

Platform governance under the Digital Services Act: a perspective on disinformation

Rita Gsenger^{a,b}

^aResearch Group Norm Setting and Decision Processes, Weizenbaum Institute, Berlin, Germany; ^bInstitute of Journalism and Communication Studies, Free University Berlin, Berlin, Germany

ABSTRACT

Online platforms have become essential infrastructures for communication and commerce, playing a central role in content governance and shaping public discourse. While their accessibility fosters communication, it also facilitates the spread of harmful information, including disinformation. To address these risks and enhance transparency and accountability, the European Union (EU) introduced the Digital Services Act (DSA), which mandates specific content moderation obligations for platforms operating within its jurisdiction. This study examines how 27 online platforms govern disinformation, employing a mixed-methods analysis of their Terms of Service and Community Guidelines. The analysis shows that ‘misleading’ content is the most frequently regulated category, with very large online platforms (VLOPs) and social media platforms exhibiting the highest levels of regulation. Although a range of sanctions exists, platforms primarily rely on content and account removal, with limited mechanisms for user participation. The study identifies four main clusters of disinformation addressed by platforms: misleading content, imposter content, pseudoscience and conspiracy theories, and manipulation. These categories illustrate specific regulatory challenges, such as health-related disinformation and identity misrepresentation. The paper situates these findings within the context of regulatory frameworks like the Digital Services Act (DSA), emphasizing the need for further research on enforcement practices and the impact of content governance.

ARTICLE HISTORY



Received 25 April 2025
Accepted 8 November 2025


KEYWORDS

Platform governance; content moderation; Digital Services Act; disinformation; regulation; social media;

Introduction

Platforms are increasingly important in our lives, as they often provide critical societal infrastructures (Plantin et al., 2018) and shape cultural production (Nieborg and Poell, 2018). Online platforms are defined as sociotechnical structures ‘designed to organize interactions between users’ (Van Dijck et al., 2018: 4) and ‘ecosystems, or [...] stand-

CONTACT Rita Gsenger  rita.gsenger@weizenbaum-institut.de  Institute of Journalism and Communication Studies, Free University Berlin, Hardenbergstr. 32, 10623, Berlin, Germany

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/1369118X.2025.2590561>.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

alone interconnected elements of content, services, application, and user communities that exist within a single corporate entity' (Flew, 2021, p. 72). Users (making decisions), governments (proposing and enforcing legislation), and the platforms themselves, who design and determine the architecture, have a cooperative responsibility for platform governance (Helberger et al., 2018). Platform governance can be understood as an 'approach necessitating an understanding of technical systems (platforms) and an appreciation for the inherently global arena within which these platform companies' function' (Gorwa, 2019b, p. 857), whereby platforms have been even called the 'new governors' (Klonick, 2018). Moreover, the relations between users, platforms operators and what has been called 'complementors' (advertisers, developers etc.) (Gorwa, 2019b, p. 857) are embedded in highly volatile power structures (Poell et al., 2019).

Platforms include various types of undesired content, such as, incivility, hate speech, extremism, and harassment (Gillespie, 2018). Especially on social media, disinformation persists due to economic and political incentives, human factors, and the lack of platform measures (Saurwein & Spencer-Smith, 2020). The European Union passed the Digital Services Act (DSA), which intends to increase platform accountability, transparency rules, influencing how platforms moderate content. The DSA is considered a landmark regulation of platform power that might have considerable influence outside the EU. That influence is often called the *Brussels effect* and was already observed regarding data protection regulations. However, the global influences of regulation content and speech are more limited due to the difference in understanding of what kind of content should be protected under free speech requirements (Bradford, 2020).

Attempting to harmonize European rules, the DSA includes a variety of measures for very large online platforms (VLOPs) and very large search engines (VLOSEs). Such an intermediary service needs to have more than 45 million monthly active users in the European Union (Art. 33 (1), DSA). It must mitigate systemic risks (Art. 34, DSA) – including disinformation campaigns (Rec. 83, 84, DSA). These mitigation measures include adapting content moderation rules (Art. 35 (1)) and applying and enforcing Terms and Conditions (Art. 14 (4)).

This study aims to understand the disinformation governance of platforms operating in Europe, including social media, search engines and product marketplaces. In the following, the current research on disinformation and platform governance is presented to highlight the disagreements regarding governance of legal but undesired speech. A study of the Terms of Service and Community Guidelines of 27 intermediaries shows, how platforms govern these types of speech and emphasize the need for more diverse and user-centric content governance mechanisms.

Literature review

Understanding disinformation

To understand the effectiveness of platform rules in countering disinformation, the object of regulation must be understood. The DSA intends to mitigate the economic and societal risks of an unchecked internet and the dissemination of disinformation (Rec. 2). Furthermore, a lack of policy intervention in the increasing dissemination of disinformation content might lead to risks to fundamental rights of information and

expression (Marsden et al., 2020). Even as illegal content is a significant risk, the DSA does not harmonize which behavior or content counts as illegal. The scope of illegal content should cover the extent of offline illegality, including products (such as narcotics), offers (such as scams) or content (such as child sexual abuse material). Therefore, the Member States need to develop the details of the regulatory scope. For illegal content, interventions concerning the content are required. That is different for expressions that are legal (Husovec, 2024a) and might be ‘lawful, but awful’ (Keller, 2022). Compared to illegal content and services, the term disinformation is less developed and no uniform definition was established (Ó Fathaigh et al., 2021). However, some content that is morally or normatively condemned (e.g., racism), might still be legal as it is protected by freedom of speech (Keller, 2022). The DSA’s recitals refer to disinformation campaigns relating to public health, public security, public discourse, political participation and equality, as these might bear risks to society (Rec. 83, DSA).

Disinformation categories are defined in the literature in various ways, often as a taxonomy with multiple sub-categories. These might include rumors, hoaxes, trolling, conspiracies, and manipulative and misleading content (Kapantai et al., 2021). Others define disinformation as the ‘motivated faking of news’ (Marsden et al., 2020, p. 2). However, no unified definition has emerged regarding the inclusion and exclusion of categories, and various terms have been used in research and the public discourse to describe the phenomenon. Most notably, Wardle and Derakhshan’s 2017 typology of the information disorder differentiates between disinformation, misinformation and malinformation. Disinformation is defined by the European Commission as ‘the creation, presentation and dissemination of verifiably false or misleading information for the purpose of economic gain or intentionally deceiving the public, and which may cause public harm’ (European Commission, 2018, p. 10). Even though the Commission considers the term as a regulatory object, it is also often used as a rhetorical weapon, for instance, to discredit political opponents (Galantino, 2023). Misinformation is defined as inaccurate information that is held confidently but ignorantly (Vraga & Bode, 2020). Furthermore, misinformation sparks outrage, which facilitates its circulation and often shared without being reading first (McLoughlin et al., 2024).

Malinformation describes information intended to harm (Wardle & Derakhshan, 2017). However, the term did not sustain a broader application. Another term often used in this context is fake news. Fake news describes false information the media or journalists (or someone pretending to be media) disseminate (Jaster & Lanus, 2018). However, disinformation, as the deliberate and organized distribution of false information to cause harm, is most crucial from the perspective of the DSA, which intends to regulate ‘disinformation campaigns’ (Rec. 68, 83, 88, DSA). The intention of disinformation is most often to manipulate (Benkler et al., 2018), to harm (Wardle & Derakhshan, 2017), or to gain some advantage, such as money (Herasimenka et al., 2023) or political influence (Keller et al., 2020). However, the intention is often unclear and inaccessible (Altay et al., 2023) and the effect was overestimated by the media, especially after the Cambridge Analytica scandal in 2016 (Karpf, 2019), as interaction with content does not equate conviction (Bail et al., 2020). The content of a larger disinformation campaign might not be directly harmful and not singular posts, but the larger narratives these posts fit into and reinforce are crucial (Wardle, 2023). Moreover, the discourse changes and actors that undermine the trust in institutions become stronger (Wardle, 2020).

Platform ecosystems and governance

The DSA seeks to regulate ‘intermediary services’ (Art. 3), which includes online platforms and search engines (Husovec, 2024b). An online platform is a hosting service that ‘at the request of a recipient of the service, stores and disseminates information to the public’ (Art. 3, lit. i, DSA).

Platforms are often classified according to their business models, which focus on their various functions to generate revenue (see, for instance, Derave et al., 2024; Nooren et al., 2018). They can also be understood according to infrastructure and sectorial functions (like health, food, and housing) (Van Dijck et al., 2018). However, not all platform types are relevant for disinformation regulation and content governance. Only platforms that host and distribute pertinent information and have some sort of network effect are essential regarding the distribution of disinformation. That excludes, for instance taxi services that might distribute wrong information about a driver.

Platforms are governed transnationally with what Gorwa (2019, p. 5) has called the ‘governance triangle’. The triangle includes NGOs (e.g., civil society organizations, researchers), firms (companies and industry associations) and states (including supranational groups). Similarly, Balkin (2018) analyzes free speech as pluralist and with regard to these three stakeholders. Platforms as intermediaries do not produce content, but they connect the users generating content with the users seeking content, i.e., they mediate (Gillespie, 2018). Moreover, they are usually not entirely liable for the content that their users create and provide (Gorwa, 2019b). The development from publishers to platforms altered the infrastructure of free expression significantly, so governments regulate in cooperation with private companies (Balkin, 2018). Platform companies assume roles similar to those of governments, making decisions that affect areas such as foreign policy or crime control (Eichensehr, 2019). Platforms, however, do not have state-like sovereignty (Eichensehr, 2019) and they are not neutral actors (Gillespie, 2018), even though they often assume a posture of neutrality or might strive towards it (Eichensehr, 2019). Moreover, users can be considered not only citizens of nation-states, but also ‘participants of a transnational social environment’ (Muniz Da Conceição, 2025, p. 19). The DSA aims to create a more trusted online environment by removing illegal content, holding perpetrators accountable and enabling users challenge platforms if they have been wrongly accused (Husovec, 2024b). Platforms engage in private governance (Balkin, 2018) by, for instance, monitoring, editing, and deleting content. Moreover, VLOPs such as Apple might use deplatforming as a measure of depriving fringe platforms access to infrastructure (e.g., by banning them from the app store) (Monaci, 2024). Various industry initiatives establishing self-regulatory measures were adopted, often responding to government pressure or to improve their public image (Gorwa, 2019). Overall, governments still regulate speech by imposing fines or injunctions, however, they also regulate speech by regulating internet infrastructure (Balkin, 2014; 2018). Therefore, their role as intermediaries is doubted even though the process of (partly) automated ex-post moderation cannot be equated to the editorial process of newspapers, movies, books, and the like. Platform companies offer various services with differing business models and technical capabilities. Moreover, they vary according to size and impact. In that regard, proportionality is crucial as more users indicate more influence on public speech (Flew, 2021). That is reflected in the stricter rules of the DSA for very large online platforms and search engines.

Depending on their functionalities, the service providers have differing due diligence obligations according to the DSA. For instance, VLOPs must conduct risk management (Art. 35) and audits (Art. 37) and make their rules and enforcement transparent (Art. 42). Hosting services that do not make information available to the public must provide notice and action mechanisms (Art. 16). The DSA foresees voluntary Codes of Conduct that should encourage compliance with measures against illegal content and systemic risks (Art. 45, DSA). The Codes enable the provision of evidence regarding the details and scope of risk assessments (Husovec, 2024a). The 2022 Code of Practice on Disinformation (European Commission, 2022) was integrated into the DSA in February 2025 and was signed by various platforms, including VLOPs such as TikTok, Meta, and YouTube (European Commission, 2023b). The DSA establishes a co-regulatory system: the risk management for VLOPs and VLOSEs is overseen by the European Commission; other providers are overseen by the Member States. Due diligence obligations are supervised conjointly by the European Commission and the Member States. Orders regarding illegal content are still made by national authorities, for which the DSA only outlines optional requirements (Husovec, 2024a). The Code was criticized for not ensuring enforcement or providing a mechanism to monitor process (Galan-tino, 2023).

Platforms have various possibilities to moderate content. The platforms publish prohibited content rules in their Terms of Service and Community Guidelines. However, these depend on the platform's values and are frequently imposed on users (Scharlach et al., 2023). Previous studies of Terms of Service have focused on various topics such as privacy (Kaur et al., 2018), consumer goods (Weiger et al., 2020), or values (Scharlach et al., 2023). Furthermore, research has focused on user perceptions, readability, and accessibility of terms of service (McDonald & Cranor, 2009). However, differences between platforms regarding their governance remain understudied. This first research question is:

RQ1. How does platform governance differ between social media platforms, search engines, and product marketplaces?

The governance of undesired content or incivility has mainly been studied in the context of hate speech (Obar & Oeldorf-Hirsch, 2020). However, the various forms of disinformation governed by platforms and their understanding have not been investigated. In light of the DSA, understanding these governance mechanisms is crucial, as they might inform compliance assessments and shed light on the platforms' understanding of disinformation. This leads to the second research question:

RQ2. How are different types of disinformation sanctioned across different types of platforms?

As outlined above, the definition of disinformation remains discussed and its various types are complex to assess (Kapantai et al., 2021). It is essential to consider platforms understanding as they influence public discourse by setting the rules for infrastructure and public spaces. Accordingly, the third research question is:

RQ3. How can the types of frequently moderated disinformation categories be understood?

Methods

To answer the research questions, I used a mixed-methods approach to broadly understand the material with a correspondence analysis to select crucial aspects for a structured content analysis (Mayring, 2015).

Data corpus

Intermediaries, which have the highest economic significance (volume of traffic and trade) and power over users, which is measured according to users' business dependence on the platforms (European Commission, 2021), were selected. That list is extended to include the intermediaries designated as VLOPs and VLOSEs by the European Commission (2023a). The combination of these metrics allows for a more diverse set of platform rules that not only includes user numbers but also economic significance.

To assemble the corpus of analysis, the Terms of Service and Community Guidelines of 27 intermediaries, which adhere to the EU jurisdiction, were collected. The first round of data collection was conducted in October and November 2022. The codebook was developed in 2023 and the data was updated in December 2023 and January 2024 (with a replacement of the data by Twitter/X). The corpus is grouped into three categories: social media, search engines, and product marketplaces (each intermediary was assigned a code during data analysis designating the type: social media (SM), product marketplace (PM), or search engine (SE) and specifying the name). The corpus included 777.125 words, including seven product market place platforms, four search engines and 17 social media platforms.

Developing the coding frame

The coding frame was developed based on a literature review, including the categories object of regulation, rules, procedures and sanctions, which designate the content governance process. The *object of regulation* meta-category focused on the literature on disinformation, adapting the typology developed by Kapantai et al. (2021). They base their typology on a corpus of eight taxonomical frameworks found in research and additional literature from other stakeholders, refining them to focus on disinformation (by excluding some categories, such as satire and parody). The codebook's *platform rules* and *procedures* categories are based on Weiger et al. (2020), who developed a coding framework by open coding ToS of consumer product websites. The sanction meta-category was developed based on an online content governance taxonomy developed by Goldman (2021), assembling remedies employed by at least one platform. These include content regulation, account regulation, monetary sanctions and visibility restrictions.

To refine the codebook for this study, two researchers with backgrounds in law and communication sciences conducted a test coding of a third of the data. Categories not found in the data were removed, and inductive categories identified as salient were added. Coding was done using MaxQDA (VERBI Software, 2024).

Coding and analysis

Both researchers recoded the first third of the material until an intercoder reliability of Krippendorff $\alpha > 0.79$ was established. Subsequently, the material was divided and

coded by one researcher each. The coding resulted in 4,054 coded text passages. To answer RQ1, frequency of codes and the connections between the object of regulation and the sanctions are calculated. Subsequently, the most sanctioned categories of disinformation are selected using a correspondence analysis, and the results provide the basis for the qualitative analysis to answer RQ3.

Four clusters (misleading, imposter, pseudoscience and conspiracy theories and manipulation) of disinformation across all three platform types are identified as the most sanctioned and extracted for qualitative analysis. The extracted text passages were organized according to contextual information provided in the data. For instance, the text passage ‘misleading content, such as spam,’ was categorized as belonging to the cluster misleading and the dimension spam. The quantitative and qualitative results are reported in the next section.

Results

Quantitative results: the platform governance process

To understand the differences in platform governance (RQ 1), comparisons were made between the categories social media, search engines and product marketplaces. Intermediaries were also compared according to size (VLOPs, VLOSEs and smaller platforms). [Figure 1](#) shows the differences in content types that are regulated.

Compared to other platform categories, regulations on VLOPs and social media platforms are the most detailed. Conspiratorial content and disinformation, in general, are regulated strongly, given that spam, for instance, falls under the category of ‘manipulation,’ which is less regulated in comparison but does constitute a considerable problem. A comparison of coded passages was conducted to answer how different types disinformation are sanctioned across platforms (RQ 2), summarized in [Figure 2](#). The types of

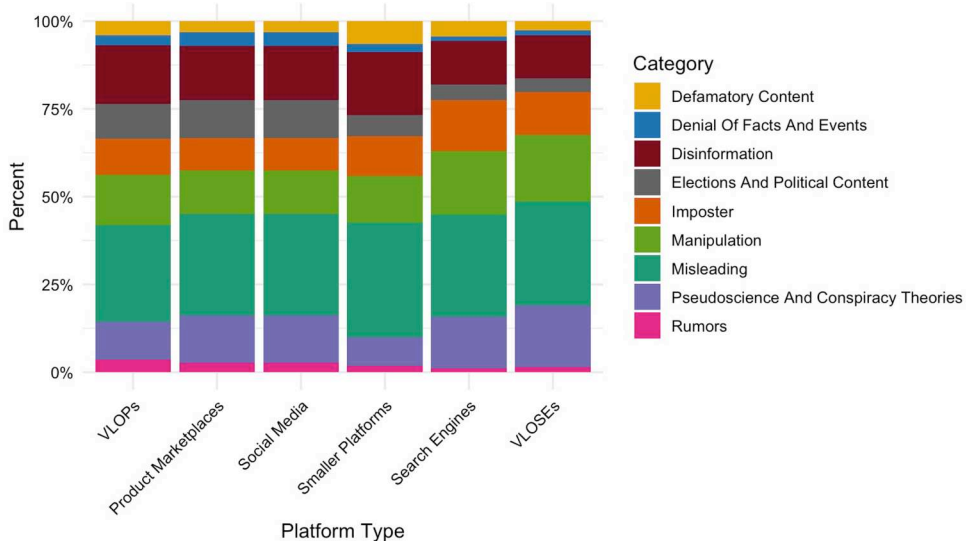


Figure 1. Frequency of disinformation types coded in platform types (See Annex for the full table).

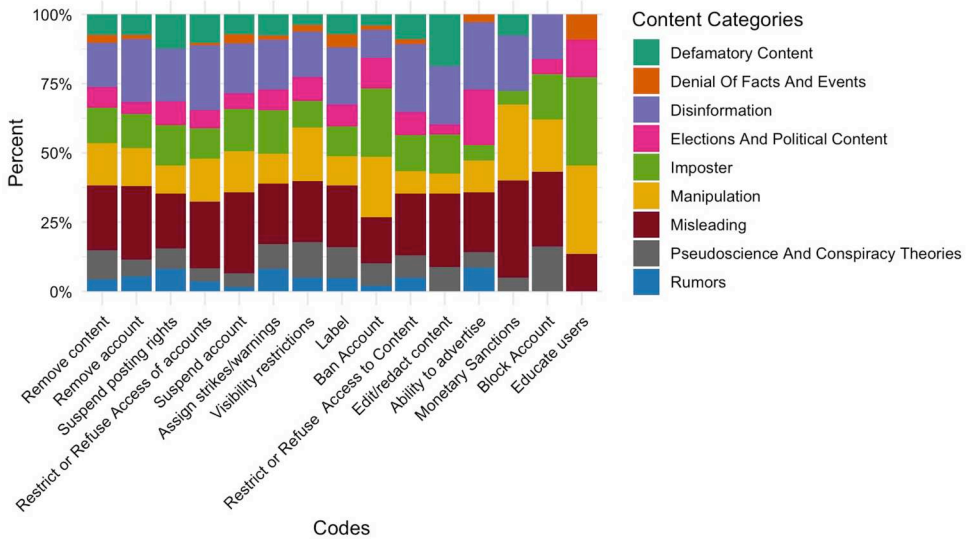


Figure 2. Most used types of sanctions of disinformation types (for the entire table, see the Annex).

sanctions for the disinformation are less varied than might be expected. Platforms have many options to sanction content but relocating content, removing credibility badges, outing/unmasking, community service, and shadowbanning (Goldman, 2021) were excluded from the analysis, as they were not found in the data. However, the reduction of visibility of accounts and content might include shadowbanning. Platforms do not use that term precisely and rarely explain the process in detail. Generally, most sanctions are communicated to the user in some way. However, the details of that communication vary considerably (see Figure 2). Deleting content is the most widely used sanction for disinformation. Other often-used measures to sanction disinformation are removing or suspending user accounts or restricting or refusing access to user accounts on the platform. The least used measure is to educate users, which refers to conversing with them about their behavior and giving them a chance to improve.

Additionally, the procedure of how content is governed needs to be explored to understand content governance better. Details about the procedure, which is how these sanctions are conducted and how the communication between platforms and users is organized, are shown in Table 1.

The categories ‘account and content regulation’ refer to the description of the process of sanctioning. Accordingly, platforms inform users about account regulation and, most often, content regulation (i.e., users receive a message if they are being blocked or their account is deleted). Most notably, VLOPs do not inform users much about the detection measures of disinformation they employ and appeal and reporting functions are not that prevalent. Furthermore, only some platforms offer user moderation possibilities, which refers to users being able to decide who they interact with on a platform. All platforms focus on excluding liability across all types of disinformation. Most notably, product marketplaces emphasize not being liable for products sold or reviews posted on their sites.

Table 1. Procedure of content moderation of platforms by platform types.

Category	VLOSEs	Smaller Platforms	VLOPs	Search Engines	Product Marketplaces	Social Media
Content Regulation	11 (22,45%)	39 (10,34%)	104 (16,51%)	19 (19,59%)	46 (12,11%)	89 (15,37%)
Account Regulation	7 (14,29%)	69 (18,30%)	137 (21,75%)	19 (19,59%)	81 (21,32%)	113 (19,52%)
Exit possibilities for Users	4 (8,16%)	23 (6,10%)	30 (4,76%)	7 (7,22%)	16 (4,21%)	34 (5,87%)
Appeal	3 (6,12%)	16 (4,24%)	46 (7,30%)	3 (3,09%)	18 (4,74%)	44 (7,60%)
Detection Measures	7 (14,29%)	18 (4,77%)	26 (4,13%)	8 (8,25%)	7 (1,84%)	36 (6,22%)
ToS can change any time	3 (6,12%)	29 (7,69%)	61 (9,68%)	10 (10,31%)	45 (11,84%)	38 (6,56%)
Liability	7 (14,29%)	89 (23,61%)	160 (25,40%)	21 (21,65%)	134 (35,26%)	101 (17,44%)
Reporting of misconduct	5 (10,20%)	48 (12,73%)	23 (3,65%)	7 (7,22%)	14 (3,68%)	55 (9,50%)
User moderation	0 (0,00%)	39 (10,34%)	12 (1,90%)	1 (1,03%)	6 (1,58%)	44 (7,60%)
Cooperation with law enforcement	2 (4,08%)	7 (1,86%)	31 (4,92%)	2 (2,06%)	13 (3,42%)	25 (4,32%)
SUM	49 (100%)	377 (100%)	630 (100%)	97 (100%)	380 (100%)	579 (100%)

Qualitative results: aspects of disinformation

The most sanctioned categories were extracted and analyzed qualitatively to answer RQ2. Which types of disinformation are most frequently moderated by online platforms and search engines? Four clusters of disinformation across all three platform types are identified. The categories were identified by extracting passages that used the words misleading, imposter, pseudoscience and conspiracy theories, and manipulation. The extracted text passages were subsequently organized according to their description. For instance, the text passage ‘misleading content, such as spam,’ was categorized as belonging to the cluster misleading and the dimension spam. The excluded dimensions could not be summarized under any specific topic but were still referred to by platforms using the cluster description. Table 2 provides an overview of the frequency of clusters and the extracted dimensions.

Misleading content and behavior

Platforms generally regulate misleading content in terms of the topics and, secondly, according to user behavior. Content that is misleading and regulated under the Terms of Services might be surprising (X), deceptive (Vimeo, Outbrain), factually inaccurate (Outbrain) or might be edited or combined to mislead users about events (TikTok). Platforms seek to address false authorships, for instance, regarding content or reviews (Facebook, Google Play). However, Facebook is the only platform that regulates the false authorship of news content. Otherwise, impersonation, for instance, by using fake accounts or creating events to impersonate others (Facebook, Instagram) or posing as someone else (X), is not permissible.

Some specific topics are mentioned repeatedly across platforms that are often misleading. These include climate change (Pinterest, Google Play), armed conflicts and war crimes (X), the voting process, election results and civic participation (Pinterest,

Table 2. Overview of clusters and frequency of topics. Duplicate passages were excluded.

	Search Engines 28	Social Media 190	Product Marketplaces 109
Misleading			
Content	14 (50%)	38 (20%)	18 (16%)
Spam	6 (21%)	32 (16%)	18 (16%)
Behavior	6 (21%)	17 (9%)	2 (2%)
Excluded	2 (7%)	103 (54%)	71 (65%)
Imposter	13	54	41
Impersonation	9 (69%)	20 (37%)	13 (31%)
Misrepresentation	1 (7%)	15 (27%)	9 (21%)
Excluded	3 (23%)	19 (35%)	19 (46%)
Pseudoscience and Conspiracy Theories	14	69	18
Health	3 (21%)	13 (18%)	5 (27%)
Vaccines	0	22 (31%)	0
Miracles	0	5 (7%)	3 (16%)
Conspiracies	0	15 (21%)	1 (5%)
Climate	0	4 (5%)	1 (5%)
Excluded	11 (78%)	10 (14%)	8 (44%)
Manipulation	15	63	61
Coordinated Behavior	2 (13%)	7 (11%)	2 (3%)
Images and Media	7 (46%)	15 (24%)	1 (1%)
Manipulation of the System	4 (26%)	11 (17%)	32 (52%)
Excluded	2 (13%)	30 (47%)	26 (42%)

Snapchat, Google Play). These types of content are mainly regulated on social media platforms and with most detail by Facebook, Instagram and YouTube. Product marketplaces do not include misleading content as part of their prohibited products to sell, except for Alibaba, which strictly regulates products against and concerning COVID-19. For some products, however, advertisements are restricted and should not be sold to children, such as products that claim extreme weight loss.

Content governance not only focuses on content that is not permissible but also on user behavior, which is equally important to consider. Misleading user behavior might include deceptive activities, including misrepresentation of affiliation or impersonation (Vimeo) to mislead other users or manipulate ratings and downloads (Apple App Store). That behavior includes primarily spam activities. Many spammers operate with e-mails or private messages (Ferrara, 2019). However, spam is prevalent on all types of platforms, such as content distributed by social bots (Latah, 2020), compromised accounts (Grier et al., 2010), astroturfing campaigns (Halperin, 2021) or clickbait news articles (Rubin, 2019). Any type of spam is usually designed to attract users with lower digital literacy levels (Redmiles et al., 2018). All investigated platforms consider spam detrimental to the user experience. Spam includes, for instance, clickbait (Google Shopping, Facebook) and misleading titles (YouTube). News organizations frequently use clickbait articles to increase website traffic (Bazaco et al., 2019). Spam mainly consists of mass messages, possibly automated, including bot networks (TikTok, Vimeo). However, these posts and messages are difficult to identify, even with automated means (Martini et al., 2021). For search engines, most notably Google, spam has also been a problem, and they have engaged in a so-called arms race with spammers who tried to optimize the search engine results in their favor to spread misinformation (Metaxa-Kakavouli & Torres-Echeverry, 2017).

Imposter

The imposter cluster is organized around inauthentic behavior, impersonation and misrepresentation. Impersonation is the most prevalent category, from not creating an

account for someone else or speaking for someone else without their permission (Instagram, Facebook), pretending to be someone else (Twitch), like a celebrity or an elected official (Tumblr), to not use others usernames (Instagram, Twitch). These practices are often prohibited due to their intent to deceive (Wikimedia, Reddit, Facebook, Bing) and they are frequently connected to fraud (Bing, Google Search). YouTube differentiates between two types of impersonation:

Channel impersonation: A channel that copies another channel's profile, background, or overall look and feel in such a way that makes it look like someone else's channel. The channel does not have to be 100% identical, as long as the intent is clear to copy the other channel. Personal impersonation: Content intended to look like someone else is posting it.

Misrepresentation regards the regulation of pretending to have a different affiliation with a person, organization or entity (Snapchat, Pinterest) or falsely representing the platform as a moderator (Reddit). If a user is misrepresenting their identity, they are, according to Facebook, '[u]sing a name that is not the authentic name you go by in everyday life' and 'using an inherently violating name, containing slurs', providing wrong information or having multiple accounts. These are connected to inauthentic behavior, as it refers to the process of misappropriating identities (X) or misleading people about 'identity purpose' or the 'origin of the represented identity' (Facebook).

Pseudoscience and conspiracy theories

The cluster is divided into health, vaccines, wrong treatment, climate, miracles and conspiracies. Pseudoscience and conspirational content were often directly mentioned regarding disinformation, especially connected with the misleading category. Therefore, considerable overlaps of the categories can be observed.

In the health cluster, general misinformation about COVID-19 is prevalent, for instance, 'products which have unfounded medical claims or claims related to COVID-19' (Aliexpress) or contradicting the remedies against and spread of COVID-19 (YouTube). Other general health topics focus on weight loss products (Google Shopping), the denial of AIDS (Google Shopping) or anything contradicting scientific consensus and best practices (Google Search). In a second cluster, miracles are restricted, which includes "'miracle cures" for medical ailments such as arthritis, diabetes, Alzheimer's disease, or cancer' and '[p]roducts that claim to be "cure-all" for several diseases' (Google Shopping). Moreover, miracle weight-loss products (Facebook, Instagram) or other miracle cures that can heal everything (X, LinkedIn, Taboola) are not permissible. Finally, any product that 'undermine[sic!] religions, or promotes cults and superstitions' (Aliexpress) is not allowed.

Similarly, disinformation about vaccines is mentioned by the majority of platforms, including anti-vaccine advocacy (Google Shopping), anti-vax information (VK) or advice (LinkedIn). Furthermore, many platforms detail that suggesting alternative treatments to vaccines, such as Vitamin C (LinkedIn) or relying on natural immunity (Vimeo), are not permissible. Spreading information about dangerous consequences of vaccines, such as sudden infant death syndrome, autism or the disease they are against (Facebook, Instagram). Public health is considered a collective right that might be at risk due to online content (Art. 34, DSA). Furthermore, spreading disinformation about vaccines can be detrimental to public health, for instance, in Nigeria, where an anti-vaccination

campaign led to the murders of vaccine workers in 2013 and finally also to the ban of polio vaccination by law (Berman, 2020). Other examples of information contradicting scientifically proven facts regarding climate change (YouTube, TikTok, Pinterest), and the misrepresentation of solutions to climate change (Pinterest). Climate change is argued to threaten health, well-being, security, and fundamental rights. It can, therefore, be understood as a systemic risk under the DSA (Griffin, 2023).

Conspiracies are often not allowed, with examples such as ‘Pizzagate, QAnon, StopTheSteal’ (YouTube), which are considered harmful. Furthermore, Google Shopping does not allow users to claim to be the victim of a conspiracy. Conspiracy theories, like QAnon can lead to political radicalization and undermine public security, as shown by the Capitol Hill riots on January 6 (Moskalenko et al., 2023). Moreover, QAnon has been interpreted as an extremist movement connected with anti-government protests (Miotto & Droogan, 2024).

Manipulation

The data in the manipulation cluster mainly focuses on two types: Coordinated inauthentic behavior to manipulate other users. That is, for instance, explicitly called out by Instagram and Facebook. That type of behavior is akin to disinformation campaigns regulated by the DSA (Rec. 69, 83, 88, DSA), which might be advanced by states (Iosifidis & Nicoli, 2021) to internally discredit politicians or the media (Hameleers, 2023), or the interference by foreign states (Wagnsson & Barzanje, 2021). Such coordinated campaigns were attempted, for instance, during the 2016 US presidential elections (Benkler et al., 2018). However, Meta seems to be the only platform company considering inauthentic coordinated behavior in a larger sense. Other platforms focus on specific behaviors such as manipulating content for one particular point of view (Wikimedia) or manipulating the algorithm with affiliate links (Pinterest). For product marketplaces, manipulative behavior is regulated regarding the site’s integrity (Aliexpress).

As a second cluster, manipulated media content, such as deepfakes or content produced by generative AI, is considered. That includes ‘no synthetic media’ of ‘public figures (contains the likeness (visual or audio) of a real person, including: (1) a young person, (2) an adult private figure, and (3) an adult public figure when used for political or commercial endorsements, or if it violates any other policy)’ (TikTok). That policy is connected to hate speech, sexual exploitation and harassment (TikTok). No manipulated images or media (Instagram, X, Google Play, Google Search), or ‘highly deceptive manipulated media’ (Facebook, Instagram) should be shared not to cause harm or confuse people (X). That explicitly includes AI produced content, that seems authentic (Instagram) or ‘distorts real-life events’ (LinkedIn). Some platforms suggest a watermark for such content (Google Play).

Discussion

This study investigates the content governance of disinformation on online platforms and search engines.

The process of moderation did not vary strongly across platform sizes and types. Most platforms have a form of communication to inform users that their account or content is sanctioned. The rules do not specify the level of detail of these communications. Appeals

can sometimes be made; these, however, differ according to the severity of the violations. Platforms must publish transparency reports, detailing their content moderation efforts (Art. 42, DSA) and they are required to notify users about their decisions and provide redress mechanisms (Art. 16(5)). As platforms assume roles similar to governments regarding free speech, they might also enable an expansion of user rights, even countering governmental tendencies (Muniz Da Conceição, 2025). However, user moderation is generally not incentivized, for instance, regarding recommender systems that contribute to platform revenue. From the platforms' perspective, uniform rules are necessary, to attract advertisers. The DSA provides empowerment as it enables users to decide for themselves what information they consume (Husovec, 2024b).

Many platforms also adopt a system of strikes. Most notably, Aliexpress has a complicated system that ranks users according to the strikes they receive. The system does not seem very transparent, and the violation of rules depends on the platforms' interpretations. Others, like YouTube, warn users whenever a strike is issued, and users can appeal. Though AI technologies often support content moderation (Gorwa et al., 2020), many platforms are not transparent and consistent in their reporting on the use of automated tools in content governance (Dergacheva et al., 2023).

The sanctions vary according to the affordances and requirements of platforms. The variances often lie in the details of measures, as some platforms, like Facebook and Instagram, provide detailed justifications and examples for disinformation. These more detailed rules might also be a tool for managing any adverse effects on the company's reputation as a tool for the spread of harmful information (Opgenhaffen, 2023). No considerable difference between larger and smaller platforms could be observed; some smaller platforms, such as Taboola, also include very miniscule rules. Content moderation seems uniform across many platforms, with removal being a standard measure. Deplatforming is most often used for illegal content, such as terrorism or pornography. However, more recently, deplatforming was expanded to content that breaches platforms' ToS (Monaci, 2024). Studies on deplatforming show that users who were deplatformed for spreading conspiracy theories make alternative accounts on other platforms (Innes & Innes, 2023). Deplatforming, however, has still proven effective in limiting the reach of disinformation (Rauchfleisch & Kaiser, 2024) and it reduces the attention given to influencers (Horta Ribeiro et al., 2025). Furthermore, deplatforming reduces the conversations about the deplatformed persons and reduces the toxicity of their followers on that particular platform (Jhaver et al., 2021). Overall, more diversity regarding content moderation would be beneficial in addressing disinformation more effectively. Appeal mechanisms have not been shown to increase user perception of fairness and transparency of decisions in a study on Facebook (Vaccaro et al., 2020). However, better communicating the moderation and justification rules would be helpful. Some platforms employ a user moderation system, such as Reddit or Wikimedia. Still, Reddit has not established a definition of disinformation. On these platforms, spaces for debate of flagged content or the role of flagging are established. Other platforms have a more 'monarchic structure', where debates about content decisions have to be led outside of the platform (Crawford & Gillespie, 2016, p. 422).

The types of disinformation are clustered according to the dimensions described by the platforms. The role of spam and its intersection with disinformation content is repeatedly shown in the data. Spammers might not be as influential as influencers or

opinion leaders. Still, they operate on a larger scale and create much content that influences discussions (Cantini et al., 2022) or disrupts the user experience (Karunakaran & Brorson, 2019). The stricter rules regarding disinformation are most likely a result of public pressure, for instance, on Meta regarding hate speech policies (Griffin, 2020). Moreover, the COVID-19 pandemic and the media attention on ‘the information disorder’ (Wardle & Derakhshan, 2017, p. 20) resulted in stricter rules regarding health information, vaccines and information about COVID-19, which are included in almost every platform.

The assessment of content or user accounts that should be regulated is often based on the expected consequences, mainly their influence and harm. Assessing harm across platforms is challenging, and the impact of content is often unclear (Altay et al., 2023). Moreover, direct effects are a lot rarer than assumed (Karpf, 2019; Bail et al., 2020) and the focus on foreign actors prevent the realization that domestic actors also undermine trust in institutions, such as scientific institutions and governments (Wardle, 2020). Particularly, the information environment on online platforms is volatile and fast-paced. While regulations do not adapt quickly, platforms can adjust their content moderation systems and contribute to a fact-based and respectful online environment.

Conclusion

This study analyzes the intended platform rules and the platforms’ understanding of disinformation. However, the study has some limitations. First, the enforcement of the regulations was not investigated. Therefore, it is unclear if the content is detected and regulated as the Terms of Service and Community Guidelines describe.

Second, another aspect crucial for disinformation is advertising. Advertisers are essential for platforms to generate revenue, and they are potential spreaders of disinformation. Therefore, advertising guidelines of platforms are also worthwhile to study complementarily to Terms of Services and Community Guidelines.

Lastly, as the DSA is a recent regulation, its impact and the details of its implementation are not yet clear. Therefore, conducting a follow-up study might be beneficial in understanding the implications of the regulation and its shortcomings.

Nevertheless, this study gives crucial insights regarding the content governance process of disinformation and sheds light on differences across platform types and sizes. These insights can be essential to understanding the perspective of platforms on disinformation and the change that the DSA might bring.

The data that support the findings of this study are available from the corresponding author, [author initials], upon reasonable request.

Acknowledgments

I am grateful for the support of my research group, Norm Setting and Decision Processes, at the Weizenbaum Institute and, in particular, for the support of Mariam Sattorov, Jasmin Bernardy, and Caroline Tomalka while conducting this research. I am grateful for the insights and feedback by Prof. Christoph Neuberger, Prof. Herbert Zech, Dr. Anna Litvinienko, Anna-Theresa Mayer, Florian Primig, Dr. Jakob Ohme, Tamer Farag, Stella Köchling, Xixuan Zhang, Franziska Martini, and Vivien Benert.

Author contributions

CRedit: **Rita Gsenger**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Disclosure statement

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

Funding

This work has been funded by the Federal Ministry of Research, Technology and Space (BMFTR) (grant number: 16DIII41 –“Weizenbaum-Institut”).

Note on contributor

Rita Gsenger is a research associate in the Research Group ‘Norm Setting and Decision-Making Processes’. She is a PhD candidate at the Institute of Journalism and Communication Studies at Free University Berlin under Prof. Dr. Christoph Neuberger. Her research focuses on issues of platform regulation, content moderation, evidence-based regulation, and the structural background of disinformation and conspiracy theories, as well as their countermeasures. Rita Gsenger studied cultural and social anthropology and philosophy at the University of Vienna. This was followed by an MA in Philosophy at the University of Innsbruck and an MSc in Cognitive Science at the University of Vienna, both with distinction. In addition, Rita Gsenger was employed as a research assistant at the Institute for Information Systems and New Media at the Vienna University of Economics and Business from 2018-2021.

References

- Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. *Social Media + Society*, 9(1), 1–13. <https://doi.org/10.1177/20563051221150412>
- Bail, C. A., Guay, B., Maloney, E., Combs, A., Hillygus, D. S., Merhout, F., Freelon, D., & Volfovsky, A. (2020). Assessing the Russian internet research agency’s impact on the political attitudes and behaviors of American twitter users in late 2017. *Proceedings of the National Academy of Sciences*, 117(1), 243–250. <https://doi.org/10.1073/pnas.1906420116>
- Balkin, J. M. (2014). Old-school/new-school speech regulation. *Harvard Law Review*, 127, 2296–2342.
- Balkin, J. M. (2018). Free speech is a triangle. *Columbia Law Review*, 7(118), 2011–2056. <https://www.jstor.org/stable/10.230726524953>
- Bazaco, A., Redondo, M., & Sánchez-García, P. (2019). Clickbait as a strategy of viral journalism: Conceptualisation and methods. *Revista Latina de Comunicación Social*, 74(74), 94–115. <https://doi.org/10.4185/RLCS-2019-1323en>
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- Berman, J. M. (2020). *Anti-vaxxers: How to challenge a misinformed movement*. The MIT Press.
- Bradford, A. (2020). *The Brussels effect*. Oxford University Press.
- Cantini, R., Marozzo, F., Talia, D., & Trunfio, P. (2022). Analyzing political polarization on social media by deleting bot spamming. *Big Data and Cognitive Computing*, 6(1), 3–16. <https://doi.org/10.3390/bdcc6010003>
- 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>

- Derave, T., Gailly, F., Sales, T. P., & Poels, G. (2024). A taxonomy and ontology for digital platforms. *Information Systems*, 120(102293), 102293–20. <https://doi.org/10.1016/j.is.2023.102293>
- Dergacheva, D., Kuznetsova, V., Scharlach, R., & Katzenbach, C. (2023). One Day in Content Moderation: Analyzing 24 h of Social Media Platforms' Content Decisions through the DSA Transparency Database. <https://doi.org/10.26092/ELIB/2707>
- Eichensehr, K. E. (2019). Digital switzerlands. *University of Pennsylvania Law Review*, 167, 665–732. https://scholarship.law.upenn.edu/penn_law_review/vol167/iss3/3/
- European Commission. (2018). A multi-dimensional approach to disinformation: report of the independent High level group on fake news and online disinformation. Publications Office. <https://data.europa.eu/doi/10.2759739290>
- European Commission. (2021). Study on “Support to the observatory for the online platform economy”: Annexes. Publications Office. <https://data.europa.eu/doi/10.2759169684>
- European Commission. (2022). 2022 Strengthened Code of Practice on Disinformation. European Commission. <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>
- European Commission. (2023a). *Digital Services Act: Commission designates first set of Very Large Online Platforms and Search Engines*. European Commission. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413
- European Commission. (2023b). *The Code of Conduct on Disinformation*. European Commission. <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>
- Ferrara, E. (2019). The history of digital spam. *Communications of the ACM*, 62(8), 82–91. <https://doi.org/10.1145/3299768>
- Flew, T. (2021). *Regulating platforms*. Polity Press.
- Galantino, S. (2023). How will the EU Digital Services Act affect the regulation of disinformation?. *SCRIPTed*, 20(1), 89–129. <https://doi.org/10.2966/scrip.200123.89>
- Gillespie, T. (2018). Regulation of and by platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media* (pp. 254–279). Sage.
- Goldman, E. (2021). Content moderation remedies. *SSRN Electronic Journal*, 1(2), 1–76. <https://doi.org/10.2139/ssrn.3810580>
- Gorwa, R. (2019b). What is platform governance? *Information, Communication & Society*, 22(6), 854–871. <https://doi.org/10.1080/1369118X.2019.1573914>
- Gorwa, R. (2019). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review*, 8(2), 1–22. <https://doi.org/10.14763/2019.2.1407>
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15. <https://doi.org/10.1177/2053951719897945>
- Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010). @Spam: The underground on 140 characters or less. *Proceedings of the 17th ACM Conference on Computer and Communications Security*, 27–37. <https://doi.org/10.1145/1866307.1866311>
- Griffin, R. (2020). How Public Pressure forced Facebook to change its policies on hate speech. *Science Po*. <https://www.sciencespo.fr/public/chaire-numerique/en/2020/07/09/how-public-pressure-forced-facebook-to-change-its-policies-on-hate-speech/>
- Griffin, R. (2023). Climate Breakdown as a Systemic Risk in the Digital Services Act. Hertie School. https://opus4.kobv.de/opus4-hsog/frontdoor/deliver/index/docId/5075/file/Climate_breakdown_as_systemic_risk_in_DSA.pdf
- Halperin, Y. (2021). When bots and users meet: Automated manipulation and the new culture of online suspicion. *Global Perspectives*, 2(1), 24955. <https://doi.org/10.1525/gp.2021.24955>
- Hameleers, M. (2023). Disinformation as a context-bound phenomenon: Toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory*, 33(1), 1–10. <https://doi.org/10.1093/ct/qtac021>
- Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, 34(1), 1–14. <https://doi.org/10.1080/01972243.2017.1391913>

- Herasimenka, A., Au, Y., George, A., Joynes-Burgess, K., Knuutila, A., Bright, J., & Howard, P. N. (2023). The political economy of digital profiteering: Communication resource mobilization by anti-vaccination actors. *Journal of Communication*, 73(2), 126–137. <https://doi.org/10.1093/joc/jqac043>
- Horta Ribeiro, M., Jhaver, S., Cluet-i-Martinell, J., Reignier-Tayar, M., & West, R. (2025). Deplatforming norm-violating influencers on social media reduces overall online attention toward them. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1–25. <https://doi.org/10.1145/3710960>
- Husovec, M. (2024a). The Digital Services Act's red line: What the commission can and cannot do about disinformation. *Journal of Media Law*, 16(1), 47–56. <https://doi.org/10.1080/17577632.2024.2362483>
- Husovec, M. (2024b). *Principles of the Digital Services Act*. Oxford University Press.
- Innes, H., & Innes, M. (2023). De-platforming disinformation: Conspiracy theories and their control. *Information, Communication & Society*, 26(6), 1262–1280. <https://doi.org/10.1080/1369118X.2021.1994631>
- Iosifidis, P., & Nicoli, N. (2021). *Digital democracy, social media and disinformation*. Routledge.
- Jaster, R., & Lanius, D. (2018). What is fake news? *Versus*, 2(127), 207–227.
- Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the effectiveness of deplatforming as a moderation strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–30. <https://doi.org/10.1145/3479525>
- Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, 23(5), 1301–1326. <https://doi.org/10.1177/1461444820959296>
- Karppf, D. (2019). On digital disinformation and democratic myths. *Social Science Research Council*. <https://mediawell.ssrc.org/articles/on-digital-disinformation-and-democratic-myths/>
- Karunakaran, S., & Brorson, E. (2019). Spam in user generated content platforms: Developing the HaBuT instrument to measure user experience. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 1981–1988. <https://doi.org/10.1109/SMC.2019.8914165>
- Kaur, J., Dara, R. A., Obimbo, C., Song, F., & Menard, K. (2018). A comprehensive keyword analysis of online privacy policies. *Information Security Journal: A Global Perspective*, 27(5-6), 260–275. <https://doi.org/10.1080/19393555.2019.1606368>
- Keller, D. (2022). *Lawful but Awful? Control over Legal Speech by Platforms, Governments, and Internet Users*. The University of Chicago Law Review. <https://lawreview.uchicago.edu/online-archive/lawful-awful-control-over-legal-speech-platforms-governments-and-internet-users>
- Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on twitter: How to coordinate a disinformation campaign. *Political Communication*, 37(2), 256–280. <https://doi.org/10.1080/10584609.2019.1661888>
- Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131(6), 1598–1670.
- Latah, M. (2020). Detection of malicious social bots: A survey and a refined taxonomy. *Expert Systems with Applications*, 151(113383), 1–21. <https://doi.org/10.1016/j.eswa.2020.113383>
- Marsden, C., Meyer, T., & Brown, I. (2020). Platform values and democratic elections: How can the law regulate digital disinformation? *Computer Law & Security Review*, 36, 105373. <https://doi.org/10.1016/j.clsr.2019.105373>
- Martini, F., Samula, P., Keller, T. R., & Klinger, U. (2021). Bot, or not? Comparing three methods for detecting social bots in five political discourses. *Big Data & Society*, 8(2), 1–13. <https://doi.org/10.1177/20539517211033566>
- Mayring, P. (2015). *Qualitative inhaltsanalyse: Grundlagen und techniken*. Beltz.
- McDonald, A. M., & Cranor, L. F. (2009). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, 4(3), 543–568.
- McLoughlin, K. L., Brady, W. J., Goolsbee, A., Kaiser, B., Klonick, K., & Crockett, M. J. (2024). Misinformation exploits outrage to spread online. *Science*, 386(6725), 991–996. <https://doi.org/10.1126/science.adl2829>

- Metaxa-Kakavouli, D., & Torres-Echeverry, N. (2017). Google's role in spreading fake news and misinformation. *SSRN Electronic Journal*, 1–31. <https://doi.org/10.2139/ssrn.3062984>
- Miotto, N., & Droogan, J. (2024). 'Stand against the wiles of the devil': Interpreting QAnon as a pseudo-christian extremist movement. *Critical Sociology*, 51(3), 503–526. <https://doi.org/10.1177/08969205241228744>
- Monaci, S. (2024). The governance of disinformation everyday practices of platform sovereignty. *International Journal of Communication*, 18, 4298–4313. <https://ijoc.org/index.php/ijoc/article/view/21896>
- Moskalenko, S., Pavlović, T., & Burton, B. (2023). QAnon beliefs, political radicalization and support for January 6th insurrection: A gendered perspective. *Terrorism and Political Violence*, 36(7), 962–981. <https://doi.org/10.1080/09546553.2023.2236230>
- Muniz Da Conceição, L. H. (2025). The quantum state of the individual in platform governance: Digital constitutionalism and global democratisation. *Information, Communication & Society*, 1–28. <https://doi.org/10.1080/1369118X.2025.2492572>
- Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society*, 20(11), 4275–4292. <https://doi.org/10.1177/1461444818769694>
- Nooren, P., Van Gorp, N., Van Eijk, N., & Fathaigh, RÓ. (2018). Should We regulate digital platforms? A New framework for evaluating policy options. *Policy & Internet*, 10(3), 264–301. <https://doi.org/10.1002/poi3.177>
- Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1), 128–147. <https://doi.org/10.1080/1369118X.2018.1486870>
- Ó Fathaigh, R., Helberger, N., & Appelman, N. (2021). The perils of legally defining disinformation. *Internet Policy Review*, 10(4), 1–25. <https://doi.org/10.14763/2021.4.1584>
- Opgenhaffen, M. (2023). Combatting disinformation with crisis communication: An analysis of meta's newsroom stories. *Communications*, 48(3), 352–369. <https://doi.org/10.1515/commun-2022-0101>
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293–310. <https://doi.org/10.1177/1461444816661553>
- Poell, T., Nieborg, D., & Van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4), 1–13. <https://doi.org/10.14763/2019.4.1425>
- Rauchfleisch, A., & Kaiser, J. (2024). The impact of deplatforming the far right: An analysis of YouTube and BitChute. *Information, Communication & Society*, 27(7), 1478–1496. <https://doi.org/10.1080/1369118X.2024.2346524>
- Redmiles, E. M., Chachra, N., & Waismeyer, B. (2018). Examining the demand for spam: who clicks? *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–10. <https://doi.org/10.1145/3173574.3173786>
- Rubin, V. L. (2019). Disinformation and misinformation triangle. *Journal of Documentation*, 75(5), 1013–1034. <https://doi.org/10.1108/JD-12-2018-0209>
- Saurwein, F., & Spencer-Smith, C. (2020). Combating disinformation on social media: Multilevel governance and distributed accountability in Europe. *Digital Journalism*, 8(6), 820–841. <https://doi.org/10.1080/21670811.2020.1765401>
- Scharlach, R., Hallinan, B., & Shifman, L. (2023). Governing principles: Articulating values in social media platform policies. *New Media & Society*, 26(11), 6658–6677. <https://doi.org/10.1177/14614448231156580>
- Vaccaro, K., Sandvig, C., & Karahalios, K. (2020). "At the end of the day Facebook does what It wants": how users experience contesting algorithmic content moderation. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1–22. <https://doi.org/10.1145/3415238>
- Van Dijck, J., Poell, T., & Waal, M. d. (2018). *The platform society*. Oxford University Press.
- VERBI Software. (2024). MAXQDA 2022 [Computer software]. <https://www.maxqda.com/>

- Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication*, 37(1), 136–144. <https://doi.org/10.1080/10584609.2020.1716500>
- Wagnsson, C., & Barzanje, C. (2021). A framework for analysing antagonistic narrative strategies: A Russian tale of Swedish decline. *Media, War & Conflict*, 14(2), 239–257. <https://doi.org/10.1177/1750635219884343>
- Wardle, C. (2020). The Media Has Overcorrected on Foreign Influence. *Lawfare*. <https://www.lawfaremedia.org/article/media-has-overcorrected-foreign-influence>
- Wardle, C. (2023). Misunderstanding misinformation. *Issues in Science and Technology*, 29(3), 38–40. <https://doi.org/10.58875/ZAUD1691>
- Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making (Report No. DGI(2017)09). Council of Europe.
- Weiger, C., Smith, K. C., Cohen, J. E., Dredze, M., & Moran, M. B. (2020). How internet contracts impact research: Content analysis of terms of service on consumer product websites. *JMIR Public Health and Surveillance*, 6(4), e23579. <https://doi.org/10.2196/23579>