POLITECNICO DI TORINO
Repository ISTITUZIONALE

The Governance of Disinformation. Everyday Practices of Platform Sovereignty

Original
The Governance of Disinformation. Everyday Practices of Platform Sovereignty / Monaci, S.. - In: INTERNATIONAL JOURNAL OF COMMUNICATION. - ISSN 1932-8036. - 18:(2024).

Availability:
This version is available at: 11583/2993002 since: 2024-10-02T08:03:29Z

Publisher:
USC Annenberg School for Communication and Journalism

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
ACM postprint/Author's Accepted Manuscript, con Copyr. autore

(Article begins on next page)

10 October 2026

International Journal of Communication 18(2024), 4298-4313 1932–8036/20240005

The Governance of Disinformation
Everyday Practices of Platform Sovereignty

SARA MONACI♦
Politecnico di Torino, Italy

Drawing from the theoretical contributions of Shoshana Zuboff (2019) and Benjamin Bratton (2016), this article discusses the emergence of platform sovereignty in the context of *deplatforming* and *deplatformization* practices. Content moderation measures can be implemented by dominant subjects under exceptional circumstances, such as the Ukraine War, or in the aftermath of a *coup d'etat*, such as the Capitol Hill assault in 2021, and they underline the sovereign role of platforms vis à vis that of national states in the context of information and disinformation governance. This article analyzes deplatforming and deplatformization practices to highlight the leading principles of platform sovereignty: the radical indifference to social media communications and the defense of the *data stack*, which enable dominant platforms to maintain their superiority in the social media ecosystem.

Keywords: platform sovereignty, digital sovereignty, disinformation, deplatforming, deplatformization

Disinformation Is Not a Priority

In recent years, disinformation governance has been at the center of public and academic debate (Bowers & Zittrain, 2020; Iosifidis & Nicoli, 2020; Saurwein & Spencer-Smith, 2020). Numerous commentators have argued that major platforms, such as Facebook, Instagram, X, and Google, cannot be considered mere infrastructures, but *moderators and gatekeepers* of content that aim to mitigate the spread of fake news, conspiracy theories, and extremist propaganda (Gillespie, 2018). Some events, such as the Cambridge Analytica scandal (Cadwalladr & Graham-Harrison, 2018), fake news dissemination during the COVID-19 pandemic (Cinelli et al., 2020), and the 2021 Capitol Hill attack (Jeppesen, Hoechsmann, VanDyke, & McKee, 2022), represented regulatory crises that invoked stricter regulations to curb the harmful effects of the power of platforms. More recently, following the outbreak of the war in Ukraine and the pervasiveness of Russian propaganda, major platforms banned content from major news sources—Russia Today (RT) and Sputnik—whose official social media profiles were sanctioned in almost all European Union states (Klar, 2022).

Sara Monaci: sara.monaci@polito.it
Date submitted: 7-15-2023

Copyright © 2024 (Sara Monaci). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at http://ijoc.org.

The regulatory crises highlighted, on the one hand, the marginality of institutions historically charged with media management—nation-states, for example—in the governance of global communication flows, which evolve around platforms than around state media. On the other hand, they underscored two other aspects that will be debated in this article. The first is the emergence of platform sovereignty in the governance of disinformation. Platform sovereignty is understood here not as an a priori quality, but as a dimension that, in line with Bratton (2016), manifests itself widely in the exercise of disinformation governance by platforms such as Facebook, X, etc. The characteristics and sources of legitimate sovereignty are discussed in the subsequent section. In general, the concept describes control over content moderation, and the prerogative to decide who or what platforms to exclude from the media ecosystem through deplatforming and deplatformization actions. In the context of Western liberal democracies, this prerogative has always belonged to nation-states, especially in times of crisis. Today, platforms exert their sovereignty on content moderation, negotiating their prerogatives with those of supranational institutions, such as the European Commission or national governments. The second factor is the inspiring principle that characterizes the exercise of sovereignty, which refers neither to *we the people* nor to *the good of institutions* (Tuccari & Borgognone, 2021), but rather seems to be connoted by what Shoshana Zuboff (2019) calls "radical indifference" toward the "first text"—the online communication that fuels public debate on social media. The "first text" may also involve the potential threats posed by extremist content, conspiracy theories, fake news, and disinformation in various forms. Nevertheless, "disinformation is not a priority" for platforms when it involves the "first text" (Zuboff, 2019, p. 524). It is managed by platforms according to the principle of radical indifference. I will argue that such indifference is, however, functional in feeding the "shadow text"—the accumulation of data and behavioral surplus by the main platforms. This process is also known as *platformization,* which is the accumulation of the behavioral surplus through a stack of data (Van Dijck, 2021). The accumulation of data is the source of platforms' profit, and it can legitimize measures aimed at preventing other platforms from doing the same. This is the case, for example, when mainstream platforms adopt measures of *deplatformization*: They deny essential services to fringe platforms, such as Gab, Parler, and BitChute. For example, in the aftermath of the Capitol Hill attack in 2021, major players such as Meta, Apple, Amazon Web Service, and Microsoft deprived Gab, Parler of basic infrastructure—cloud services, access to third-party funding, presence in-app stores—for collecting data and processing behavioral surplus.

Other content moderation measures exemplify the exercise of platform sovereignty, for example, *deplatforming,* which involves removing specific social media profiles and their associated content. It is usually applied to illegal content, such as pornography, terrorist threats, or copyright infringements. More recently, it has also been applied to disputable claims of pirated content, nudity, or content transgressions that are simply against a platform's terms of service (Mirrlees, 2021). Deplatforming is often the result of reports made by the users themselves or by trusted flaggers; it could also be the outcome of the application of laws, such as those in Europe that punish anti-Semitic expressions or profiles that glorify jihadist terrorism. These are measures adopted by platforms in the exercise of their full sovereignty. It is worth noting that often, such measures—especially those that prevent the accumulation of behavioral surpluses, such as deplatforming—are temporarily adopted by platforms. After a few months, profiles that were once banned may be allowed back into the public arena, given their significant number of followers and their potential to generate traffic and data (Are & Briggs, 2023). Currently, there is no publicly accessible documentation of the measures adopted by these platforms on individual profiles. It is possible to observe

the effects of deplatforming through ex-post surveys, research on specific profiles, and actions taken and documented by other sources (Innes & Innes, 2021; Rogers, 2020). These facts highlight the ambiguity of content moderation measures: Although they censor official Russian social media profiles, they leave room for thousands of individuals to easily replicate the same contents that fuel the same propaganda.

This article does not intend to discuss whether free expression should be censored; rather, it examines the type of sovereignty exercised by platforms and the guiding principles behind it. By exploring a few key concepts and different theoretical approaches, this article contributes to the debate and discusses the legitimacy of platform sovereignty. Sovereignty will be identified as an emerging phenomenon in the governance of disinformation, thanks to the practices of deplatforming and deplatformization. The assumption here is that dominant platforms adopt these measures to balance content moderation policies with domination over the instrumentalizing power of the data stack.

My analysis aims to contribute to this special issue by interrogating communicative approaches to platform study and opening up new possibilities for research within a more articulated interpretive framework that can answer some of the epistemological questions posed by the contemporary digital ecosystem.

Platform Sovereignty and the Source of Legitimacy

Benjamin Bratton's (2016) seminal work is credited with the most systematic and articulate—and at times hostile—discussion about sovereignty in relation to algorithmic platforms. The distinctive feature of platforms is computation, which, according to Bratton, has a political character—computation itself is *governance*. Bratton's (2016) view is decidedly radical and goes beyond the concept of the *state as a machine* in the sense of Weber's (2019) bureaucracy or the *state machine* in Althusser's (2014) philosophy. He emphasizes an identity relationship between *computation and governance* that echoes the visionary positions of 1970's cybernetics. As with Michel Foucault, technologies are not representative of governance, but *are* governance. Unlike Foucault's (2002) archaeology, however, the priority of technologies is not human discourses and bodies, but the computation of all information in the world and the reduction of the world itself to information. Despite the political nature of computational governance, Bratton emphasizes that platform sovereignty is not necessarily the expression of deliberate choices and practices enacted by platforms to determine the exclusive dominance of digital environments. According to Bratton (2016),

> We can identify platform sovereignty as a still immature combination of legally articulated political subjectivity (one sometimes determined by geographic position and sometimes not) and an infrastructural sovereignty produced in relation to the platform infrastructures of planetary-scale computation, regardless of whether these are privately or publicly owned. (p. 44)

Unlike that of states and markets, platform sovereignty can be planned or unplanned, universal or specific, generative or reactive, and technologically determined or politically guaranteed. Its sovereignty can conflict with other sovereignties, as seen in the 2009 Google-Sino War (de Seta, 2021) and, more recently, Facebook's interference in electoral behaviors revealed by the 2018 Cambridge Analytica scandal (Mare, 2018). Bratton's (2016) reflection highlights a new geopolitical dimension of platform sovereignty, which should be

understood as a specific dimension of a larger incidental megastructure—*the stack*—the topological model of contemporary planetary computation. Bratton's positions have sparked a lively debate, considering both the model's ambition and its radical or *accelerationist* position. For instance, Tiziana Terranova (2014) criticized the drifts of a computational totality-to-come underlying a Black Stack to invoke the possibility of constructing a new order for the post-capitalist common: A Red Stack that could function as an infrastructure for autonomist politics and an escape route from the neoliberal paradigm of computation.

Decidedly different is Shoshana Zuboff's (2019) approach to the political role of platforms. In her work, *The Age of Surveillance Capitalism* (Zuboff, 2019), she analyzes the emergence of dominant actors—the major platforms, such as Google, Facebook, etc.—whose political objective is not at all incidental. Notwithstanding the rhetorical practices of technological *inevitabilism* enthusiasts, the dominant platforms' objective is clear and traceable to the monopoly of instrumentalizing power—that is, the power to structure and instrumentalize social behavior to modify, predict, monetize, and control it (Zuboff, 2019, p. 370). This is made possible by the control of the *behavioral surplus*: the ability to process data of billions of users through predictive models able to forecast electoral behaviors, buying habits, vacation intentions, etc. In the vast and, until a few years ago, virgin territory of surplus data, the major platforms have planted their flags (and their algorithms): "[...] *it is as if google and c. had found an unknown continent, colonized it by dint of declarations, and with these they legitimized their own power and sovereignty* [...]" (Zuboff, 2019, p. 241; emphasis in original). What, then, is the source of the legitimacy of this enormous power and the sovereignty derived from it? The legitimacy stems from what Zuboff (2019) calls *the division of learning*: the ability of companies, services, and governments to know and learn from data on billions of individuals, and to transform that knowledge into strategic and predictable patterns of social behavior. This division of learning has two levels: the *first text* represented by the accessible space of social communication and the digital public sphere mediated by platforms and *shadow text*, which is created by processing millions of messages, likes, retweets, and shares—the *data stack*—into predictive models of social behavior. It is the latter text, not the former, that provides platforms with added value and legitimizes their sovereignty over data. Cognitive and interpretative skills, exclusive to a small group of experts, enable certain entities—surveillance platforms—to impose their sovereignty on personal and collective data. It is the behavioral surplus processed by people's everyday interactions that constitutes the shadow text and is fundamental for the platforms' business. There is, thus, an ontological difference between the first text, which involves the sphere of communication on platforms such as Facebook, X, etc., and the shadow text, which is constituted by data manipulation of the behavioral surplus that people leave behind in their everyday use of social media. While the former concerns the public dimension of political debate and is accessible by society as a whole, the latter is understandable only by a small group of experts—*the clergy of surveillance capitalism*—able to create economic value from data. Subjective ideas, dreams, images, reactions, opinion wars, and passionate statements exchanged on social pages are nothing but a *data stack* fed to algorithms that can process complex models of social behavior. Such models can predict preferences for our user profiles, including what they will like, who or what they will avoid, which restaurant they will choose, or which political party they will support in the next election.

The source of the legitimacy of platform sovereignty lies entirely in the answers to these crucial questions: Who knows? Who decides? Who decides who decides? It is the platforms themselves, by virtue of the division of learning and a substantial cognitive asymmetry between the first text and the shadow text,

that maintain sovereignty over the behavioral surplus. What are the impacts of the division of learning on content moderation and disinformation governance?

The next section will describe how radical indifference has to be understood as the ordering principle of content moderation; subsequently, the concepts of deplatforming and deplatformization as practices specific to platform sovereignty in disinformation governance will be discussed.

Content Moderation Through Radical Indifference

As mentioned earlier, much attention has been devoted by academics and others to the increase of toxicity in social media public debates and the need to introduce measures aimed at the governance of disinformation. However, less attention has been paid to analyzing the general effect of the exponential increase in online debate on public opinion, and the less obvious dynamics that generate this scenario. Philosophical reflection—the ability to grasp contemporary phenomena from a systemic perspective—contributes significantly in this regard. Peter Sloterdijk (2018) observed that the devices of *bourgeois public opinion*—the contemporary media—cannot, in any way, serve as collectors of public outrage or enablers of political action. Modern media can trigger affective epidemics because all issues spread according to the principle of viral infection. At the same time, however, they neutralize their significance. According to Sloterdijk (2018), the production of indifference, which erases the distinction between first-order issues and secondary matters, is the democratic mission of contemporary media.

Zuboff's (2019) reflection on the same issue assumes a more radical view: If for Sloterdijk (2018), indifference is the outcome of a confused media sphere where opinions mingle without any ordering principle of a political or moral nature, in Zuboff's (2019) view, it is indifference itself that becomes the ordering principle of content from the point of view of platforms and the exercise of their sovereignty. To explain this view, it is necessary to recall what has been written about the division of learning and the articulation between the first text and the shadow text. Given the strategic priority of shadow text, platforms use the first text as a neutral set of meaningless data. As Zuboff (2019) states:

> The instrumentalizing power cultivates an unusual way of knowledge that combines the formal indifference of the neoliberal perspective with the observational perspective of radical behaviorism. Through the capacity of the Big Other, instrumentalizing power reduces human experience to observable and measurable behavior, while at the same time remaining deliberately indifferent to the meaning of that experience. I call this new form of knowledge radical indifference. (p. 393)

Recalling Skinner's (2011) studies on radical behaviorism, Zuboff (2019) glimpses a perfect implementation of this in the datafication of subjects' behaviors. This process, however, mainly concerns the shadow text—that is, the elaborations of the behavioral surplus. The first text, communication—the locus of symbolic meaning production—has an essentially ancillary role in the shadow text. The meanings produced by communication are the object of the radical indifference of instrumentalizing power because their role is merely functional in feeding the shadow text. Without the communicative interactions in the first text, platforms would not have the raw material to process into the shadow text; for this reason, the

first text is the object of radical indifference and minimal interference by the instrumentalizing power of the platforms. We can, thus, understand why, in content moderation—commonly understood as *moderation of communicative content*—the influence of surveillance capitalism on the division of learning is observed.

As stated at the outset, in recent years, many have urged major platforms to moderate their content. This has led to the implementation of measures and actions aimed at removing formally illegal content, such as those related to child pornography and violence of all kinds, verbal and visual. However, this has left the platforms with a sovereign role—based on radical indifference—, of managing all other types of content that, according to a recent formulation, may represent systemic risks (Tommasi, 2023). Apparent in the moderation of such content is a trade-off between the least possible number of removed posts—because the imperative remains not to limit the accumulation of data stacks that feed the shadow text—and the need to mitigate the spread of toxic or illegal content, as urged by the European Union and citizen organizations pushing for a more inclusive digital debate. Thanks to the principle of radical indifference, it is possible to understand how platforms exert their sovereignty in the governance of disinformation. On the one hand, they are indifferent to communicative facts; on the other hand, they are very careful in applying more binding measures (e.g., deplatformization) toward possible competitors in the digital ecosystem.

The next sections will describe deplatforming and deplatformization as specific areas to observe the dynamics of sovereignty and the implementation of the principle of radical indifference.

Deplatforming

To preserve the integrity of their own environments, most mainstream social media platforms deploy diverse intervention strategies, targeting inappropriate content (e.g., through flagging, demotion, or deletion) and those who share it through temporary or permanent deplatforming. Deplatforming and *flagging content* are considered explicit content moderation measures. Users become aware of such measures once they are put in place, either because they have been previously warned or because they can no longer access their social media profiles. However, it has been noted that other content moderation measures—such as shadow banning or demotion—can be implemented without users' knowledge. Such measures involve limiting the visibility of certain users' content, and given their total opacity, they result in an increasing and inexorable overshadowing of certain profiles considered problematic (Leerssen, 2023; Myers West, 2018). This article will not discuss the invisible moderation measures, but it will focus on deplatforming as a paradigmatic and deliberative action openly initiated by the main social platform. Considering that it is the ultimate sanction formally implemented by Facebook, X, etc., it can be conceived of as a concrete and observable practice of platform sovereignty over users' contents.

As mentioned above, the term deplatforming commonly refers to the removal of an individual user's account by a social network, e.g., Facebook (Rogers, 2020). Deplatforming can be seen as the last resort in content moderation (Johansson et al., 2022)—the final step in a multi-level process. Each platform has a different set of rules, reified in the Terms of Service (ToS), that define what is allowed and what violations incur penalties. Usually, platforms first issue warnings, such as flagging or removing content or temporarily suspending a user's account. What types of violations legitimize deplatforming? Until 2016, it was primarily a strategy to force a content creator to refrain from posting illegal content, such as pornography, terrorist threats,

or copyright violations—material contrary to the rule of law. Later, it was also applied to controversial claims of pirated content, nudity, or content transformations that were simply contrary to the law. In the wake of the United States and United Kingdom elections in 2016, platforms began to take greater responsibility for removing fake news and misinformation, as well as inflammatory, discriminatory, and hateful messages posted by far-right and alt-right figures. In other words, the act of deplatforming has gradually covered a wider range of materials and been applied to the broader media and cultural context of which social media are a part. Since 2021, exceptional circumstances have revealed the sovereign role of platforms—as opposed to that of nation-states and beyond—in banning important news outlets from public debate. The first episode of global significance was the deplatforming of the account of Donald Trump, former President of the United States, following the Capitol Hill attack on January 6, 2021 (Di Salvo, 2021). The assault on the U.S. government headquarters in the aftermath of Trump's electoral defeat in the 2020 election was an exceptional event in recent political history. The episode highlighted how the influence exerted by the former president through his X channel, thanks also to the many lies spread during the COVID-19 pandemic, cultivated a subversive political base capable of committing heinous acts of violence: Five individuals were killed during the assault. The platforms' intervention was immediate: Donald Trump's profile was deleted from X, Facebook, Instagram, and YouTube. Other social platforms, such as TikTok and Twitch, owned by Amazon followed suit by imposing a similar injunction. The censorship of the leader of the largest democracy in the West was a worldwide event that underscored the sovereign role of platforms in undertaking unprecedented measures. From that moment on, platforms were no longer conceived as mere moderators but as *the new gatekeepers of free expression* (Macedo, 2022). Despite the media reach of the action, Trump's deplatforming did not prevent thousands of less prominent profiles from posting content inspired by QAnon's conspiracy theories and their xenophobic and misogynistic positions. Several studies showed that deplatforming actions had a significant effect on platforms such as Facebook and Instagram, but a very limited impact on X and YouTube, which remained highly privileged channels for disinformation (La Gatta, Luceri, Fabbri, & Ferrara, 2023; Rogers, 2020). Thus, the sovereign decision to censor the former President of the United States was not matched by adequate long-term moderation strategies involving thousands of profiles fueling the first text with the same kind of disinformation. The fact that this kind of content became highly viral (Braun & Eklund, 2019; Stöcker, 2020) confirms Zuboff's (2019) contention that in the long run, the principle of content moderation is indifferent to the quality or toxicity of the debate, as long as the shadow text is continuously fed by the interactions of millions of conspiracist and misogynist profiles.

A more recent crisis further underscores the quality of platform sovereignty. The invasion of Ukraine by the Russian military in February 2022 opened a warfront in the heart of Europe and created a new breeding ground for the spread of propaganda and disinformation. From a media governance perspective, the war has led to what has been called a Digital Iron Curtain (Mackinnon & Gramer, 2022), resulting in the Russian government blocking major Western media outlets—including social platforms—and prompting the European Union to impose sanctions on Russian media. Big techs, such as Facebook, X, and YouTube, promptly deplatformed the accounts of major Russian media outlets in the European Union (EU): Russia Today (RT) and Sputnik. Other platforms, such as Reddit, went a step further by announcing that they were restricting Russian state media outlets universally across the platform in all geographies. On a global scale, Facebook and X said they would demote content from Russian state media to make it less accessible. An important study carried out by Sisu et al. (2022) revealed, however, that the actions taken by the EU member states toward RT and Sputnik mainly targeted media companies and Internet providers. Most governments did not propose or introduce domestic legislation aimed at regulating platforms, social media

accounts, or TV channels in response to the war in Ukraine. None of the EU member states had specific procedures in place about the regulation of social media corporations in times of crisis. While there are no social-media-specific coordination procedures or regulations in the vast majority of EU countries, there are more general emergency powers of government that relate to the media in times of crisis. These possibilities for restrictions generally cover public interests, such as national security. None of these laws, however, has been used by national states to regulate social media platforms during the Ukraine conflict. Sisu et al. (2022) highlighted two important points. The first is the marginal role of national governments with respect to social platforms in the governance of information flows, despite the international crisis situation. Their research noted both the legal vacuum about the regulation of social platforms and the passivity of national governments in adopting restrictions already provided by law for the protection of national security understood as public interest.

As a matter of fact, dominant platforms acted in the full sovereignty of their powers by deciding who was in or out. The decision to censor Russian information outlets, however, aligns with the already discussed principle of radical indifference. Indeed, such censorship did not prevent thousands of profiles of ordinary individuals, in the West as elsewhere, from posting daily reruns of the same content as published by RT or Sputnik on fringe channels, such as Rumble, BitChute, or Telegram. Telegram, in particular, was often used in Western countries to spread hyper partisan messages moderated by Facebook or X. In the Italian debate, for example, it was used to spread content relating to Russian propaganda on the Ukraine War (Monaci & Persico, 2023). Rogers (2020) also observed that many users who were deplatformed from the main platforms for their extreme positions (alt-right, extreme right, etc.) migrated to Telegram to circumvent the content moderation policies applied by Facebook, X, etc. Moreover, Telegram has, so far, not been included by the European Commission in the Digital Service Act (DSA) among the number of Very Large Online Platforms (VLOP). Those platforms, according to the DSA, have to observe several content moderation policies aimed at preventing the spread of propaganda and fake news, especially in times of war. Because of the limited number of Telegram's active users in Europe (less than 45 million), the platform remains relatively untouched by the Digital Service Act (DSA) regulations (Choo, 2024). Moreover, various studies on the spread of disinformation during the outbreak of the war in Ukraine noted the massive presence of Russian propaganda on X (Geissler, Bär, Pröllochs, & Feuerriegel, 2023) and the spread of conspiracy theories by pro-Russian groups—for example, the one concerning the presence of U.S. government-funded bio-laboratories in Ukraine (Collins & Collier, 2022). In general, both at the outbreak of the conflict and thereafter, there is evidence of a high level of engagement with Russian state-sponsored media and other domains known to push unreliable information (Chen & Ferrara, 2023).

What, then, is the logic of deplatforming adopted by sovereign platforms? Despite the paradigmatic case of the censorship of major Russian media outlets, the logic of damage control leads to turning a blind eye to the debate in the first text: disinformation produces engagement, virality, and interactions that feed the shadow text, which represents the real value for the instrumentalizing power.

Deplatformization

As Van Dijck, de Winkel, and Schäfer (2021) state, deplatformization is different from deplatforming: It involves a systemic effort to push radical right-wing platforms back to the fringes of the

ecosystem by denying them the infrastructural services needed to survive online. The assault on Capitol Hill prompted similar actions by tech giants. Rising radical right-wing platforms were banned from the digital ecosystem in different ways. Parler was banned from Google's and Apple's app stores, while Amazon Web Services (AWS) denied his cloud services. Parler was left frantically looking for a new infrastructural refuge, as other platforms, including Gab and BitChute, had done before. In this sense, deplatformization is indicative of the hegemonic control wielded by Big Tech platforms, particularly the members of the GAFAM (Google, Amazon, Meta, Apple, and Microsoft), and is reflective of the restructuring of online life around the infrastructure, economic model, and governmental frameworks of platforms. This strategy is implemented by dominant platforms, cutting off cloud services and funding opportunities for marginal platforms through circuits (e.g., PayPal) or economic services (crowdsourcing, crowdfunding). Conceptually, it is possible to define deplatformization as a *digital embargo* aimed at depriving marginal environments of their main means of existence in the online ecosystem. This article does not intend to discuss the legitimacy of deplatformization actions that are often taken in the wake of dramatic events—for example, the blacking out of 8chan following the mass shooting in El Paso, New Mexico (United States) in August 2019 (Hill, 2023) or the removal of cloud services by Amazon Web services vis-à-vis the Parler platform in the aftermath of the Capitol Hill assault (O'Brien, 2021). This article discusses such deplatformization practices as evidence of the sovereignty of dominant platforms.

Deplatformization can be achieved in three ways: blocking access to networked distribution, demonetizing, and disabling infrastructural services. The first strategy—blocking access to networked distribution—impedes platforms' ability to attract users and publish content widely. After the attack on Capitol Hill, both Gab and Parler were banned from the Apple App Store and the Google Play Store. Subsequently, other services followed Apple's example, and on January 9, 2021, Amazon denied Parler access to its cloud service Amazon Web Service (AWS), arguing that Parler's refusal to curb violent content was a real risk to public safety (Romero, 2021). This moderation measure significantly limited the distribution of fringe platforms in smartphone devices and restricted their use to web interfaces. A second deplatformization strategy is demonetization, which can be achieved in various ways, such as by removing a payment service or denying access to fundraising activities. An interesting case exemplifying such deplatformization patterns is BitChute, an alternative video-sharing platform founded in 2017 by Ray Vahey in the United Kingdom. It has grown to a size slightly bigger than Gab, and the two platforms partly overlap in their user bases. For users banned from YouTube, BitChute typically provides an alternative channel. Because of BitChute's permissiveness toward extreme hateful and conspiracy content, X began blocking posts linked to the site, limiting their access to distribution channels. BitChute was banned by PayPal in 2018, causing the demonetization of its users, who financed their videos by linking to fundraising websites, such as SubscribeStar and cryptocurrency processors (Blake, 2018). The third strategy to push fringe platforms away from the mainstream ecosystem has been to disconnect them from infrastructural services, including domain registrars, cloud analytics, and storage services. For its data storage and analytics, Gab has been ousted by all mainstream web-hosting services, including Microsoft Azure in 2018 and AWS in 2019. Since then, Gab has no longer hosted its service in the cloud but has moved to renting hardware in an undisclosed data center, according to the Wall Street Journal (McMillan & Tilley, 2021). In January 2019, BitChute also moved its domain services to Epik after being banned by major infrastructural services.

While such actions taken unilaterally by mainstream platforms refer to their disinformation governance strategies, they also highlight how deplatformization, unlike deplatforming, is aimed at eroding the shadow text of fringe platforms from depths and limiting their substantial ability to survive in the digital ecosystem. The strategies described are aimed not only at removing services but also at preventing *platformization,* limiting the accumulation of a data stack by fringe platforms. Once the fringe environments cannot be accessible to users through their mobile apps, they are also prevented from accessing funding opportunities through crowdsourcing, crowdfunding, or subscription initiatives. More importantly, they are disconnected from the core infrastructure services underlying the stack, including domain registrars, cloud analytics, and storage services. It is not just the infrastructure services but their shadow text that is at risk. In fact, the erosion of users, funding sources, and cloud access that can capitalize on user data is a specific strategy adopted by mainstream platforms to dismantle the shadow text of other platforms competing against them in the profitable disinformation market. Deplatformization, thus, identifies the sovereign role of mainstream platforms in a digital ecosystem in which fringe environments are pushed further to the margins while a few actors maintain their dominance over instrumentalizing power. The publicized principle of keeping the public sphere clean seems, thus, a minor goal compared with limiting the accumulation of behavioral surpluses by potential competitors.

Despite the enforcement by tech companies of various deplatformization strategies, connections between mainstream environments and controversial platforms are not completely severed. Four American researchers found that, despite YouTube's efforts to deplatform extreme-right users who subsequently moved to BitChute, there were still substantial links between the two platforms: Over 25% of URLs found in BitChute's video descriptions point back to YouTube (Childs, Buntain, Trujillo, & Horne, 2022). Another study on the spread of pandemic-related conspiracy theories observed multiple self-branding operations by content creators who articulated their online presence on Gab and BitChute and through active links to YouTube and X (Mahl, Zeng, & Schäfer, 2023). Moreover, recent research related to the social media debate on the Ukraine war (Monaci & Persico, 2023) found, among the dominant profiles on X, numerous links to the fringe Rumble platform openly visible and accessible by users. In many cases, profiles were pushed back to, but not over, the edge; they remained connected to the centralized ecosystem so that mainstream operators could still strategically profit from their information flows. An aggressive oligopoly dominates the ecosystem of online platforms tolerating the presence of fringe environments—ideal breeding grounds for hate speech and conspiracy theories—as merely functional to maintaining a superficial relationship with active multi-platform user traffic.

Conclusions

Discussing sovereignty as an emergent quality of dominant platforms may seem a daunting and obvious task at the same time. Sovereignty could be taken for granted since the size and role of platforms are self-evident in relation to the number of their users, volume of messages, data exchanged, and their pervasive impacts on our daily lives. It is, however, also a daunting task since sovereignty, as analyzed in the essay, emerges not so much in the inclusive power of platforms, but in their measures of exclusion, explicit censorship, or concealment of content and individuals, doomed to slip into the marginal areas of online public debate. Such measures—deplatforming and deplatformization—normally do not make the news, but represent daily, yet opaque, practices of a new kind of sovereignty that erodes that of institutions traditionally dedicated to

moderating the public sphere: media, political state authorities, cultural and/or religious institutions. An unprecedented form of technological sovereignty is emerging, despite the fact that for several years, Google, Facebook, etc. had systematically denied their role as *media* to show themselves as neutral technological infrastructures, substantially indifferent to their impacts on the perceptions, opinions, and behaviors of millions of people. It was not until 2016—with the first code of conduct negotiated by the major platforms with the European Commission—that they undertook content moderation actions toward online expressions of violent extremism, xenophobic or anti-Semitic positions, hate speech, etc. (Moore & Tambini, 2022). Such actions were prompted by the geopolitical condition of those years threatened by multiple Jihadist terrorist attacks, and the rise of a populist, xenophobic extreme right in Europe and the United States. The political context gave rise to two concomitant processes: the adoption of content moderation policies, also encouraged by the European Commission (EC), and the emergence of fringe platforms (4Chan, GAB, 8Kun) as hotbeds for individuals banned from mainstream social networks. Then, a dilemma arose for Facebook, Google, etc. about how to preserve sovereignty over communication and data accumulation and simultaneously comply with the directives of institutions, such as the EC, which advocated for a more incisive role by the online players. Exercising their sovereignty according to a double standard seemed the most effective option. On the one hand, platforms played their role as moderators of content in times of obvious crisis. The Ukraine War or the Capitol Hill assault had a global geopolitical resonance and required specific, firm, and exemplary actions. Nevertheless, such actions can represent a cost in terms of strategic assets. The censorship of Donald Trump, for example, was not only a limitation on the former leader's visibility, but it also represented a cost to Twitter in terms of the platform's popularity. Let us not forget how Twitter itself took advantage of the former U.S. President's reputation to feed its own popularity. Moreover, thanks to such initiatives, platforms catalyzed public opinion through some exemplary measures, concealing the fact that while the opinion leaders of the American far-right were banned from social media, millions of profiles of private individuals continued to spread toxic contents, feeding the platforms' data stack. On the other hand, with the rise of fringe environments, large platforms have adopted more radical measures: They deprived them of network infrastructure and access to funding sources to prevent them from growing and capitalizing on their data stack. Thanks to this form of *digital embargo* toward small competitors, dominant platforms turned out to be the new *super powers* in the social media ecosystem; they act as the United States, the United Kingdom, China, or Russia in the global geopolitical arena with respect to states subject to political or military reprisal, as is the case in Gaza or Taiwan. This is the result of the double standard in the governance of disinformation as performed by the dominant platforms, seizing the opportunity to make exemplary actions of questionable impact while preserving their own sovereignty on the data stack.

References

Althusser, L. (2014). *On the reproduction of capitalism: Ideology and ideological state apparatuses*. London, UK: Verso Books.

Are, C., & Briggs, P. (2023). The emotional and financial impact of de-platforming on creators at the margins. *Social Media + Society, 9*(1), 1–12. doi:10.1177/20563051231155103

Blake, A. (2018, November 14). BitChute, YouTube alternative, cries foul over apparent punt from PayPal. *The Washington Times*. Retrieved from https://www.washingtontimes.com/news/2018/nov/14/bitchute-youtube-alternative-cries-foul-over-appar/

Bowers, J., & Zittrain, J. L. (2020). Answering impossible questions: Content governance in an age of disinformation. *Harvard Kennedy School Misinformation Review, 1*(1), 1–8. doi:10.37016/mr-2020-005

Bratton, B. H. (2016). *The stack: On software and sovereignty*. Boston, MA: MIT Press.

Braun, J. A., & Eklund, J. L. (2019). Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism. *Digital Journalism, 7*(1), 1–21. doi:10.1080/21670811.2018.1556314

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. Retrieved from https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Chen, E., & Ferrara, E. (2023, June 12). Tweets in time of conflict: A public dataset tracking the Twitter discourse on the war between Ukraine and Russia. *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2023)*. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/22208

Childs, M., Buntain, C., Z. Trujillo, M., & Horne, B. D. (2022, June). Characterizing YouTube and Bitchute content and mobilizers during us election fraud discussions on Twitter. *Proceedings of the 14th ACM Web Science Conference, 2022*. Retrieved at https://dl.acm.org/doi/abs/10.1145/3501247.3531571

Choo, F. Y. (2024, May 28). *EU in touch with Telegram as it nears criterion for EU tech rules*. Reuters. Retrieved from https://www.reuters.com/technology/eu-touch-with-telegram-it-nears-criterion-eu-tech-rules-2024-05-28/

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., … Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports, 10*(1), 1–10. doi:10.1038/s41598-020-73510-5

Collins, B., & Collier, K. (2022, March 14). *Russian propaganda on Ukraine's non-existent "biolabs" boosted by U.S. far right*. NBC News. Retrieved from https://www.nbcnews.com/tech/internet/qanon-ukraine-biolabs-russian-propaganda-efforts-boosted-us-far-right-rcna19392

de Seta, G. (2021). Gateways, sieves and domes: On the infrastructural topology of the Chinese Stack. *International Journal of Communication, 15*, 2669–2692.

Di Salvo, P. (2021). "Deplatforming," l'attacco a Capitol Hill e la nuova sfera pubblica privatizzata ["Deplatforming," the Capitol Hill assault and the new privatized public sphere]. *Studi Culturali, 18*(3), 449–458.

Foucault, M. (2002). *Archaeology of knowledge*. London, UK: Routledge.

Geissler, D., Bär, D., Pröllochs, N., & Feuerriegel, S. (2023). Russian propaganda on social media during the 2022 invasion of Ukraine. *The European Physical Journal. EPJ Data Science, 12*(1), 12–35. doi:10.1140/epjds/s13688-023-00414-5

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.

Hill, S. (2023). "Definitely not in the business of wanting to be associated": Examining public relations in a deplatformization controversy. *Convergence. The International Journal of Research into New Media Technologies,* 1–21. Advance online publication. doi: 10.1177/13548565231203981

Innes, H., & Innes, M. (2021). De-platforming disinformation: Conspiracy theories and their control. *Information, Communication & Society, 26*(6), 1262–1280. doi:10.1080/1369118X.2021.1994631

Iosifidis, P., & Nicoli, N. (2020). *Digital democracy, social media and disinformation*. London, UK: Routledge.

Jeppesen, S., Hoechsmann, M., VanDyke, D., & McKee, M. (2022). *The Capitol riots: Digital media, disinformation, and democracy under attack*. London, UK: Routledge.

Johansson, P., Enock, F., Hale, S., Vidgen, B., Bereskin, C., Margetts, H., & Bright, J. (2022). *How can we combat online misinformation? A systematic overview of current interventions and their efficacy*. Retrieved from https://arxiv.org/abs/2212.11864

Klar, R. (2022, March 4). Tech companies seek to choke out Russian state media. *The Hill*. Retrieved from https://thehill.com/policy/technology/596813-tech-companies-seek-to-choke-out-russian-state-media/

La Gatta, V. L., Luceri, L., Fabbri, F., & Ferrara, E. (2023). The interconnected nature of online harm and moderation: Investigating the cross-platform spread of harmful content between YouTube and Twitter. *Proceedings of the 34th ACM conference on hypertext and social media* (HT' 23). Retrieved from https://dl.acm.org/doi/pdf/10.1145/3603163.3609058

Leerssen, P. (2023). An end to shadow banning? Transparency rights in the Digital Services Act between content moderation and curation. *Computer Law & Security Review, 48*. doi:10.1016/j.clsr.2023.105790

Macedo, S. (2022). Lost in the marketplace of ideas: Towards a new constitution for free speech after Trump and Twitter? *Philosophy & Social Criticism, 48*(4), 496–514. doi:10.1177/01914537221089363

Mackinnon, A., & Gramer, R. (2022, August 4). West seeks to pierce Russia's digital iron curtain. *Foreign Policy*. Retrieved from https://foreignpolicy.com/2022/04/08/west-russia-digital-iron-curtain-media/

Mahl, D., Zeng, J., & Schäfer, M. S. (2023). Conceptualizing platformed conspiracism: Analytical framework and empirical case study of BitChute and Gab. *New Media & Society*. Advance online publication. doi:10.1177/14614448231160457

Mare, A. (2018). Politics unusual? Facebook and political campaigning during the 2013 harmonised elections in Zimbabwe. *African Journalism Studies, 39*(1), 90–110. doi:10.1080/23743670.2018.1425150

McMillan, R., & Tilley, A. (2021, January 12). Parler faces complex, costly route to getting back online. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/parler-faces-obstacles-to-getting-back-online-11610474343

Mirrlees, T. (2021). GAFAM and hate content moderation: Deplatforming and deleting the alt-right. In M. Deflem & D. M. D. Silva (Eds.), *Media and law: Between free speech and censorship* (pp. 81–97). Leeds, UK: Emerald Publishing Limited. doi:10.1108/S1521-613620210000026006

Monaci, S., & Persico, S. (2023). La disinformazione in tempo di guerra. Il ruolo delle piattaforme sottotraccia nell'ecosistema dei social network [Disinformation in times of war. The role of fringe platforms in the social networks ecosystem]. *Comunicazione Politica, 24*(2), 271–296.

Moore, M., & Tambini, D. (Eds.). (2022). *Regulating big tech: Policy responses to digital dominance*. Oxford, UK: Oxford University Press.

Myers West, S. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society, 20*(11), 4366–4383. doi:10.1177/1461444818773059

O'Brien, M. (2021, January 12). *Right-wing app Parler booted off internet over ties to siege*. Associated Press News. Retrieved from https://apnews.com/article/google-apple-amazon-parler-app-367b6dff19026c3811651aa01bde58f2

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication, 35*(3), 213–229. doi:10.1177/0267323120922066

Romero, L. (2021, January 12). Experts say echo chambers from apps like Parler and Gab contributed to attack on Capitol. ABC News. Retrieved from https://abcnews.go.com/US/experts-echo-chambers-apps-parler-gab-contributed-attack/story?id=75141014

Saurwein, F., & Spencer-Smith, C. (2020). Combating disinformation on social media: Multilevel governance and distributed accountability in Europe. *Digital Journalism, 8*(6), 820–841. doi:10.1080/21670811.2020.1765401

Sisu, M., Benedek, W., Fischer-Lessiak, G., Kettemann, M., Schippers, B., & Viljanen, J. (2022). *Governing information flows during war: A comparative study of content governance and media policy responses after Russia's attack on Ukraine*. Hamburg, Germany: Verlag Hans-Bredow-Institut. doi:10.21241/ssoar.78580

Skinner, B. F. (2011). *About behaviorism*. New York, NY: Knopf Doubleday Publishing Group.

Sloterdijk, P. (2018). *Zorn und Zeit. Politisch-psychologischer Versuch* [Anger and time. A political and psychological reflection]. Frankfurt am Main, Germany: Suhrkamp Verlag.

Stöcker, C. (2020). How Facebook and Google accidentally created a perfect ecosystem for targeted disinformation. *Proceedings of Disinformation in Open Online Media* (MISDOOM 2019). Retrieved at https://link.springer.com/chapter/10.1007/978-3-030-39627-5_11

Terranova, T. (2014). Red Stack attack! Algorithms, capital and the automation of the common. In R. MacKay & A. Avanessian (Eds.), *#Accelerate#: The accelerationist reader* (pp. 379–399). Boston, MA: MIT Press.

Tommasi, S. (2023). *The risk of discrimination in the digital market: From the Digital Services Act to the future*. Berlin, Germany: Springer Nature.

Tuccari, F., & Borgognone, G. (Eds.). (2021). *La sovranità. Trasformazioni e crisi in età* contemporanea [Sovereignty. Transformations and crisis in the contemporary age]. Roma, Italy: Carocci.

Van Dijck, J. (2021). Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society, 23*(9), 2801–2819. doi:10.1177/1461444820940293

Van Dijck, J., de Winkel, T., & Schäfer, M. T. (2021). Deplatformization and the governance of the platform ecosystem. *New Media & Society, 25*(12), 3438–3454. doi:10.1177/14614448211045662

Weber, M. (2019). *Economy and society. A new translation*. Boston, MA: Harvard University Press.

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