Governing Social Media's Opinion Power: The Interplay of EU Regulations[*]

Urbano Reviglio, Konrad Bleyer-Simon, Sofia Verza

Abstract

Mainstream social media platforms nowadays play an unprecedented gatekeeping role in public discourse. Through rules set by design, terms and conditions, content moderation practices and algorithmic systems, these platforms wield significant influence over individual users and public opinion. While the control over social media's gatekeeping function was initially entrusted to self-regulation under the presumption of neutrality, it is increasingly acknowledged that their role is pivotal for the whole media ecosystem. This article examines the evolving policy landscape in EU media governance, offering a retrospective analysis of how policymakers have addressed social media's "opinion power"— and a prospective analysis of the most recent regulatory developments, most importantly the Digital Services Act and the European Media Freedom Act. Finally, we conclude highlighting limitations, challenges, and opportunities of the EU regulatory framework.

Table of contents

Keywords

social media – platform governance – EU media policy – media regulation – content moderation – public opinion

1. Introduction

Since the early 2000s, the control over the gatekeeping role of social media platforms has been left to self-regulation presuming their content neutrality[1]. As a matter of

---

[1]   Note that in this article the terms "social media" and "platforms" are used interchangeably to

fact, they were considered technology providers, not media companies[2]. Nowadays it is commonly recognized that social media's content moderation and content curation fulfill a fundamental role in the overall media ecosystem, so their role is not to be considered neutral at all. Indeed, through prioritization, removal, labeling, amplification or reduction of content visibility, social media make "editorial" choices which are key in defining which information people can see in their social media feeds[3]. Though these actions may not fit the classic definition of editorial choices and lack comparable regulation, they eventually influence not only individuals and public opinion but also how news is produced, distributed and consumed, as well as the practices of political communication and the formation of users' political preferences.[4]

The increased awareness of the risks that an unregulated social media environment can pose to democracy, consequently induced a shift in internet governance, including a focus on a range of platforms and especially social media.[5] The European Union (EU) led the way in the global regulatory landscape. In 2016, it regulated the protection of privacy and the handling of data with the General Data Protection Regulation (GDPR)[6]. In 2018, it has initiated a self-regulatory endeavor to curb disinformation with the first Code of Practice on Disinformation (CoP)[7] – later extended in the Strengthened Code of Practice on Disinformation (2022), soon to be transformed into a co-regulatory code of conduct. In 2022, the Digital Services Act (DSA)[8] and Digital Market Act (DMA)[9] introduced a set of obligations to safeguard competition, as well as set the ground for the safety and transparency of the online environment, while the European Media Freedom Act (EMFA)[10], enacted in 2024, includes among other things, protections for journalistic content published through social media. In the same year, the European Commission enacted a Regulation on the transparency

---

generally refer to large social media companies. While these often include platforms classified as "Very Large Online Platforms" under the EU Digital Services Act, the usage extends to major social media platforms that might not meet this specific regulatory classification.

[2]   R. Caplan-P. M. Napoli, *Why media companies insist they're not media companies, why they're wrong, and why it matters*, in *Medias Res*, 60, 2018.

[3]   T. Gillespie, *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*, New Haven and London, 2018.

[4]   M. Moore-D. Tambini, *Regulating Big Tech: Policy Responses to Digital Dominance*, New York, 2021.

[5]   Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[6]   Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[7]   EU 2022 *Strengthened Code of Practice on Disinformation*.

[8]   Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).

[9]   Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

[10]   Regulation (EU) 2024/1083 of the European Parliament and of the Council establishing a common framework for media services in the internal market and amending Directive 2010/13/EU (European Media Freedom Act), 20 March 2024.

and targeting of political advertising (TTPA)[11], which is particularly relevant in this context as a significant share of content on social media is sponsored, and even organic content can be amplified through payments. The measures just listed were complemented by a set of guidelines and official communications of the Commission, as well as efforts in neighboring fields, such as the Artificial Intelligence Act (AIA, 2024)[12] that introduces transparency obligations for artificial intelligence services.

This article discusses the policy shift occurring in this area of platform governance by analyzing the emerging European legal framework, and how this regulates the "opinion power" of social media, and the challenges this entails for media pluralism.[13] The research question guiding this analysis is the following: how are the new EU regulations and related policies expected to influence the governance of social media's internal pluralism, particularly users' diversity of exposure? To answer this question, in Chapter 2 we discuss what social media's opinion power is and how we conceive it. This allows us to disassemble the main components of this power, specifying how social media affect internal pluralism and diversity of exposure. Chapter 3 offers an overview of how such opinion power has been regulated in the last two decades in the EU to highlight the challenges faced, and which ones are still relevant. In Chapter 4, we analyze how the emerging European governance regime regulates such opinion power and, in Chapter 5, we discuss whether this is capable of properly redistributing opinion power between different stakeholders. Finally, preliminary conclusions are drawn.

2. Unpacking social media's opinion power

Historically, media exerted an extraordinary influence over the formation of individual and public opinion[14]. Such influence unfolds in various ways; not only by informing but also by entertaining, distracting, and persuading citizens (i.e., infotainment and propaganda), by facilitating or constraining the diffusion of information (e.g., gatekeeping theory), by choosing which political issues are most salient (i.e., agenda setting theory), or by indirectly dissuading (perceived) minoritarian opinions to be expressed (i.e., the spiral of silence theory). To regulate and minimize these forms of

---

[11]   Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising.

[12]   Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

[13]   It is important to note that, while this article examines the legal interconnections opportunities related to the governance of the opinion power exercised by social media platforms to outline the main challenges and opportunities, it does not aspire to provide a comprehensive analysis of this complex and evolving landscape.

[14]   M. E. McCombs-D. L. Shaw, *The Agenda-Setting Function of Mass Media,* in *The Public Opinion Quarterly*, 36 (2), 1972, 176 ss.; E. Noelle-Neumann, *The Spiral of Silence. Public Opinion-Our Social Skin*, Chicago, 1972; N. Postman, *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*, 1985; N. Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media*, 1988.

influence and favor a diverse media environment, the European policy community has traditionally put an emphasis on media pluralism, namely ensuring the availability of a range of different points of view in the media environment[15]. This principle is enshrined in international documents, such as in art. 11(2) of the Charter of Fundamental Rights of the European Union.

In European media policy debates, media pluralism has usually been invoked around issues of media ownership concentration, the role of public service media, and media subsidies[16]. It is a multi-faceted notion that can be understood in various ways. Three dimensions are probably the most critical ones; "external pluralism" which refers to how plural the structure of the media market is; "internal pluralism" which generally refers to the plurality of content and viewpoints that are provided by a single media company; finally, adapting the latter to the digital environment where content abundance has led to the deployment of personalization algorithms, the focus shifted to the diversity of content that users are ultimately exposed to, what is referred to as "exposure diversity"[17]. Nowadays much of the public debate occurs online and much of the news is indeed accessed or found through social media.[18] Much discussion has centered around the negative effects of social media to political discourse, notably the reduction of information diversity to which individuals are exposed to and that they eventually consume (so-called filter bubbles and echo chambers)[19] and the amplification of disinformation, conspiracy theories, and sensational and divisive content.[20] Of course, we still lack conclusive evidence about the impact these phenomena ultimately have, and it is not even clear how exposure diversity could be achieved given the notion of 'diversity' is multidimensional and unsuited for algorithmic operationalizations.[21] This is why in this paper we opted to focus more broadly on social media's "opinion power", how this may affect exposure diversity, and how – from a regulatory perspective – such power could be constrained, or even leveraged to promote more diversity.

---

[15] European Commission: Directorate-General for Communications Networks, Content and Technology, P. Parcu-E. Brogi-S. Verza, et al., *Study on media plurality and diversity online* – Final report, Publications Office of the European Union, 2022.

[16] K. Karppinen, *Problem definitions in European policy debates on media pluralism and online platforms*, in T. Dwyer-D. Wilding, *Media Pluralism and Online News*, Bristol, 2023, 96 ss.

[17] N. Helberger-K. Karppinen-L. D'acunto, *Exposure diversity as a design principle for recommender systems*, in *Information, communication & society*, 21(2), 2018, 191 ss.

[18] N. Newman-R. Fletcher-C.T. Robertson-K. Eddy-R. Kleis-Nielsen, *Reuters Institute Digital News Report 2023*.

[19] The two phenomena are similar but substantially different. Filter bubbles are conceived as cultural and ideological bubbles in which individuals continue to see and consume content that reinforces its opinions and interests. Echo chambers refer to a group situation where established information, ideas, and beliefs are uncritically spread and amplified, while dissenting views and arguments are ignored. The crucial difference is that the former may not depend on the user's autonomy and awareness – therefore it is mainly caused by technological affordances – while the latter pre-exists the digital age and thus it is primarily driven by social relations.

[20] U. Reviglio, *The Algorithmic Public Opinion: a Policy Overview*, in *osf.io*, 5 October 2022.

[21] F. Loecherbach-J. Moeller-D. Trilling-W. van Atteveldt, *The unified framework of media diversity: A systematic literature review*, in *Digital Journalism*, 8(5), 2020, 605 ss.

But what is "opinion power"? It can be generally defined as the ability to influence the processes of individual and public opinion formation. This is not a new term but "a normatively and constitutionally rooted notion that captures the core of media power in democracy and substantiates why that power must be distributed"[22]. Social media platforms, however, have given rise to new forms of influence and dominance over public opinion. How social media's opinion power eventually is exercised, how it differs from traditional gatekeeping power, and what legal provisions and policy interventions are needed to redistribute such power in line with the new and updated normative frameworks is still unclear. These questions also require continuous reexamination in the rapidly changing media-landscape.[23] . It is out of the scope of this article to systematically examine such concept[24] Nonetheless, we briefly describe our current understanding, guiding our subsequent analysis. We outline, on the one hand, the key elements in which such opinion power is exerted and, on the other hand, how this could play out beyond traditional gatekeeping, salience, and exposure.

To effectively understand social media's opinion power and analyze the impact of EU regulations, it is useful to deconstruct the concept of opinion power it as being composed of four main intertwined components: platform design, terms of service, content moderation, and content curation[25]; (i) "Platform design" is the informational architecture — most importantly the user interface design[26] — which prescribes and favors certain behaviors instead of others (also referred to as "affordances"[27]) which may eventually affect news exposure and consumption; (ii) the "terms of service" of social media services represent a form of privatized governance of the rules between the platform and the user, and between users.[28] This also includes the "community guidelines" that platforms usually set and regularly update, and that also define rules that ultimately prohibit certain content and behaviors and shape content moderation practices; (iii) "content moderation" involves all the actions and strategies undertaken by platforms to moderate content according to the platforms' rules and to national laws.[29] This also includes diverse algorithmic systems detecting content to moderate; in

---

[22] *Ibid.*

[23] J. Schlosberg, *Digital Agenda Setting: Re-examining the Role of Platform Monopolies*, in M. Moore-D. Tambini (eds.), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*, New York, 2018, 202 ss.

[24] For such debate see T. Seipp et al., *Dealing with opinion power in the platform world*, cit.

[25] These components are all explicitly defined in the Digital Services Act and they are provided in footnotes.

[26] Art. 3(m) DSA defines "online interface" as «any software, including a website or a part thereof, and applications, including mobile applications». Admittedly, this definition is rather broad and does not highlight the design ability to prescribe values and afford actions.

[27] T. Bucher-A. Helmond, *The affordances of social media platforms*, in *The SAGE handbook of social media*, 2018, 233 ss.

[28] Art. 3(u) DSA defines "terms and conditions" as «all clauses, irrespective of their name or form, which govern the contractual relationship between the provider of intermediary services and the recipients of the service».

[29] T. Gillespie, *Custodians of the Internet,* cit., 4; To be clear, content moderation on social media extends beyond content that is posted by users but includes the moderation of user accounts, comments, direct messages (DMs), live streams, advertisements, hashtags, search results, user interactions, content

particular, (iv) the algorithmic systems that recommend content (i.e., "recommender systems") are especially influential as they embed specific values of content curation, ultimately determining news visibility and exposure.[30] Of course, there are additional mechanisms that can influence social media's opinion power, such as monetization and advertising models or partnerships and publisher deals. However, these may have more indirect effects and are more closely related to issues of competition and copyright. Exploring how the four highlighted components are regulated provides a more practical and straightforward approach to gain insights into the governance of social media's influence on exposure diversity.

It is important to acknowledge how social media's opinion power is not limited to "which" content users see and "how much" exposure they (do or don't) have but, importantly, "how" and "when" content is shown or discovered. Social media have been reported to design their interfaces and order content in specific ways to affect users' behaviors and optimize their engagement.[31] The resulting addictive power and its ability to influence news consumption habits is still an underexplored area. Social media companies regularly conduct experiments on users and test algorithmic changes (i.e., A/B testing).[32] The understanding derived and its potential to enable them to imperceptibly influence patterns of news consumption is evident, though far from being fully understood. Think of the potential manipulative power of social media's recommender systems. By recommending content that users find strongly disagreeable, they can increase political polarization[33]; by recommending a lot of conflicting news accounts, they can generate "reality apathy" (i.e., people do not care about what is

---

flagging, bot activity, amongst others, ensuring a comprehensive content oversight. Art.3(u) DSA defines "content moderation" as «the activities, whether automated or not, undertaken by providers of intermediary services, that are aimed, in particular, at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service, including measures taken that affect the availability, visibility, and accessibility of that illegal content or that information, such as demotion, demonetisation, disabling of access to, or removal thereof, or that affect the ability of the recipients of the service to provide that information, such as the termination or suspension of a recipient's account».

[30]   Art. 3(s) DSA defines "recommender system" as «a fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service or prioritize that information, including as a result of a search initiated by the recipient of the service or otherwise determining the relative order or prominence of information displayed».

[31]   V. R. Bhargava-M. Velasquez, *Ethics of the attention economy: The problem of social media addiction*, in *Business Ethics Quarterly*, 31(3), 2021, 321 ss.

[32]   See R. J. Deibert, *The road to digital unfreedom: Three painful truths about social media*, in *Journal of Democracy*, 30(1), 2019, 25 ss; the persuasive and manipulative capability of social media can also be exemplified with two famous experiments conducted by Facebook more than a decade ago; one is the experiment on 61-million-person in social influence and political mobilization that showed their potential to nudge citizens to vote (R. M. Bons-C. J. Fariss-J. J. Jones-A. D. Kramer-C. Marlow-J. E. Settle-J. H. Fowler, *A 61-million-person experiment in social influence and political mobilization*, in *Nature*, 489(7415), 2012, 295-298); the other is the infamous study on the "emotional contagion effect" that showed how small changes in the algorithms can manipulate emotions on a mass level (A. D. Kramer-J. E. Guillory-J. T. Hancock, *Experimental evidence of massive-scale emotional contagion through social networks*, in *Proceedings of the National Academy of Sciences*, 111(24), 2014, 8788-8790).

[33]   C. A. Bail-L. P. Argyle-T. W. Brown-J. P. Bumpus-H. Chen-M. F. Hunzaker-A. Volfovsky, *Exposure to opposing views on social media can increase political polarization*, in *Proceedings of the National Academy of Sciences*, 115(37), 2018, 9216-9221.

true or not)[34]; conversely, by suggesting conspiracy theories, they nudge users towards making sense of a complex reality through emotionally appealing – albeit simplistic or even untrue – explanations (e.g., the "rabbit hole effect" on YouTube)[35]; by changing the order of political candidates information in search queries, they can manipulate voting intentions, or even favor a specific viewpoint on a topic to people who have not yet formulated a strong opinion (e.g., the "search engine manipulation effect")[36]. Online platforms can influence what information their users consume in other subtle ways as well, notably through the search suggestions or autocomplete functions.[37] For example, TikTok has been recently shown to provide search suggestions through the "Others Searched For" function that may lead to questionable information or send users down contentious political rabbit holes.[38] Furthermore, it should be considered that only a minority of users regularly and proactively look for political news online, and as such, most users encounter political news in social media only or mainly incidentally.[39] This grants these platforms greater control over whether and how much political news a user receives, potentially creating "social media news deserts" where certain user groups are minimally or not at all exposed to political and public affairs content.[40] In this context, it is also important to mention the activities of third-party actors (foreign interference, bot activities, etc.), whose activities are aimed at manipulating other users, as platform design and the choice (or lack) of responses can have a significant impact on users' exposure to content. To effectively regulate the opinion power of social media, it is crucial to address its impact on the exposure and the consumption of diverse content in the short- as well as in the long-term, ultimately influencing not only opinion formation but, more broadly, and more subtly, individual and collective worldviews.

3. A brief history of EU social media governance

Historically, the debate on social media governance pivoted on the question of whether platforms can be held accountable for the content shared through them, legally and ethically. The European liability regime was influenced by the U.S. model of the Com-

---

[34]   L. Thorburn-J. Stray-P. Bengani, *Is Optimizing for Engagement Changing Us?*, 10 October 2022.

[35]   M. Yesilada-S. Lewandowsky, *Systematic review: YouTube recommendations and problematic content*, in *Internet policy review*, 11(1), 2022.

[36]   R. Epstein-J. Li, *Can biased search results change people's opinions about anything at all? A close replication of the Search Engine Manipulation Effect (SEME)*, in *Plos one*, 19 (3), 2024.

[37]   ; Autocomplete functions have been already subject of a case law in 2013 before the German Federal Court of Justice. See S. Wünsch, *Google don't complete. Deutsche Well*, in *dw.com*, 15 May 2013.

[38]   M. Schüler-M. Degeling,-S.Romano-K. Meßmer, *Other search for … the opposition party*, in *tiktok-audit. com,* 16 July 2024.

[39]   R. Fletcher-R. K. Nielsen, *Are people incidentally exposed to news on social media? A comparative analysis. New Media & society*, 20(7), 2018, 2450 ss.

[40]   M. Barnidge-M. A. Xenos, *Social media news deserts: Digital inequalities and incidental news exposure on social media platforms*, in *New Media & Society*, 26(1), 2024, 368 ss.

munications Act, Section 230 (1996).[41] Under the E-Commerce Directive (2000/31/EC)[42], the early 2000s were marked by a relatively hands-off approach to online content. The case law of European Courts has been important in tracing the borders of platforms' liability exemptions, especially with regards to intellectual property rights and thanks[43] . In this situation, much of online content moderation remains at the discretion of platforms also in the EU. The set of rules that users are required to follow on the services are shared with them through terms of service, community guidelines and other internal policies. These rules are usually prepared and updated by the legal teams of the platforms – generally based in the country where the platform's headquarters are established, and reflecting the values and legal environment of that country — which is in most cases the United States.[44] .

Many of today's large online platforms started out as startups without clear use purposes or monetization strategies, not to mention an understanding of the potential role they might play in society. As such, their rules were developed "in an *ad hoc* manner", often by reacting to some imminent threats, among other things, driven by a «desire to prevent fraud, to assuage advertisers, avoid lawsuits»[45]. This approach can be also referred to as the "first wave" of content moderation governance[46]. In the mid to late 2000s and more prominently around the 2010s, detection algorithms and more proactive efforts towards content moderation were undertaken. Moreover, platforms started adapting these rules to the national legal frameworks they operate in, adding to the list of non-acceptable behaviors those considered illegal or harmful in specific country contexts. While some non-tolerable activities that might lead to certain forms of punitive action against the content or its publisher were illegal in most jurisdictions (such as inciting hatred), platforms also had the opportunity to ban certain legally acceptable forms of conduct on their platforms, from pictures containing nudity[47] and

---

[41] Notably, this was supported by the assumption that the size and speed of online user activity made it impossible to monitor online content effectively. Platforms in the U.S. still nowadays enjoy the same status as Internet Service Providers (ISPs), considered as mere hosting providers of content created by someone else. In addition to Section 230, Section 512 of the Digital Millennium Copyright Act (DMCA) is also relevant, providing the "Safe Harbor Protection" to online platforms, shielding them from liability for copyright-infringing content uploaded by users, provided they comply with notice-and-takedown procedures.

[42] In particular, art. 14 under the E-Commerce Directive (2000/31/EC) requires action from platforms when they become aware of the existence of illegal content on their services. Platforms must act quickly to remove illegal content to maintain their liability protection.

[43] See for example EctHR, *Delfi AS v. Estonia,* app. no. 64669/09 (2015); CJEU, C-324/09, *L'Oréal and Others* (2011) in contrast to *Tiffany (NJ) Inc. v. eBay Inc.* 600 F.3d 93 (2nd Cir.2010); CJEU, C-131/12, *Google Spain* (2014); T-201/04, *Microsoft v. Commission*, mentioned in M. Cantero Gamito, *Regulation of Online Platforms* in J. Smits-J. Husa-C.Valcke-M. Narciso, *Elgar Encyclopedia of Comparative Law*, Cheltenham-Northampton, 2021.

[44] R. Gorwa, *The Politics of Platform Regulation: How Governments Shape Online Content Moderation*, Oxford, 2024, 21.

[45] *Ibid.*, 13.

[46] This was basically characterized by privatized vertical procedures that apply legislative-style rules drafted by platforms to individual cases and hears appeals from those decisions, deciding on a case-by-case basis. This approach still dominates content moderation practices nowadays. See E. Douek, *Content moderation as systems thinking*, in *Harvard Law Review*, 136 (2), 2022, 526 ss.

[47] Inevitably, sometimes users oppose such content moderation decisions. Consider the #Freethenipples

satire, for example, to the publication of fabricated and misleading information, that got to be known later as disinformation.

As social media platforms grew massively, the self-regulation approach has been increasingly questioned. The international policy community, media activists and scholars began to be concerned of the rising gatekeeping power of mainstream social media platforms, both for the moderating practices themselves and for their opacity, indicating a growing consensus for a regulatory change [48]. In this period an historical shift took place where traditional forms of speech regulation, which typically involve legal restrictions and eventually government censorship (what has been described as "old-school speech regulation") contrast with newer forms that arise from the nature of social media platforms, algorithmic content moderation, and the actual privatization of public discourse (i.e., "new-school speech regulation") [49].

European policymakers, civil society and scholars soon had to realize how complicated such exercises were, as there is indeed content on social media that does not violate existing laws but still could cause significant harm to society. In particular, policy makers were concerned about the proliferation of untrue statements that might risk the integrity of elections or hamper public health authorities' efforts to handle pandemics. Indeed, content moderation has always been particularly challenging as it requires striking a balance between free speech and other conflicting interests, such as reputation, public safety, or crime prevention, taking into account [50] At the same time, there has been a lack of specific forms of regulation that could be used as a role model for content on social media, as traditional media laws and policies are often inadequate in light of and the emergence of new actors involved in it. The calls for increased action created an environment in which platforms proactively formulated their own rules on what is allowed and what is unacceptable on their services – [51].

In 2015, the European Commission published its Digital Single Market Strategy[52], preluding the enactments of the DSA, DMA, AIA and EMFA. A pivotal shift towards a more active governance of social media was marked by the voluntary EU Code of conduct on countering illegal hate speech online (2016)[53]. Major platforms like Facebook, Twitter, and YouTube agreed to review and remove illegal hate speech within 24 hours. This initiative reflected the EU's growing concern for the societal impacts of

---

movement which gained traction particularly around mid 2010s and advocated for the normalization of female breasts challenging societal restrictions on their visibility. It generally argued that the censorship and sexualization of female breasts are discriminatory and perpetuate harmful stereotypes.

[48] E. Morozov, *The net delusion: The dark side of Internet freedom,* New York City, 2012; C. Fuchs, *Social Media: a Critical Introduction,* 2013; T. Gillespie, *Custodians of the Internet,* cit.; J. Meese- S. Bannerman, *The Algorithmic Distribution of News,* in *Policy Responses,* 2022.

[49] J. M. Balkin, *Old-school/new-school speech regulation*, in *Harvard Law Review,* 127, 2013, 2296.

[50] I. Nenadić-S. Verza, *European Policymaking on Disinformation and the Standards of the European Court of Human Rights*, in E. Psychogiopoulou-S. de la Sierra, *Digital Media Governance and Supranational Courts Selected Issues and Insights from the European Judiciary*, Cheltenham, 2022, 175 ss.

[51] T. Flew, *Regulating Platforms*, Cambridge, 2021.

[52] Communication COM(2015) 192 final from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 6 May 2015 on A Digital Single Market Strategy for Europe.

[53] See *EU Code of Conduct on Tackling Illegal Hate Speech*, 2016.

online hate, setting the stage for more stringent measures.[54] At times, national measures also played a role in this process: in 2017 the German Network Enforcement Act (NetzDG) introduced strict requirements for social media platforms to remove "obviously illegal" content within 24 hours under threat of fines of up to €50 million. This law was a critical moment for EU content moderation policy, emphasizing the responsibility of platforms in policing content, and influenced discussions across Europe about balancing free speech with the need to control harmful online behavior.[55] In 2018, the revelation that the company *Cambridge Analytica* had harvested the personal data of millions of Facebook users without consent and used it for political profiling and targeting – among others, on behalf of pro-Brexit and pro-Trump campaigners – catalyzed a broader reckoning regarding the role of digital platforms in societal harm, mis- and disinformation, and electoral interference. Following this, the EU's Code of Practice on Disinformation (CoP) was introduced in 2018, addressing the spread of manipulative content in the context of new techniques and tactics, reflecting concerns exacerbated by the scandal's revelations. This policy instrument, however, was still non-binding and voluntary. In 2018, the General Data Protection Regulation (GDPR) came into effect, representing a fundamental step in strengthening human rights in the digital realm as well as illustrating the growing EU digital regulatory worldwide influence (i.e., Brussels effects)[56].

In this same period, the debate on content moderation practices became particularly intense in relation to "high-intensity events" (elections, terrorist attacks, natural disasters, pandemics), as certain kinds of content communicated in these contexts, even if otherwise legal, may pose a risk to human life and interfere with human rights, including the right to receive and impart information and to form and develop an opinion[57]. Key examples are concerns related to COVID-19 disinformation[58] and geopolitical pressures for information sovereignty. The latter have been on the European policy agenda at least since 2015, when the European Council asked the High Representative of the EU for Foreign Affairs and Security Policy to address information manipulation attempts originating from Russia. This was followed by a number of measures

---

[54] Other relevant regulations affecting content moderation include: the Directive on combating terrorism 2017/541, which provides for similar obligations against public online incitement to acts of terrorism; The revised Audio-visual Media Service Directive (AVMSD) 2018/1808, which includes new obligations for video-sharing platforms to tackle illegal online content (such as terrorist content, child sexual abuse material, racism and xenophobia) and specific categories of hate speech; The Directive on Copyright in the Digital Single Market 2019/790 which establishes obligations for copyrighted-materials.

[55] R. Gorwa. *Elections, institutions, and the regulatory politics of platform governance: The case of the German NetzDG*, in *Telecommunications Policy*, 45(6), 2021, 102 ss.; V. Claussen, *Fighting hate speech and fake news. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation*, in *Rivista di diritto dei media*, 3, 2018, 110 ss.

[56] A. Bradford, *The Brussels effect: How the European Union rules the world,* New York, 2020.

[57] I. Nenadić-Verza, *European Policymaking on Disinformation,* cit.

[58] For example, the Communication on Tackling COVID-19 disinformation – Getting the facts right in 2020 as well as the European Democracy Action Plan, in the same year). See Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, *Tackling Covid-19 Disinformation – Getting The Facts Right*, JOIN/2020/8 Final, of 10 June 2020.

related to so-called hybrid threats, such as the establishment of the East Strategic Communication Task Force in the same year.[59]

In the early 2020s, we entered a new era for content moderation governance, which reflects a growing understanding of the complexity of human communication online and the limits of previous moderation technologies. Facebook, for example, established in 2020 the "Oversight Board" as an independent body that can issue non-binding recommendations to Facebook and Instagram, or even make binding decisions on whether specific content should be allowed or not on the platforms[60]. The Digital Services Act (DSA), in particular, enshrines this policy paradigm shift. Enacted in 2022, it aspires to promote a European digital sovereignty by establishing systemic regulation of platforms like those operated by Meta or Alphabet to address their impact on public discourse regulating various governance areas previously almost untouched, including terms of service, content moderation, recommender systems, and interface design. In 2024, the European Union also enacted the European Media Freedom Act (EMFA), a legislation designed to safeguard media freedom and pluralism, partially on digital platforms too. The same year marked a significant milestone with the introduction of the Artificial Intelligence Act (AIA), the first major regulation globally to specifically govern the use of artificial intelligence, that aims to address risks specifically posed by AI applications. Across all these regulations, the governance of content moderation has evolved from a focus on removing explicitly illegal content to strengthening the users' autonomy, among other things through algorithmic transparency and data protection, and to mitigate the risk posed by harmful content, such as hate speech and disinformation, by considering them as symptoms of so-called systemic risks, that stem from the design or functioning of platforms.[61]

Despite this policy shift, the EU's response to the challenges posed by large social media platforms was rather slow. This can be traced back to the history of Internet devel-

---

[59] G. Abbamonte -P. Gori, *European Union*. In: O. Pollicino (ed.), *Freedom of Speech and the Regulation of Fake News,* Antwerp*,* 2023, 129 ss.

[60] Criticism stems from the fact that the OB depends on Meta's funding (it is funded by an independent trust established by Meta); also, it is not that diverse in terms of geographical and gender background; and despite being an innovative body for a company like Meta, its case-by-case ex post review of content moderation practices – setting precedents for future similar cases- exemplifies the standard picture approach of judicial review–style solutions, thus falling in the traditional approach of speech regulation. On the other hand, subjecting platforms' content moderation decisions to judicial review may be expensive and time-consuming. A relationship between the OB's rulings and the domestic legal frameworks consists in the obligation for Meta to implement the OB's ruling unless doing so "could violate the law" in the relevant jurisdiction; moreover, ideally the OB cannot review cases that clearly violate national laws and could make Meta and its employees legally vulnerable. See D. Wong- L. Floridi, *Meta's oversight board: A review and critical assessment*, in *Minds and Machines*, 33(2), 2023, 261 ss.

[61] The DSA, for example, outlines four main categories of "systemic risks". A first category concerns the risks associated with the dissemination of illegal content, such as the dissemination of child sexual abuse material or illegal hate speech. A second category concerns the actual or foreseeable impact of the service on the exercise of fundamental rights, including but not limited to freedom of expression and of information, including media freedom and pluralism. A third category of risks concerns the actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes, as well as public security. A fourth category of risks stems from similar concerns relating to the design, functioning or use, including through manipulation, of VLOPs and VLOSEs with an actual or foreseeable negative effect on the protection of public health, minors and serious negative consequences to a person's physical and mental well-being, or on gender-based violence.

opment itself which has been driven by the US and its values.[62] In the early days, social media were widely seen as a powerful force for good, promoting free expression, fostering connections among people, and driving a global democratic revolution. Once the US-based digital platforms came to dominate the market showing their negative consequences and potential risks, there was a lack of collective determination and financial resources, in addition to the complexities in the coordination of the EU's digital policies and the digital companies' lobbying pressure. Over time, a recurring cycle has been observed where public outrage over specific incidents temporarily disrupted the status quo, prompting platforms to make only superficial adjustments that appease public sentiment without leading to substantial regulatory changes[63]. Public pressure and sustained negative coverage, though, seem to have had a role in shaping the governance of online platforms[64]. Ultimately, the history of the regulation of social media's opinion power is not only the history of unprecedented socio-technical challenges, but also a history of U.S. technological and cultural hegemony over the EU, and how big tech companies' increasing influence over people's life have been slowly recognized, increasingly opposed and, finally, regulated.[65]

4. The interplay between EU regulations

At the time of writing, the DSA is the main EU regulation that can be deployed to govern social media's opinion power and its influence on media pluralism. It is complemented by other regulatory tools, primarily the European Media Freedom Act (EMFA, 2024), the Regulation on the Targeting and Transparency of Political Advertising (2024), the Artificial Intelligence Act (AIA) (2024), the Strengthened Code of Practice on Disinformation (CoP, 2022), the General Data Protection Regulation (GDPR, 2016), amongst others.

The DSA thus updates the European legal framework on platform liability, which was previously prescribed by Directive 2000/31 on E-commerce. It develops a more thorough regulation on the techniques and the decision-making processes employed for content moderation. It contains many reporting and transparency obligations[66] By establishing Very Large Online Platforms (so-called VLOPs) and Very Large Search

---

[62] H. Nieminen-C. Padovani-H. Sousa, *Why has the EU been late in regulating social media platforms?*, in *Javnost-The Public*, 30(2), 2023, 174 ss.

[63] M. Ananny-T. Gillespie, *Public platforms: Beyond the cycle of shocks and exceptions*, IPP2016 The Platform Society, 2016.

[64] N. Marchal-E. Hoes-K. J. Klüser-F. Hamborg-M. Alizadeh-M. Kubli-C. Katzenbach, *How Negative Media Coverage Impacts Platform Governance: Evidence from Facebook, Twitter, and YouTube,* in *Political Communication,* 2024, 1 ss.

[65] A similar dynamic has been empirically observed in the United Kingdom. See M. Kretschmer-U. Furgal-P. Schlesinger, *The emergence of platform regulation in the UK: an empirical-legal study,* in *Weizenbaum Journal of the Digital Society–special issue:"Democracy in Flux–Order, Dynamics and Voices in Digital Public Spheres"*, 2022.

[66] See, in particular, arts. 15 (Transparency reporting obligations for intermediary services), 24 (Transparency reporting obligations for online platforms) and 42 (Transparency reporting obligations) and the EU official website with the DSA transparency reports: *transparency.dsa.ec.europa.eu.*

Engines (VLOSEs)[67] as new legal subjects, the DSA acknowledged the special role and reach of social media, which therefore shall «pay due regard to freedom of expression and of information, including media freedom and pluralism » (Recital 47). Platforms that have at least 45 million monthly active users (10% of EU population) (art. 33) have additional duties regarding activities that directly and indirectly influence media pluralism. VLOPs and VLOSEs are required to conduct periodical assessments for "systemic risks" potentially caused by their services (art. 34) and to take appropriate measures to mitigate them (art. 35). Content that is harmful, but not necessarily illegal, as well as a number of systemic risks that represent a challenge to freedom of expression, media pluralism, informed citizenship and electoral integrity fall under the "duty of care" of VLOPs (Recitals 80-83). The DSA also lists factors to consider when assessing risks (such as content moderation systems, recommender systems, advertising systems, and "data related practices") and a list of possible mitigation measures (such as adapting interface design, terms and conditions, content moderation processes, etc. under art. 35 DSA).

Furthermore, the DSA evolves the traditional "Notice-and-Takedown" system establishing a more detailed "Notice-and-Action" framework (art. 16 DSA). As well as under the e- Commerce Directive (ECD, 2000)[68], any individual can notify companies about specific supposedly illegal content, and companies are in charge of deciding whether to remove the content or not. Furthermore, the DSA evolves the traditional "Notice-and-Takedown" system contained under the e- Commerce Directive (ECD, 2000), in which any individual could notify companies about specific supposedly illegal content, and companies were in charge of deciding whether to remove the content or not. This approach led to low quality notifications, increasingly sent by algorithms, and not humans, and to the over-removal of content to avoid liability and reduce costs. The DSA thus establishes a more detailed "Notice-and-Action" framework (art. 16 DSA)[69].

Positive developments can be observed in the DSA in this regard: firstly, a new subject is introduced: "trusted flaggers", which are entities recognized by online platforms for their expertise and reliability in identifying content that violates community standards or legal requirements (art. 22). Moreover, platforms shall also include in their general terms and conditions information on the restrictions they impose on the use of their services, applying them in a transparent and non-discriminatory manner and with due regard for fundamental rights such as the right to freedom of expression and the freedom and pluralism of the media (art. 14). Such information shall concern, inter alia, the policies, procedures, measures and tools used for the purposes of content moderation, including algorithmic decision-making and human verification, as well as the procedural rules of their internal complaint-handling system. Article 17 also

---

[67]   Find here the *list of designated very large online platforms and search engines under DSA*.

[68]   Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce).

[69]   A. de Streel-M. Husovec, *The e-commerce Directive as the cornerstone of the Internal Market, Policy* Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies, European Parliament, 2020.

requires social media to provide "statements of reasons", meaning a clear and specific justification for a series of restrictions imposed on the grounds that the information provided by the recipient of the service constitutes illegal content or is incompatible with their general conditions. These restrictions include removal of content, disabling access to content, or demoting content[70], suspension, termination or other limitations of service provision, monetization opportunities or even the whole account. Finally, a specific internal complaint-handling system and out-of-court dispute settlement bodies related with the notice-and-action mechanism have been introduced, updating art. 17 ECD. The dispute settlement bodies (art. 20 and 21 DSA), are established by the EU member states and the national Digital Service Coordinators, do not establish precedents, handle higher numbers of complaints and not just selected cases, and have no binding powers on platforms[71].

The DSA is complemented by the European Media Freedom Act (EMFA), especially for the protection of news media pluralism. The EMFA aims to protect and improve the media sector in the EU, given the crucial democratic implication of guaranteeing access to free, plural and independent news[72]. For the scope of this article, it is especially worth considering the inability of the media to fulfill their social role of providing quality and independent news if they are overwhelmed by the competition of digital actors that impact their revenues and news distribution[73]. Especially relevant in this context is art. 18 which regulates the relationship between VLOPs and news media services ("media service providers" (MSP), art. 2 EMFA), providing a so-called "special treatment for the media" or "media privilege"[74]. In particular, it regulates the suspension of the «provision of its online intermediation services» based on its terms and conditions, and outside of the systemic risk cases listed in the DSA. The goal is to

---

[70] Demotions are reductions of content visibility. Another common term is "shadowban" which specifically refers to demotions that have not been announced by a platform and, thus, they are only suspected. To better understand the topic, see T. Gillespie, *Do Not Recommend? Reduction as a Form of Content Moderation*, in *Social Media+ Society*, 8(3), 2022, and P. Leerssen, *An End to Shadow Banning? Transparency rights in the Digital Services Act between content moderation and curation*, in *Computer Law & Security Review*, 48, 2023, 105790.

[71] This process could also lead to issues related to workload, de facto limiting the efficiency and the effectiveness of ODS. The list of certified out-of-court dispute settlement bodies is available online.

[72] M. Monti, *The missing piece in the DSA puzzle? Article 18 of the EMFA and the media privilege*, in *Rivista Italiana di Informatica e Diritto*, 2, 2024.

[73] The proposal acts on several aspects: the concentration and transparency of media ownership, the governance of media policies by European and national regulatory authorities, journalists' safety from surveillance through spywares, editorial independence, and the relationship between media service providers and platforms in the context of content moderation. See E. Brogi-D. Borges-R. Carlini-I. Nenadic-K. Bleyer-Simon-J. Kermer-U. Reviglio-M. Trevisan-S. Verza, *The European Media Freedom Act: media freedom, freedom of expression and pluralism*, 2023.

[74] A similar clause had been discussed as an obligatory "media exemption" or "non-interference principle" in the phase of drafting the DSA, encompassing terms and conditions and notice-and-action procedures. But at that time, no political agreement was reached on this issue; however, a debate started which fed into the debate over art. 18 EMFA. See C. Papaevangelou, *'The non-interference principle': Debating online platforms' treatment of editorial content in the European Union's Digital Services Act*, in *European Journal of Communication*, 38(5), 2023, 466 ss. On the media privilege in EMFA, see D. Tambini, The EU is taking practical measures to protect media freedom. Now we need theory, in *cmpf.eui.eu*, 9 May 2023; M. Z. van Drunen-C. Papaevangelou-D. Buijs- R. Ó. Fathaigh, *What can a media privilege look like? Unpacking three versions in the EMFA*, in *Journal of Media Law*, 15 (2), 2023, 152 ss.

offer protection against the unjustified removal by VLOPs, in case the media content was produced in line with professional standards[75]. To minimize the impact of any restriction to that content on users' right to receive and impart information, and to preserve media outlets and journalists from unjustified content removals or suspensions, VLOPs should submit their statement of reasons to the MSPs prior to the suspension or restriction of visibility taking effect, and MSP's complaints to platforms should be handled with priority «to minimize the impact of any restriction to that content on users' right to receive and impart information» (Recital 50). Thus, art. 18 EMFA aims to contribute to protect media pluralism by establishing a privileged procedure for content moderation of news content produced by MSPs over other types of content, ultimately leading to a prominent and diverse provision of news content offered by independent sources, that are already subject to other regulations, either hard media law or self-regulation aimed at guaranteeing the quality of their content.

The DSA also introduces specific regulations for recommender systems. It provides the first detailed legal definition of recommender systems worldwide[76] (art. 3 (s)) and requires platforms to (1) notify its users when these systems are being used, (2) disclose the employed algorithmic parameters in plain and intelligible language, (3) allow users to manually alter the criteria used for content recommendations (art. 27), and (4) allow users to opt-out from recommender systems based on profiling (art. 38). Furthermore, under the DSA independent audits of algorithmic systems are mandated. According to art. 37 platforms must undergo, at their own expense and at least once a year, independent audits aimed at assessing compliance with the regulation. These will complement and interact with the assessment and mitigation of systemic risks, which includes risks to media pluralism, in accordance with the delegated regulation on the performance of audits (DRPA) (2024/436) (arts. 13 and 14).[77] To implement and monitor the DSA provisions, the European Centre for Algorithmic Transparency (ECAT) has also been established, which is committed to research algorithmic systems for policy purposes, through platform assessment and investigations, scientific projects, networking, and community building. This legal framework complements the GDPR which guarantees European consumers a set of individual rights in relation to the collection of data useful for user profiling, thus relevant in relation to recommender systems on social media platforms. Importantly, the GDPR regulates more in detail profiling and, even if it is not explicitly mentioned, the "right to an explanation"[78].

Social media's recommender systems are also regulated under the Artificial Intelligence Act (AIA). At present, however, they are classified only as "minimal risk" AI systems,

---

[75] Media service providers should self-declare to VLOPs ex art. 18 (1), stating among other things, to abide by standards of editorial independence and not to provide AI generated content without editorial review.

[76] Except, to our knowledge, a more general definition in China's regulation. U. Reviglio-G. Santoni, *Governing Platform Recommender Systems in Europe: Insights from China*, in *Global Jurist*, 23(2), 151 ss.

[77] Commission Delegated Regulation (EU) 2024/436 of 20 October 2023 supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council, by laying down rules on the performance of audits for very large online platforms and very large online search engines.

[78] L. Edwards-M.Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, in *Duke Law & Technology Review*, 16 (18), 2017, 18 ss.

which are systems free to use for which the only suggested regulatory action is the promotion of voluntary codes of conduct. AIA's framework, in fact, defines four levels of risk in AI: unacceptable, high, limited and minimal or no risk, carrying with them different obligations. Initially, recommender systems were deemed as high risk by the EU parliament[79]. This would have led to a whole series of additional obligations for platforms that would have increased the safety, transparency and accountability of these systems, most importantly risk management (art. 9) and technical requirements (art. 15). However, the EC is empowered to adopt delegated acts to amend the list of AI systems categorized under each risk typology (art. 73). The possibility therefore remains that such systems in the future may be subject to more stringent provisions[80]. There are, however, already applicable norms of the AIA. First, according to art. 5, AI models using "subliminal techniques" beyond a person's consciousness or that are intentionally manipulative or designed to exploit a person's vulnerability in a manner that causes or likely to cause physical or psychological harm are to be banned. This would prevent recommender systems from manipulating users. In parallel, the DSA addresses the issue of "dark patterns" (art. 25), which states that social media's interface should not be designed in a way that hinders users' ability to make informed decisions[81]. Additionally, the emerging framework established by the DSA for data access for research (art. 40), which also operates in tandem with the GDPR rules on research processing, is another fundamental tool to improve the understanding of social media functioning, and how this affects media pluralism, eventually providing evidence for policymaking.

Furthermore, a particularly sensitive area of content moderation is undoubtedly disinformation. In this governance area, the central instrument of the EU is the Code of Practice on Online Disinformation (CoP), a self-regulatory instrument set up in 2018 by the leading online platforms and the advertising industry established and evaluated through a process guided by the European Commission (EC). While acknowledged as a significant first step, the CoP in its initial execution contained some critical flaws and limitations.[82] Because of the noted shortcomings, the EC modified this self-regulatory instrument in a "strengthened" version (2022) and is expected to convert it into a code of conduct under the DSA. By signing the CoP platforms such as Facebook and TikTok have also committed to making changes to their algorithms based on so-called "trust indicators" that would reduce the risk of users being misled by ambiguous content. Signatory platforms should amplify "authoritative information", allowing users to introduce trust signals into recommendation systems, and providing metrics to evalu-

---

[79] See *Commission welcomes political agreement on Artificial Intelligence Act,* press release, 9 December 2023.

[80] N. Helberger, *FutureNewsCorp, or how the AI Act changed the future of news*, in *Computer Law & Security Review*, 52, 2024, 105915.

[81] Recital 67 DSA refers to dark patterns and lists a number of unfair practices that platforms often engage in.

[82] E. Culloty-K. Park-T. Feenane-C. Papaevangelou-A. Conroy-J. Suiter, *Covidcheck: assessing the implementation of EU code of practice on disinformation in relation to Covid-19*, Project Report. Broadcasting Authority of Ireland and FuJo, 2021; I. Nenadic- E. Brogi- K. Bleyer-Simon, *Structural indicators to assess effectiveness of the EU's Code of Practice on Disinformation*, European University Institute, Working Paper, 2023.

ate the effectiveness of fact-checking, and information on user engagement with the options provided to modify the output of algorithms. Trust indicators are expected to provide the basis for platforms to improve the discoverability of trustworthy content sources and decrease the visibility (demotion) of their untrustworthy counterparts[83]. Political advertising has already become a key component of platform regulation in the context of the initial Code of Practice on Disinformation, as well as its later iteration. As content aimed at one's friends and followers can easily be turned into advertising, by boosting or amplifying a post through payments, it becomes hard to determine the differences between organic content or professional communications and political advertising. As previously said, it was especially the scandal around Cambridge Analytica that made it clear how risky political advertising can be, and how broad the concept of political advertising can become, depending on its definition and on the context.[84] While the CoP still depended on the voluntary cooperation of online platforms, the EU regulation on the transparency and targeting of political advertising, which will entry into force in 2025, has introduced mandatory rules. These cover the labeling and targeting of campaign messages, as well as give authorities the opportunity to impose sanctions on those who violate the rules, including the technology companies that provide advertising services. As such, users of online services are expected to be exposed to less deceptive and manipulative advertisement and will be given tools to understand why certain messages target them.

Regulatory areas	Main regulatory instruments	Main provisions
Liability	E-Commerce Directive (2000/31/EC) DSA (2022/2065) AVMSD (2018/1808) Directive on Combating Terrorism (2017/541) Directive on Copyright in the Digital Single Market (2019/790)	Art. 14 E-Commerce Directive Arts. 4-10 DSA (Chapter 2 – Liability of providers of intermediary services)

---

[83] In the CoP, "trustworthiness" refers to the source or publisher of information. An information publisher can be trusted when the chance that users will be exposed to false or misleading content from that source is relatively low. Furthermore, a reputable publisher is expected to have a process in place to make sufficient and timely corrections, in case it publishes false or misleading content.

[84] In its narrowest sense, political advertising refers to advertising placed by political parties and candidates running for office, with the aim of securing votes. However, Cambridge Analytica's advertisements were not directly placed by a party or candidate, and were not always advocating for a vote – sometimes, they just wanted to convince certain voters to stay at home, and not cast their votes. See C. Timberg- I. Stanley-Becker, *Cambridge Analytica database identified Black voters as ripe for 'deterrence,' British broadcaster says*, in *washingtonpost.com*, 29 September 2020.

Content moderation	DSA (2022/2065) EMFA (2024/1083) CoP (2022) Political Ads (2024/900)	Art. 16 DSA (Notice and action mechanisms) Art. 17 DSA (Statement of reasons) Art. 22 DSA (Trusted flaggers) Art. 18 EMFA (Content of media service providers on very large online platforms)
Content curation	DSA 2022/2065 DRPA (2024/436) EMFA (2024/1083) AVMSD (2018/1808)	Art. 27 DSA (Recommender system transparency) Art. 38 DSA (Recommender systems) Art. 3 EMFA (Right of recipients of media services) Commitments 18-22 CoP Art. 7a AVMSD (N.B. enforced by member states and only for video-sharing platforms, i.e., YouTube) Art. 13 AVMSD (N.B. enforced by member states)
Transparency & accountability	DSA (2022/2065) GDPR (2016/679) DRPA (2024/436) CoP (2022) EMFA (2024/1083)	Art. 15 DSA (Transparency reporting obligations for intermediary services) Art. 24 DSA (Transparency reporting obligations for online platforms) Art. 37 DSA (Independent Audits) Art. 40 DSA (Data access and scrutiny) Art. 42 DSA (Transparency reporting obligations) Art. 18 EMFA, para. 2 and 8 (Content of media service providers on very large online platforms)
Privacy and manipulation	GDPR (2016/679) AIA (2024/1689) Data Governance Act (2022/868) Data Act (2023/2854)	Art. 22 GDPR (Automated individual decision-making, including profiling) Art. 25 DSA (Online interface design and organisation) Art. 5 AIA (Prohibited AI Practices)
Risk management	DSA (2022/2065) DRPA (2024/436) AIA (2024/1689)	Art. 34 DSA (Risk assessment) Art. 35 DSA (Mitigation of risks) Art. 36 DSA (Crisis response mechanism) Art. 9 AIA (Risk management system)

Table 1. An overview of the mentioned provisions affecting social media's opinion power and exposure diversity.

The 2022 Strengthened CoP (just like its 2018 precedent) includes commitments for signatories that require them to propose appropriate definitions for both political and so called "issue advertising" – the latter being paid content not clearly advocating for the support of the candidate, but still capable of influencing electoral decisions –, as well as asking for increased transparency on many aspects.[85] The Regulation (EU) 2024/900 on the transparency and targeting of political advertising is even clearer, and provides a number of binding transparency and integrity obligations, highlighting the need for a comprehensive assessment of what constitutes political advertising.

Finally, addressing the concentration of opinion power in a few platforms may also require actions related to competition law. However, traditional media concentration law seems to be largely ineffective in this domain.[86] In this sense, the Digital Markets Act (DMA) should be considered as the EU regulatory tool targeting dominant platforms, imposing stricter obligations to counterbalance market concentration. The DMA designates platforms providing core services as "gatekeepers" if these have a "significant impact on the internal market" (art. 3)[87] One of the DMA's novel goals is to achieve a contestable platform market for effective market pluralism, leveling the playing field and lowering entry barriers in the platform market[88]. Provisions such as the prohibition of self-preferencing (i.e., favor one's content) and the interoperability of gatekeeper could also have spillover effects on the exposure to media diversity. Furthermore, other EU regulations such as the Data Governance Act 2022/868 and the Data Act 2023/2854 which emphasize fairer access to and sharing of data could be effective in countering platforms' dominant data power and, indirectly, opinion power. From the perspective of external media pluralism, it will be interesting to observe to what extent the DMA and other regulations aimed at restoring competition will foster

---

[85] This is fundamental when it comes to social media's opinion power: if there is no regulation of political advertising on social media in the EU member states (as regulation in EU member states still focuses on traditional legacy media) or the definition of political advertising allows for loopholes (making it possible for manipulative ads to go under the radar), certain members of the online audience will find themselves more exposed to manipulation on issues that are most important for opinion formation related to elections – not even knowing about it.

[86] T. Seipp et al., *Dealing with opinion power in the platform world*, cit.

[87] A platform service is presumed to be a gatekeeper if the company, in each of three consecutive financial years, achieved an annual turnover within the EU of at least EUR 7,5 billion, or if the company's average market capitalisation or its equivalent fair market value amounted to at least EUR 75 billion in the last financial year, or if it had on average at least 45 million monthly active end users established or located in the Union in the last financial year, and at least 10 000 yearly active business users established or located in the Union (art. 3 DMA). As of now, four social media platforms have been designated as core platform services of gatekeepers, namely TikTok, Facebook, Instagram, and LinkedIn. Here can be found the full list of designated gatekeepers.

[88] Two concrete benefits for pluralism could be expected from (i) requirements related to fairness and transparency in crawling, indexing, and ranking, namely requiring non-discrimination in content organization; (ii) privacy protection obligations, strengthening consent requirements from users for targeted content. Thus, the DMA directly addresses the core platform service of data trade, whereas the DSA tries to regulate the user's experience that is provided in exchange of such data. See J. Bayer, *Digital Media Regulation within the European Union. A Framework for a New Media Order*, Baden Baden, 2024, 265. According to Bostoen, «the DMA's contestability goal is reminiscent of the pluralism pursued by the Audiovisual Media Services Directive», see F. Bostoen, *Understanding the Digital Markets Act*, in *The Antitrust Bulletin*, 68(2), 2023, 263 ss.

diversification in the news industry, especially considering how newsrooms are dependent on platforms' logics, so these may have spillover effects.

5. Limits, challenges, and opportunities of the EU governance emerging model

While several mechanisms of transparency have been established by the DSA, one of the main concerns is that these will remain a form of "transparency theater", namely that these measures serve more to legitimize online platforms than to exert pressure on their power structure, considering that they are mostly framed as publicity, procedural fairness, and access to datasets, but they do not necessarily solve the problem of information asymmetry[89] nor the limited resources dedicated to content moderation.[90] While transparency disclosures could be appealing for various subjects such as researchers, lawyers and policymakers, they provide relatively little useful information, especially to users.[91] More broadly, it can be questioned if the emerging EU model enables a governance of content moderation which is iterative and dynamic, forcing social media to truly engage in a dialogue about the value judgements behind their choices, and explain each of them[92]. There are, for example, concerns regarding the implementation's effectiveness of provisions related to content visibility. If art. 17 DSA were applied rigorously and all demotions duly disclosed, it would still not be effective enough to avoid shadowbans. Indeed, what makes demotions particularly problematic is that they can be technically undetectable (for example, social media can minimize the risk of detection by demoting content gradually over time rather than instantaneously). Amplified content, on the other hand, must not even be disclosed in the DSA, albeit such strategy (sometimes referred as "shadow-promotion") has been repeatedly observed, even during Russia's war of aggression in Ukraine[93]. VLOPs' operations in extra-EU countries also remain outside the scope of the legislation, albeit these may still affect European citizens' public opinion. These represent critical limitations of EU regulation. It can even be questioned whether to make VLOPs truly accountable, the amplification of content should be disclosed, beyond the disclosure of general criteria of recommender systems (art. 27 and recital 70 DSA) and despite

[89]   See for example: M. Maroni, *'Mediated transparency': The Digital Services Act and the legitimisation of platform power*, in *Legal Studies Research Paper Series*, University of Helsinki, 2023.

[90]   To illustrate the shortage of human content moderators for languages other than English, the *DSA transparency reports* indicate that in Italy, Alphabet has 229 moderators for YouTube, Meta has 164 for Facebook and Instagram, TikTok employs 439 moderators, and Twitter has just one moderator for the Italian language. All things considered, this amounts to fewer than 850 individuals who understand the Italian language moderating content for approximately 140 million accounts. This indicates that in Italy there is roughly one moderator for every 168,000 accounts.

[91]   A. Trujillo-T. Fagni-S. Cresci, *The DSA Transparency Database: Auditing self-reported moderation actions by social media*. arXiv preprint arXiv:2312.10269, 2023.

[92]   E. Douek, *Content moderation as systems thinking*, cit., 533.

[93]   S. Romano-N. Kerby-M. Schüler-D. Beraldo-I. Rama, *The Impact of TikTok Policies on Information Flows during Times of War: Evidence of 'Splinternet' and 'Shadow-Promotion' in Russia*. AoIR Selected Papers of Internet Research, 2023.

content amplification is hard to assess[94].

Limitations and challenges of the current regulatory framework have been already observed in the recent Israeli-Palestinian conflict. The risk of censorship that shadow-bans entail, for example, when content about the Israeli attacks in the Palestinian territories and posts in support of Palestine have been reportedly demoted by social media.[95] In this case, social media's opinion power remained substantially unchallenged. At the same time, the spread of illegal content following Hamas' terrorist attack in Israel contributed to the first enforcement of the DSA[96]. Criticism was moved by digital rights' civil society organizations towards both the VLOPs involved and the EC for the actions undertaken which may disproportionately affect oppressed groups and human rights defenders[97]. This example shows the difficulties in defining what constitutes violent and harmful content under the DSA. In the occasion of very conflictual and polarizing events/issues, the DSA enforcers' (i.e., public authorities) framing of the situation might boost a certain narrative over another, adding to the issue of opaque content moderation choices by private social media platforms.[98]

Illegitimate influence over public opinion could be unveiled through algorithmic auditing (art. 37 DSA).[99] Their outcome, however, can be limited due to the platform providers' influence in the market.[100] Moreover, due to information asymmetries and superior technological capabilities[101], VLOPs are distinctly better equipped than national governments to monitor disinformation, foreign interference, and other malicious activities. Their advantage lies not only in accessing extensive private and real-time social media data but also in their technical expertise. This critical role inevitably enhances their influence over the formation of public opinion, and it is still unclear whether the EU governance model will be able to make their power fully accountable — as well as

---

[94]   L. Thorburn-J. Stray-P. Bengani, *What will "amplification" mean in court,* in *techpolicy.press,* 19 May 2022; L. Belli- M. Wisniak, *What's in an Algorithm? Empowering Users Through Nutrition Labels for Social Media Recommender Systems,* in *knightcolumbia.org,* 22 August 2023.

[95]   K. Paul, *Instagram users accuse platform of censoring posts supporting Palestine,* in *theguardian.com,* 18 October 2023. Two years earlier, Meta was accused of censoring posts in support of Palestine on its platforms in occasion of the Sheikh Jarrah crisis in 2021. Facebook's Oversight Board entrusted a company to conduct a due diligence exercise that led to recommendations for Meta. See H. Elmimouni et al., *Shielding or Silencing?: An Investigation into Content Moderation during the Sheikh Jarrah Crisis,* Proceedings of the ACM on Human-Computer Interaction, Volume 8, Issue GROUP, Article No.: 6, 1 ss.

[96]   The EC submitted formal requests of information to X, Youtube, TikTok and Meta on risk assessment and mitigation measures against illegal content and disinformation (based on art. 67 DSA); later, it adopted a Communication against hate speech which established a dedicated network of trusted flaggers specialized in antisemitic content online; finally, formal proceedings (based on art. 66 DSA) were opened also against X and TikTok. full list of the main enforcement activities.

[97]   Accessnow, *Precise interpretation of the DSA matters especially when people's lives are at risk in Gaza and Israel,* in *accessnow.org,* 18 October 2023.

[98]   K. Bleyer-Simon-U. Reviglio, *Defining Disinformation across EU and VLOPs,* European Digital Media Observatory, forthcoming.

[99]   P. Terzis-M. Veale-N. Gaumann, *Law and the Emerging Political Economy of Algorithmic Audits,* in *The 2024 ACM Conference on Fairness, Accountability, and Transparency,* June 2024, 1255 ss.

[100]   J. Laux-S. Wachter-B. Mittelstadt, *Taming the few: Platform regulation, independent audits, and the risks of capture created by the DMA and DSA,* in *Computer law & Security review,* 43, 2021, 105613.

[101]   J. Black, *Decentring regulation: Understanding the role of regulation and self-regulation in a 'post-regulatory'world,* in *Current legal problems,* 54(1), 2021, 103 ss.

the emerging powers of the EU institutions.

The role of users in improving content moderation and identifying illegitimate influence over public opinion could also be leveraged. As said above, the out-of-court dispute settlement bodies can be a powerful instrument. However, it has been stressed that there might be issues related to workload, limiting in practice the efficiency and effectiveness of these bodies; however, there is hope that there will be a pull effect: the more users turn to these bodies, the greater the pressure on platforms to comply with the decisions.[102] Other forms of participatory governance should be considered in future policy developments. Most importantly, "social media councils" could be established, which are independent, multi-stakeholder bodies designed to oversee, advise, and sometimes enforce accountability in how social media platforms manage content moderation (similar to X's Community Notes).[103] Users could be further involved, among others, in countering and mitigating impacts of online misinformation[104], or even providing algorithmic feedback; indeed, retrospective, deliberative judgment on previous recommendations could indeed help users to align their preferences with the output of recommender systems.[105] Furthermore, a crowd-sourced approach to identify illegitimate influence such as shadowbans could be envisioned, where users signal their suspicions or evidence, especially during conflicts and high intensity events. In practice, they could already submit these in the DSA whistleblowing platform[106], the decision to disclose this information remains at the Commission's discretion. Developing more transparent and inclusive mechanisms could strengthen the EU's governance approach to social media's opinion power.

Fundamentally, to "redistribute" social media's opinion power, users should be able to consciously decide for themselves what content they want to see in social media. Although art. 27 DSA provides the criteria that recommender systems should be adjustable by users to their preferred preferences, we still don't know what is meant with "criteria" and "preferences" and, in fact, these same criteria are left to VLOPs to choose; moreover, the design of these same options remains at their discretion. Therefore, art. 27 DSA is still unclear in its implementation. Even if a delegated act or a code of conduct would clarify this, what options would be offered to users are clearly hard to

---

[102] H. Ruschemeier-J. P. Quintais-I. Nenadic-G. De Gregorio-N. Eder, *Brave New World. Out-Of-Court Dispute Settlement Bodies and the Struggle to Adjudicate Platforms in Europe*, in *verfassungsblog.de,* 10 September 2024. According to the authors of this blog post, out-of-court dispute settlements bodies have different options, ranging from a limited mandate that only covers the content moderated and not the justification provided by the platform, to a full review of, for example, all the requirements of art. 17 DSA (statements of reason). Also, ODS bodies are likely ill-suited to carry out assessments related to misinformation. Therefore, it would be advisable for them to cooperate with fact-checking organisations as well as with journalists and news media organizations.

[103] M.C. Kettemann-W. Schulz. *Platform://Democracy: Perspectives on Platform Power, Public Values and the Potential of Social Media Councils*, 2023.

[104] The Global Partnership on AI, *Responsible AI for Social Media Governance: A Proposed Collaborative Method for Studying the Effects of Social Media Recommender Systems on Users,* 2021 Report. The Global Partnership on AI.

[105] J. Stray-A. Halevy-P. Assar-D. Hadfield-Menell-C. Boutilier-A. Ashar-N. Vasan, *Building human values into recommender systems: An interdisciplinary synthesis*, in *ACM Transactions on Recommender Systems*, 2(3), 2024, 1 ss.

[106] See *digital-services-act-whistleblower.integrityline.app.*

establish. The risk of further promoting personal relevance at the expense of exposure to more diverse content needs to be carefully weighed by considering customization options. To further empower users, EMFA could complement the DSA's endeavor, since it has recognized the right to customize the media offer (art. 20). This, however, only applies to audiovisual media such as smart TV interfaces and applications. And yet, art. 3 EMFA asserts the right of recipients (i.e., users) « to have access to a plurality of editorially independent media content and ensure that framework conditions are in place (…) to safeguard that right, to the benefit of free and democratic discourse.». In conjunction with the spirit of art. 20 EMFA, as well as art. 10 of the European Convention of Human Rights (ECHR) recognising the right to freely receive and impart information and Article 11 of the EU Charter of Fundamental Rights (CFR), art. 3 can complement artt. 27 and 38 DSA on recommender systems in its future implementation.[107] Furthermore, the forthcoming list of media service providers that will result from self-declarations to VLOPs may also represent an additional option for users to customize their experience by receiving only, or mostly (i.e., prioritize), content from media following professional standards. While this may empower users, however, it may also further strengthen VLOP's opinion power, as they ultimately select these "professional" media service providers.[108]

With regards to the limitations of art. 18 EMFA, the idea of providing a "privilege" for media content is aimed at balancing the asymmetrical relation of power between media service providers, especially smaller ones, and VLOPs.[109] Critics, however, point out the ample discretion left to platforms when accepting the MSPs' self- declarations[110]. As a matter of fact, platforms can- but are not bound- to consult public authorities when deciding if to accept or to reject the status of MSPs declared by a user of their services. Moreover, art. 18 might divert attention from the content of communications towards the standing of the source – as even renowned news media outlets can publish untrue content – and possibly provides a loophole that can be exploited by content publishers that only formally comply with quality standards. The adherence to professional standards should indeed be the basic requirement for the identification of the media service providers that will enjoy the guarantees granted by the EMFA Regulation, also in light of the standards elaborated by Recommendation 2022/1634 "on internal safeguards for editorial independence and ownership transparency", accompanying EMFA[111]. To some extent, this is explicitly specified by art. 18 (1d) EMFA

---

[107] U. Reviglio-M. Fabbri, *Navigating the Digital Services Act: Scenarios of Transparency and User Control in VLOPSEs' Recommender Systems,* in *Proceedings of the 18th ACM Conference on Recommender Systems.*

[108] How EMFA's provisions will be implemented in the single EU Member States, however, remains to be seen, considering that EMFA is a "Regulation of principles"; namely, it is directly enforceable in the national legal frameworks, as EU Regulations are, but is very open in its possible interpretations, as if it was a Directive. See E. Brogi et al., The European Media Freedom Act, cit.

[109] O. Pollicino-F. Paolucci, *Unveiling the Digital side of Journalism: Exploring the European Media Freedom Act's opportunities and challenges*, in *La Revue des Juristes de Sciences Po*, 1, 2024.

[110] D. Tambini, *The EU is taking practical measures to protect media freedom. Now we need theory*, CMPF Discussion Series, 9 May 2023; M. Z. van Drunen-C. Papaevangelou-D. Buijs-R. Ó. Fathaigh, *What can a media privilege look like? Unpacking three versions in the EMFA*, in *Journal of Media Law*, 15(2), 2023.

[111] S. Verza, *What is journalism in the digital age? Key definitions in the European Media Freedom Act,* forthcoming in *Rivista Italiana di Informatica e Diritto.*

as one of the criteria for media service providers to be able to self-declare as media. A related debate in this context is the prominence of Public Service Media and public-interest content online. This is primarily governed by the Audiovisual Media Services Directive (AVMSD) but it does not apply in the online environment.[112] For example, Article 7a of the AVMSD requires EU Member States to ensure the appropriate prominence of audiovisual media services of general interest, but its enforcement is limited to video-sharing platforms (e.g., YouTube). The EMFA and the DSA lack effective mechanisms to guarantee that PSM and public value content are given due prominence online. While the DSA primarily focuses on mitigating harmful and does not place positive obligations on platforms to prioritize public interest content, the EMFA fails to address the need for prioritizing public value content in digital spaces. However, the implementation of the DSA and the revision of the EMFA offer an opportunity to close these accountability gaps within this governance framework.[113]

In the realm of social media's potential manipulative power, the main provisions highlighted above - the prohibition of dark patterns under art. 25 DSA and of subliminal techniques under art. 5 AIA - present significant challenges. On the one hand, dark patterns and nudging techniques constantly change, so there are concerns that it may be hard to identify and promptly ban new ones.[114] On the other hand, AI harms remain hard to detect, and the same harms they cause would be hard to prove.[115] Given the potential risks highlighted in the first paragraph of this article, the fact that the AIA does not consider recommender systems to be a high risk AI technology represents a substantial limitation of the EU approach to tame and redistribute social media's opinion power. To effectively prevent manipulation, more experimental research is needed. This is particularly true for the experimental protocols of VLOPs.[116] At present, however, the data access for research granted by the DSA shows several limitations and, above all, does not allow for experimental tests in social media. Article 40 DSA can represent a watershed in the understanding of how platforms can set the news agenda and, eventually, for collecting evidence for policymaking. However, data access is limited to "vetted researchers", which means researchers affiliated with universities and independent from commercial interests, while excluding traditional watchdogs such as journalists and NGOs. Moreover, the norm protects platform data, user privacy, and trade secrets and further restrictions delimit the grounds that justify data access (most importantly, the "systemic risks"'), what data can be accessed, and more generally how these are managed afterwards.

Finally, in the realm of advertising, there are also possible challenges related to paid

---

[112]   E. M. Mazzoli, *The politics of content prioritisation online governing prominence and discoverability on digital media platforms* (Doctoral dissertation, London School of Economics and Political Science), 2023.

[113]   K. Rozgonyi, *Accountability and platforms' governance: the case of online prominence of public service media content,* in *Internet Policy Review*, 2023, 12(4).

[114]   P. Cesarini, *Regulating Big Tech to Counter Online Disinformation: Avoiding Pitfalls while Moving Forward,* in *medialaws*.eu, 2021.

[115]   D. Acemoglu, *Harms of AI*, National Bureau of Economic Research, Working Paper 29247, September 2021.

[116]   D. Knott-J. Pedreschi- S. Stray-S. Russell, *The EU's Digital Services Act must provide researchers access to VLOPs' experimental protocols*, in *informationdemocracy.org*, June 2024.

and amplified content with the potential of manipulating opinions. The discussion on political and issue advertising has already shown that the categories are not clear-cut, and a lot of new ways of manipulating political opinion can become possible. In traditional journalism, the boundaries between editorial and commercial content are being blurred by native advertising and sponsoring of content, among other things. The targeting and microtargeting of advertising are by now subject to increased scrutiny, but new forms of personalizing advertising messages might emerge in the future. The evolution of sponsored formats with the potential of influencing political processes will therefore require policymakers and other stakeholders to consider political advertising a moving target,and constantly update its definitions to prevent misuses and undue manipulation of opinions.

6. Final remarks

Despite moderately positive expectations of the EU policies' potential in the areas of social media governance and media pluralism, more pessimistic viewpoints have also been raised. Notably, influential EU leaders have suggested the possibility of using the DSA during periods of civil unrest to shut down social media platforms completely.[117] How far this model might go remains to be seen. Given the nascent stage of the DSA's enforcement, and many of the other relevant regulations entering into full force, it would be premature to assess the overall effectiveness as well as the potential unintended consequences of this emerging governance model in taming and redistributing social media's opinion power.[118] Furthermore, despite the EU's regulatory efforts and its rather ambitious, comprehensive approach to digital governance, the influence of U.S. technological and political hegemony remains a subtle yet pervasive factor shaping these strategies. Indeed, almost all social media platforms are US-based and potentially

---

[117] C. Goujard-N. Camut, *Social media riot shutdowns possible under EU content law, top official says*, in politico. eu, 10 July 2023; L. Kayali-El. Bertholomey, *Macron floats social media cuts during riots*, in *politico.eu,* 5 July 2023.

[118] Key elements of the framework are still being rolled out, starting with the first algorithmic audit set for August 2024, which will serve as an initial litmus test for the DSA's ability to open social media's black boxes. Furthermore, the Digital Services Coordinators (DSCs), namely the national authorities critical to the enforcement and oversight of the DSA, are not expected to be fully operational until 2025, as well as the European Board for Digital Services, composed of the DSCs. Similarly, EMFA establishes the European Board for Media Services, gathering representatives of national media authorities (substituting ERGA – the European Regulators Group for Audiovisual Media Services). Due to the complex interplay between the DSA and EMFA regarding fundamental rights and the platforms' operationalisation of the systemic risks, competent authorities will play a crucial role in the process of enforcement. Their role, along with the appointment of trusted flaggers who will assist in identifying and addressing compliance issues, marks a significant step in operationalizing the Act. (See I. Nenadic- E. Brogi, *The Game of Boards: The role of authorities in concerting the Digital Services Act and the Media Freedom Act for protecting media freedom,* in *medialaws.eu,* 28 August 2024.) Additionally, the transition of the Code of Practice on Disinformation into a formal Code of Conduct in January 2025 will introduce enforceable obligations specifically tailored to combat disinformation. This shift, coupled with art. 40 of the DSA allowing access to data for researchers becoming fully operable by the end of the year, underscores the gradual implementation process.

aligned with and weaponized for U.S. interests[119], and it is unlikely to expect a change in this status quo with the creation of a "EU silicon valley", considering the hurdles posed by regulatory differences and bureaucratic complexities across EU member states or even the talent and infrastructural gaps.[120] This dominance, coupled with the rapidly evolving media landscape and the lack of robust evidence to guide effective policymaking, makes regulating social media's opinion power particularly challenging. Additionally, social media's considerable political and technical power may enable them to effectively lobby against and often circumvent regulatory measures. Even if users rely on their service for essential functions such as business, social interaction and, indeed, access to information, these continue to exercise a form of privatized governance through the terms and conditions they impose on users, creating a contractual relationship marked by structural information asymmetries where social media have significantly more information than users. These asymmetries, along with the platforms' business models that prioritize user engagement—a strategy that seem to contribute to most of the unintended and harmful consequences of social media[121]—remain largely unaddressed by current regulations. Most of these Regulations, in fact, find their legal basis in art. 114 of the Treaty on the Functioning of the European Union (TFUE), namely in the objective of harmonizing the internal market.[122] This market-based rationale is arguably not the perfect fit for comprehensively addressing

---

[119] In the realms of national security (e.g., terrorist propaganda, immigration security or foreign influence operations) and public health (e.g., COVID-19), there has been a natural collaboration between U.S. social media companies and the U.S. government. While direct evidence is lacking, the history of secret surveillance initiatives like the PRISM project provides a basis for legitimate speculation that such partnerships might also extend to manipulating other content, particularly in foreign countries and for supporting U.S. interests. This conjecture is further bolstered by the U.S. government's well-documented history of conducting "psychological operations" (PSYOPS), which encompass a set of techniques used by military and non-military organizations to manipulate public perceptions through the deliberate use of information, misinformation, and communication strategies, ultimately influencing decision-making processes and behaviors. Various investigations indeed highlighted shady U.S. army operations conducted by thousands of people secretly employed (see N. Fielding - I. Cobain, _Revealed: US spy operation that manipulates social media_, in _The Guardian_, March 17 2011; William M. Arkin, _Exclusive: Inside the Military's Secret Undercover Army_, in _Newsweek_, May 17 2021). Further suspicions of collusion between the U.S. government and social media also emerged from the 'Twitter files', which revealed a close relationship with the FBI, as well as from the fact that various former CIA agents are working, or have worked, at Meta for issues related to content moderation (see M. Koening, _Spooks infiltrate Silicon Valley_, in _Dailymail_, December 22 2022).

[120] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, _Report on the state of the Digital Decade 2023_.

[121] P. Bengani-J. Stray-L.Thorburn, _What's right and what's wrong with optimizing for engagement. Understanding Recommenders,_ 27 April 2022.

[122] See P. Parcu et al., _Study on media plurality and diversity online,_ cit., 36 ss.; S Vries- O. Kanevskaia-R.de Jager, _Internal Market 3.0: The Old "New Approach" for Harmonising AI Regulation_, in _European Papers,_ 8(2), 2023, 590. It is fundamental to acknowledge that the EU is essentially an economic regulator. As van Drunen et al. argued «the more the EU enacts rules that affect public communication, the more urgent it becomes to integrate the relevant sets of expertise into EU decision making, strengthen the procedures that anticipate broader impact on the marketplace of ideas, as well as re-think more generally the legitimacy the European Union has for adopting speech-related measures under the legal bases to regulate the internal market and protection of personal data». See M. van Drunen-N. Helberger-W. Schulz-C. de Vreese_, The EU is going too far with political advertising!_, in _dsa-observatory.eu,_ 16 March, 2023.

complex phenomena impacting on human rights and societal dynamics caused by the platformization of the public sphere. In fact, it legally recognizes the strong ramifications of power of private actors for citizens' fundamental rights, public interests and social values, breaking down the traditional public-private divide.

To counteract the roots of social media's harms and effectively tame social media's opinion power, more structural interventions could be deployed such as reforming the ad-driven business model, crowd-sourcing content moderation, social media full interoperability, mandating the disclosure of platforms' experiments and non-engagement signals in recommender systems[123], that can be designed to favor societal cohesion[124] or other democratic values[125], but even developing a public service social media or creating a market of recommender systems and let users choose which one to employ[126]. Without a more ambitious interpretation of these EU regulations, as well as the implementation of more structural and radical policies, the EU's governance of social media platforms is unlikely to mitigate the risks stemming from their opinion power—which, as highlighted throughout this paper, not only determines "personalized agenda-settings" but it is closely and concerningly tied to their ability to shape users' worldviews, to manipulate their information behavior and to manage the challenges of contemporary information warfare.

7. Conclusions

This paper has delineated and explored the key elements of the individual and public opinion-shaping power of social media and their interaction with the past and emergent regulatory landscape of the EU. Over time, the EU governance model has become more stringent and dynamic: the terms of service and community guidelines across social media have increasingly been scrutinized and standardized; content moderation practices have evolved from reactive to more proactive measures; recommender systems have been subject to regular transparency provisions and algorithmic audits; and lastly, platforms' interface design is being rethought to prioritize user autonomy, compliance, and the mitigation of systemic risks. The emerging EU regulatory model, however, has also been widely criticized. Above all, because it does not directly "fix" the business model of social media and its inherent objective to maximize for "user engagement" - with all the undesirable consequences this seems to lead to – but it tackles platforms' opinion power mainly by creating indirect incentives, such as mandating transparency measures and the assessment and mitigation of "systemic risks", in addi-

---

[123]   T. Cunningham-S. Pandey- L.Sigerson-J. Stray – J. Allen - B Barrilleaux - B. Rezaei, *What We Know About Using Non-Engagement Signals in Content Ranking,* in *arXiv,* 2024.

[124]   A. Ovadya-L. Thorburn, *Bridging systems: open problems for countering destructive divisiveness across ranking, recommenders, and governance,* in *arXiv,* 2023.

[125]   J. Stray-A. Halevy-P. Assar-D. Hadfield-Menell-C. Boutilier-A. Ashar -N. Vasan, *Building human values into recommender systems: An interdisciplinary synthesis,* in *ACM Transactions on Recommender Systems,* 2(3), 2024, 1-57.

[126]   J. M. Marella, *Middleware Technologies: Towards User-Determined News Curation in Social Media,* in *Cath,* in *UJL & Tech,* 31, 2022, 95.

tion to fines. While this model may be largely beneficial to the "public sphere", in the discussion chapter we briefly highlighted many of the legal and technical challenges and opportunities it faces.

All in all, the EU governance model's effectiveness in governing social media's "opinion power" remains to be seen and it would be premature to draw definitive conclusions. The rationale behind regulations such as the DSA, EMFA, and partly the AIA, nevertheless, must be recognised as an explicit will from the side of EU institutions to rebalance very large private platforms' powers over public opinion, regaining control over public interest objectives and overcoming the traditional presumed neutrality of online platforms regarding content moderation. And yet, many of the limitations we discussed possibly derive from the market-oriented logic of the recent EU media regulations, as articulated in their legal basis -Article 114 TFUE - which may have constrained opportunities for a more profound reconsideration of the dynamics of social media's opinion power.

Our retrospective analysis suggests that despite the gradual and reactive evolution of the EU regulatory model, public pressure has eventually led to more stringent regulation and public oversight. While the effectiveness of transparency measures and user empowerment provisions can still be contested at present, the significant volume of information that will be eventually disclosed can nonetheless be expected to sustain, and possibly enhance, the regulatory impact of this governance model. To comprehensively assess such impact on opinion power in general, and media pluralism in particular, it will be crucial to expand the analysis beyond a user-centered perspective on diversity exposure to understand the overall media ecosystem, including newsrooms, as well as the impacts on external pluralism of new EU laws that act also on competition issues.