufrufen*. Retrieved April 5, 2022, from https://www.youtube.com/intl/ALL_de/howyoutubeworks/progress-impact/responsibility/#removal-by-views

Zarsky, T. (2016). The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values, 41*(1), 118–132. https://doi.org/10.1177/0162243915605575

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs.

Zuboff, S. (2022). Surveillance capitalism or democracy? The death match of institutional orders and the politics of knowledge in our information civilization. *Organization Theory, 3*(3), 1–79. https://doi.org/10.1177/26317877221129290

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Lina Warnke is a research associate at the Helmut Schmidt University Hamburg (Germany). Her research interests focus on sustainability reporting, corporate governance and corporate social responsibility.

Anna-Lena Maier is a project and change manager at the ma-co maritimes competenzcentrum GmbH and oversees transformation projects related to digitalization and the energy transition. She also is a lecturer and researcher covering topics such as business and government, business and peace, political CSR and responsible management. She received her PhD from the University of Hamburg (Germany), where she taught several courses on (international) strategic management, business ethics and CSR.

Dirk Ulrich Gilbert is a professor of business ethics and management at the University of Hamburg (Germany). He received his PhD from the University of Frankfurt and held positions at the University of New South Wales and the University of Nuremberg. His most recent research focuses on international accountability standards, labour rights in global supply chains, political CSR, and responsible management education. He has published in internationally acclaimed journals, such as Business Ethics Quarterly, Business and Society, Academy of Management Learning and Education, Journal of Management Inquiry, Management International Review, and the Journal of Business Ethics.

Authors and Affiliations

Lina Warnke[1] · **Anna-Lena Maier**[2] · **Dirk Ulrich Gilbert**[3]

✉ Lina Warnke
lina.warnke@hsu-hh.de

Anna-Lena Maier
anna-lena.maier@ma-co.de

Dirk Ulrich Gilbert
dirk.gilbert@uni-hamburg.de

[1] Faculty of Economics and Social Sciences, Helmut-Schmidt-Universität / Universität Der Bundeswehr Hamburg, Holstenhofweg 85, 22043 Hamburg, Germany

[2] Ma-Co Maritimes Competenzcentrum, Köhlbranddeich 30, 20457 Hamburg, Germany

[3] Department of Socioeconomics, Faculty of Business, Economics and Social Sciences, Universität Hamburg, Von-Melle-Park 9, 20146 Hamburg, Germany