COUNTERING 'LAWFUL BUT AWFUL' DISINFORMATION ONLINE: EU-REGULATIONS TARGETING DISINFORMATION ON MAJOR SOCIAL MEDIA PLATFORMS

THERESE ENARSSON[*]

This article examines the European Union's regulatory framework for addressing disinformation, with particular focus on the Digital Services Act (DSA) and the responsibilities of Very Large Online Platforms (VLOPs) and so-called 'lawful but awful' disinformation – harmful yet legal content. The study highlights how the EU avoids content censorship by emphasising systemic risk assessment and mitigation. Notably, disinformation is rarely defined with precision yet is framed as especially problematic when linked to democratic processes, political advertising or crises.

A key finding is the Commission's reliance on soft law to steer platform behaviour. By pushing for soft law instruments the EU can expand some normative influence without crossing into formal content regulation. While VLOPs face substantial responsibilities, the enforcement of these measures remains uncertain. These frameworks reflect a shift toward platform governance through risk-based regulation and soft law, with the Commission increasingly shaping the moderation of harmful online content without legislating censorship.

1 INTRODUCTION

1.1 BACKGROUND

Disinformation has been identified as the most pressing global risk in the short to medium term by the World Economic Forum, also amplifying other risks like state-based armed conflict and extreme weather.[1] In Europe, the European Commission has also identified disinformation as a threat to society and has responded through different means, by using a 'whole-of-society approach',[2] stating that many actors and sectors play a role in countering disinformation. The focus in this article is to analyse one main and fundamental approach by the EU to targeting disinformation – the use of regulatory measures to increase the responsibilities of Very Large Online Platforms (VLOPs).[3]

[*] Associate Professor, Department of Law, Umeå University. ORCID-id: 0000-0003-1184-1385; LinkedIn: https://se.linkedin.com/in/therese-enarsson-bb8a6251; therese.enarsson@umu.se. This work was supported by the Swedish Research Council [grant numbers 2020-02278 and 2022-05414]. No potential conflict of interest was reported by the author.

[1] Mark Elsner, Grace Atkinson, and Saadia Zahidi, 'Global Risks Report 2025' (*World Economic Forum*, 15 January 2025), 4, 8 and 13 <https://www.weforum.org/publications/global-risks-report-2025/digest/> accessed 1 September 2025.

[2] European Commission, 'Strategic Communication and Countering Information Manipulation' (2025) <https://commission.europa.eu/topics/countering-information-manipulation_en> accessed 1 September 2025.

[3] Very Large Online Platforms are here defined in accordance with art 33(1) of the DSA, as online platforms with more than 45 million users. (The same amount of average monthly users defines a Very Large Online Search Engine 'VLOSE').

The dissemination of false and deceptive information to influence public opinion, incite violence or achieve commercial gain is not a new phenomenon. Nor is the widespread circulation of disinformation via social media platforms. What has changed, however, is the arena in which disinformation spreads and the ways it shapes its impact. The digital sphere not only accelerates the dissemination of disinformation but also enables it to reach wide audiences and target specific groups with precision. This has made disinformation a more pervasive and intense societal phenomenon.[4]

One notable example of the challenge posed by disinformation is the phenomenon of filter bubbles – situations where algorithms tailor users' experiences to their existing preferences, thereby reinforcing pre-existing beliefs and perspectives.[5] As a result, people may become increasingly misinformed, caught in a bubble of disinformation – often a mixture of half-truth, falsehoods, and value judgments – reinforced by algorithms and echoed within communities of like-minded individuals.[6] This environment can even draw users into a 'rabbit hole' of disinformation, pushing them toward deeper misinformed, or even radicalized, states.[7]

Disinformation often spreads through processes of algorithmic recommendation and amplification, gradually fostering discord, polarization and radicalization. It does so by employing manipulative tactics designed to maximize user engagement and platform usage.[8] It has also been shown that disinformation plays on strong – and often negative – emotions, which in turn increase both the reach of such content and the intensity of users' interaction with it.[9] Beyond emotional manipulation, the digital sphere enables new forms of behaviour and communication that traditional media cannot, with algorithms amplifying content and non-human actors, such as bots, disseminating material at scale, often simulating human interaction. This combination significantly increases the risks to democratic processes, electoral integrity, and civic discourse.[10]

In response to the societal challenges posed by disinformation, the Digital Services Act (DSA) stands out as the most significant legislative measure, establishing the world's first binding regulatory framework that compels the largest digital platforms to take responsibility for the spread and impact of harmful content online. This makes the EU the first jurisdiction in the world to set such standards. The DSA specifically places stricter responsibilities on

[4] Emiliana De Blasio and Donatella Selva, 'Who Is Responsible for Disinformation? European Approaches to Social Platforms' Accountability in the Post-Truth Era' (2021) 65 American Behavioral Scientist 825, 828.

[5] Eli Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think* (Penguin Books Ltd 2011).

[6] Carlos Diaz Ruiz and Tomas Nilsson, 'Disinformation and Echo Chambers: How Disinformation Circulates on Social Media Through Identity-Driven Controversies' (2023) 42(1) Journal of Public Policy & Marketing 18.

[7] Becca Lewis, 'Rabbit Hole: Creating the Concept of Algorithmic Radicalization' in Johan Farkas and Marcus Maloney (eds) *Digital Media Metaphors* (Routledge 2024).

[8] Will Mbioh, 'Beyond Echo Chambers and Rabbit Holes: Algorithmic Drifts and the Limits of the Online Safety Act, Digital Services Act, and AI Act' (2024) 33(3) Griffith Law Review 189.

[9] Steffen Steinert and Matthew James Dennis, 'Emotions and Digital Well-Being: On Social Media's Emotional Affordances' (2022) 35 Philosophy & Technology 36; Mbioh (n 9).

[10] Philip N Howard and Bence Kollanyi, 'Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum' (*SSRN*, 20 June 2016), 5 <https://papers.ssrn.com/abstract=2798311> accessed 1 September 2025; Andrew Peck, 'A Problem of Amplification: Folklore and Fake News in the Age of Social Media' (2020) 133(529) Journal of American Folklore 329.

VLOPs.[11] In the DSA, it is pointed out that these dominant platforms '[…] can be used in a way that strongly influences safety online, the shaping of public opinion and discourse'.[12] This is a relevant backdrop to understanding how the regulation is supposed to impact illegal or harmful content online, such as disinformation.

As part of the same emerging regulatory field as the DSA, the Council of the European Union approved the EU Artificial Intelligence Act (the 'AI Act') in May 2024.[13] This is the first comprehensive legal framework in the world regulating AI. This regulation will also have some impact on how VLOPs are expected to work with content that can be considered disinformation, most importantly generative AI such as *deepfakes* – manipulated content depicting real people or events that does not necessarily appear to be manipulated to a regular social media user online.[14]

Besides these overarching frameworks, there are several instruments, both hard and soft law, which the Commission itself identifies as targeting disinformation by increasing responsibilities for VLOPs.[15] They especially target areas of hate speech,[16] terrorist propaganda[17] and political advertisement.[18] The latter area has only in recent years received attention as part of overarching EU regulation targeting the online sphere due to the security implications of misleading advertisement.[19]

For the purposes of this article, there is a particularly important soft law instrument addressing disinformation – the Code of Practice on Disinformation – which, as of July 1st, 2025, became a formal code of conduct under the DSA.[20] This instrument is described as becoming '[…] a relevant benchmark for determining DSA compliance regarding disinformation risks for the providers of VLOPs and Very Large Online Search Engines (VLOSEs) that adhere to and comply with its commitments' by the European Commission,

[11] There are specific responsibilities placed on very large online search engines in the DSA as well, but search engines will not be included in this study.

[12] 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) [2022] OJ L277/1 recital 79.

[13] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised 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) [2024] OJ L2024/1689.

[14] ibid recital 134.

[15] European Commission, 'Strengthening Online Platforms' Responsibility' <https://commission.europa.eu/topics/countering-information-manipulation/strengthening-online-platforms-responsibility_en> accessed 1 September 2025.

[16] European Commission, 'The EU Code of Conduct on Countering Illegal Hate Speech Online' <https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en> accessed 1 September 2025.

[17] Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online [2021] OJ L172/79.

[18] Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising [2024] OJ L2024/900 (Regulation on the transparency and targeting of political advertising).

[19] Benjamin Farrand, 'Regulating Misleading Political Advertising on Online Platforms: An Example of Regulatory Mercantilism in Digital Policy' (2024) 45(5) Policy Studies 730, 730–731.

[20] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (16 June 2022), 1(h) <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation> accessed 1 September 2025; European Commission, 'The Code of Conduct on Disinformation' (13 February 2025) <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation> accessed 1 September 2025.

and it is expected to play a significant role in understanding the demands under the DSA.

This wide regulatory approach targeting disinformation is an indicator of an equally wide concept – consisting of acts with vastly different gravity and intent. Disinformation can often be placed in the category of online content that is 'lawful but awful', since disinformation is not necessarily illegal, and it can be understood as the act of intentionally spreading misleading information, potentially causing harm.[21] It is important to note, however, that disinformation may also constitute illegal content, such as hate speech or terrorist propaganda.[22]

Yet, disinformation can be perceived as something much wider than just the spread of illegal material, relating to anything from health and religion to beauty products. It that sense, *disinformation* as a term can be said to have an umbrella character, and disinformative content can spread in different ways with different intent and impact.[23] Legally, more specific definitions of disinformation and its possible illegality must therefore be determined at the national level of EU Member State, some of which already have regulation in place to counter or ban certain defined forms of disinformation,[24] and in some cases through EU-level legislation.

The issue of illegality is closely linked to the liability of VLOPs for the content they host. Under the EU's approach, VLOPs are not held liable for failing to moderate content unless they have been notified of its illegality.[25] This also means that platforms are not explicitly obligated to monitor their services,[26] even if case law from the ECtHR and the ECJ indicates that a level of monitoring is necessary, though only for some illegal content.[27]

The Commission has also emphatically stated that it has no intention of censoring content or deciding what the truth is online. Focusing on platform regulation and platforms' responsibilities to maintain a safe space for users, instead of targeting specific content, has clearly been a way for the EU to balance different interests and rights.[28] Perhaps as a natural

[21] For more on this concept of lawful but awful content, see Anastasia Kozyreva et al, 'Resolving Content Moderation Dilemmas between Free Speech and Harmful Misinformation' (2023) 120(7) Proceedings of the National Academy of Sciences e2210666120.

[22] The matter of defining illegality in hate speech is however not straight forward either, but will not be further discussed in this article. It is analyzed in-depth in Therese Enarsson, 'Navigating Hate Speech and Content Moderation under the DSA: Insights from ECtHR Case Law' (2024) 33(3) Information & Communications Technology Law 384.

[23] Tarlach McGonagle and Katie Pentney, 'From Risk to Reward? The DSA's Risk-Based Approach to Disinformation' in Maja Capello et al (eds), *Unravelling the Digital Services Act package* (European Audiovisual Observatory 2021) 44.

[24] Like Germany with The Network Enforcement Act 'NetzDG - Gesetz Zur Verbesserung Der Rechtsdurchsetzung in Sozialen Netzwerken' <https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html> accessed 1 September 2025, and France with its' legislation: LOI n°2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information (1) 2018 (2018-1202) stating that judges can authorizes removing political disinformation in relation to election campaigns.

[25] Digital Services Act (n 12) Article 6, see also Article 74(1).

[26] Marcin Rojszczak, 'The Digital Services Act and the Problem of Preventive Blocking of (Clearly) Illegal Content' (2023) 3(2) Institutiones Administrationis – Journal of Administrative Sciences 44, 46-47.

[27] See Enarsson (n 22); *Delfi AS v Estonia* App no 64569/09 (ECtHR, 16 June 2015) para 47; Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* EU:C:2019:821 paras 39, 41, 45 and 46; Clara Rauchegger and Aleksandra Kuczerawy, 'Injunctions to Remove Illegal Online Content under the Ecommerce Directive: Glawischnig-Piesczek' (*SSRN*, 31 August 2020) <https://papers.ssrn.com/abstract=3728597> accessed 1 September 2025.

[28] Giovanni De Gregorio and Oreste Pollicino, 'The European Approach to Disinformation: Policy Perspectives' (Institute for European Policymaking, June 2024), 2

response to the vague nature of disinformation as a phenomenon and the regulatory instruments targeting it, the DSA – the most comprehensive framework targeting the responsibilities of VLOPs – was criticized even before coming into full effect for not providing clear demands or tools for addressing disinformation, especially such content that might pose a grave risk to society and be harmful, even when not clearly illegal.[29]

It is clear that disinformation is a key concern for the EU in safeguarding democracy and free elections, but it remains unclear how VLOPs should define disinformation and whether, or to what extent, they are obligated to address it. This article contributes to a deeper understanding of how the EU addresses disinformation through regulatory measures – and whether these efforts extend beyond clearly unlawful content. This can add to research on regulatory responses to disinformation, illuminating the role VLOPs are intended to play in targeting such harmful but lawful content.

1.2 AIM, METHOD AND CONTRIBUTION

This article will therefore focus on three, intertwining, questions:

1) *How is disinformation specified or framed in new regulatory instruments?*

   This research question focuses on the content, actions or types of disinformation that are targeted in these instruments but *not* classified as illegal per se (i.e. hate speech and terrorism propaganda).

2) *What regulatory demands are directed at VLOPs to target or counter disinformation on their platform?*

   This research question specifically targets correlating demands for actions taken by VLOPs as a response to disinformation, as identified above. What content, actions or types of disinformation is to be targeted and how?

3) *What does this indicate regarding the role of VLOPs in countering the dissemination of disinformation in the online sphere, and in turn, what does that say about the regulatory approach used by the EU to target 'lawful but awful' content?*

   This final analysis aims to shed light on whether specific disinformative actions, content or contexts can be identified in these instruments, spanning beyond clearly unlawful content and aiming to clarify the perceived and intended role of VLOPs in the whole-of-society approach to disinformation.

To address these questions, a textual and qualitative analysis of legal material will be conducted, combining a doctrinal analysis with a normative evaluation, focusing on the DSA and, to a lesser extent, the AI Act and the Regulation on the transparency and targeting of political advertising.[30] The AI Act has a limited interplay with regulating disinformation on social media platforms but will be analysed in relation to generative AI, especially regarding risk assessment and risk mitigation in relation to, and as regulated by, the DSA. The Regulation on the Transparency and Targeting of Political Advertising will also be considered

<https://iep.unibocconi.eu/sites/default/files/media/attach/PB19_Disinformation_%20Pollicino.pdf> accessed 1 September 2025.

[29] Ethan Shattock, 'Fake News in Strasbourg: Electoral Disinformation and Freedom of Expression in the European Court of Human Rights (ECtHR)' (2022) 13(1) European Journal of Law and Technology 1.

[30] Regulation on the transparency and targeting of political advertising (n 18).

as part of mapping the emerging regulatory framework on disinformation and will be analysed in relation to the context of disinformative and misleading political ads. The obligation for VLOPs to curtail such content through content moderation is regulated in the DSA, however.[31]

Using that as a point of departure, it will also be necessary to branch out to soft-law instruments, since the area of online regulation is highly directed towards self- and co-regulatory instruments, mainly codes of conduct.[32] Here the Code of Practice on Disinformation will be of particular importance.[33] Taken together, this article aims to establish what legal rules, requirements, or limits are placed on VLOPs to counter disinformation, and how far that reaches beyond clearly illegal content.

Because the DSA and the AI Act are both new regulations in an emerging field, the interpretation and outcome of their implementation could also impact how legislation in other jurisdictions around the world sets out to regulate this worldwide phenomenon.[34]

2 REGULATORY FRAMING OF DISINFORMATION

2.1 DISINFORMATION AS SOCIETAL RISK FOR DEMOCRATIC AND ELECTORAL PROCESSES

As mentioned, the Commission has clearly stated that it will never decide what the truth is online.[35] This has been elaborated on by Martin Husovec, who suggested that demanding VLOPs to target specific, in itself *lawful* content, would be crossing a red line:

> Whenever the Commission tackles problems that are posed also by *lawful* expressions, the mitigation measures that it requires from companies must remain strictly content-neutral. Any attempt by the Commission to prescribe content-specific measures for legal content, such as forcing companies to ban some specific lawful content in terms and conditions, would mean that the Commission assumes the role of a surrogate legislature regarding content. The DSA offers no empowerment for such formal rulemaking. Doing so would mean that the Commission oversteps its competencies. It crosses a red line.[36]

Without being able to define what specific expressions could be disinformative, apart from already unlawful content under national or European legislation, these new regulatory

[31] Liubomir Nikiforov, 'Transparency in Targeting of Political Advertising: Challenges Remain' (*SSRN*, 1 November 2024) <https://papers.ssrn.com/abstract=5054430> accessed 1 September 2025.
[32] Rachel Griffin, 'Codes of Conduct in the Digital Services Act: Functions, Benefits & Concerns' (2024) 2024 Technology and Regulation 167.
[33] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20).
[34] Martin Husovec, 'Rising Above Liability: The Digital Services Act as a Blueprint for the Second Generation of Global Internet Rules' (2023) 38(3) Berkeley Technology Law Journal 884; Anupam Chander, 'When the Digital Services Act Goes Global' (2023) 38(4) Berkeley Technology Law Review 1067; Daphne Keller, 'The EU's New Digital Services Act and the Rest of the World' (*Verfassungsblog*, 7 November 2022) <https://verfassungsblog.de/dsa-rest-of-world/> accessed 1 September 2025; Mateusz Łabuz, 'Regulating Deep Fakes in the Artificial Intelligence Act' (2023) 2(1) Applied Cybersecurity & Internet Governance 1.
[35] Gregorio and Pollicino (n 28).
[36] Martin Husovec, 'The Digital Services Act's Red Line: What the Commission Can and Cannot Do about Disinformation' (2024) 16(1) Journal of Media Law 47, 55.

frameworks still clearly target the spread of disinformation. Therefore, to clarify this broader targeting of disinformation and the demands placed on VLOPs to counter it, it is vital, as a first step, to recognize how disinformation is framed or understood in instruments targeting the phenomenon.

The AI Act and, even more extensively, the DSA frame disinformation as a separate phenomenon from illegal content, and disinformation is understood and described as relating to certain contexts and actions that result in the spread of harmful content. The aim of the DSA – to create safe and trustworthy spaces online – expands the scope beyond simple illegal acts or content.[37] It is clear from the first recitals in the DSA, which frame the aim and purpose of the regulation, that disinformation is viewed and phrased as a risk for society. In recital 2 the efforts or concerns of member States to '*tackle illegal content, online disinformation or other societal risks*' are mentioned, and addressing disinformation as a *societal risk* that can cause *societal harm* is a theme that continues throughout the DSA.[38] This is framed in slightly different ways throughout the DSA, providing guidance on how disinformation is to be understood.

One specific context or type of action that can pose a risk is highlighted in recital 69 DSA, where particular weight is given to users' vulnerabilities being exploited or manipulated through advertising techniques, stating that '[i]n certain cases, manipulative techniques can negatively impact entire groups and amplify societal harms, for example by contributing to disinformation campaigns or by discriminating against certain groups'. Here online platforms are also identified as constituting a higher societal risk than other environments due to the potential for being used for that kind of manipulation.[39] Online platforms are therefore to use neither micro-targeting nor base advertising on the profiling of users, in line with the General Data Protection Regulation (GDPR).[40]

The impact and potential harm of advertising is also reiterated in the Regulation on the Transparency and Targeting of Political Advertising that came into force in 2024.[41] Here, a clear focus is – of course – transparency, more precisely regarding who is advertising, who is paying for ads and for which election. This ties together with safeguarding free elections. An overall objective in the EU is that elections should be free from disinformation and misleading political ads, and the Regulation on the Transparency and Targeting of Political Advertising mirrors that overall aim.[42]

[37] Digital Services Act (n 12) recitals 9, 69, 84, 104. See also Artificial Intelligence Act (n 13) recital 110.

[38] Digital Services Act (n 12) recitals 2, 9, 69, 83, 88, 95, 104.

[39] ibid recital 69.

[40] 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) [2016] OJ L119/1 Article 4(4).

[41] Regulation on the transparency and targeting of political advertising (n 18) recital 6.

[42] Directorate-General for Communications Networks, Content and Technology (European Commission), 'A Multi-Dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation' (Publications Office of the European Union 2018) <https://data.europa.eu/doi/10.2759/739290> accessed 1 September 2025; European Commission, 'Strengthening Online Platforms' Responsibility' <https://commission.europa.eu/topics/countering-information-manipulation/strengthening-online-platforms-responsibility_en> accessed 1 September 2025. There is also a new code of conduct underway, specifically targeting political advertisement, potentially adding to the field of countering disinformation, as well as protecting free elections, see Julian Jaursch, 'Addressing Online Advertising Transparency in the DSA and Beyond' (*Tech Policy Press*, 16 December 2024)

In the Regulation on the transparency and targeting of political advertising, disinformation and misleading advertisements are not defined – but through understanding the purpose of the regulation, conclusions can still be drawn on the type of content that is targeted. Here the focus is to prevent deceiving the public while enhancing open political campaigns and debates, and by doing so also protecting a well-functioning democratic society.[43] This is to be done by limiting election interference and opaque political advertising. It is worth noting, however, that the regulation has already received criticism for referring obligations on content moderation to the DSA and for not providing sufficient guidance for moderation systems – automatic or human-based systems – to successfully identify political ads and make nuanced moderation decisions. Moderation of political content will be especially challenging during the most sensitive periods around elections, with massive amounts of content to moderate, which enhances the need for clear guidelines for VLOPs.[44]

The focus on democratic processes and elections is also prominent in the AI Act. While the AI Act takes a back seat to the DSA in relation to more general obligations to fight disinformation, it specifically highlights the prevalence of disinformation in the context of democracy and free elections.[45] The interplay between the DSA and the AI Act is visible in the phrasing of the act, where the importance of effective implementation of the DSA in relation to detection of the use of AI systems is reiterated, stating that VLOPs must identify and mitigate risks connected to AI-manipulated or AI-generated content, '[…], in particular risk of the actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes, including through disinformation'.[46]

More specifically, disinformation in the AI Act mainly relates to the prevalence of bots and deepfakes on online platforms. This is defined in Article 3(60) and targets AI-generated or manipulated content that can be interpreted as genuine by those who engage with it. The definition in Article 3 (60) encompasses all types of content – from image, audio or video – that depict or resemble existing persons, material objects or events. It is important to note that, to constitute a deepfake, it must resemble or depict something or someone existing in real life, but with a purpose, or at least potential, to mislead.

Fighting disinformation is thus given a prominent – but vague – place in the DSA and the Regulation on the Transparency and Targeting of Political Advertising and AI Act alike, with no clear definitions of the concept. Instead, these regulations focus on the risks that disinformation can cause to society, democracy and electoral processes as well as the responsibility of VLOPs to be aware of and mitigate those risks.

2.2 DISINFORMATION AS *CONTEXT* AND *ACTIONS*, NOT CONTENT

As a stepping stone toward understanding the obligations to counter disinformation beyond

<https://techpolicy.press/addressing-online-advertising-transparency-in-the-dsa-and-beyond> accessed 1 September 2025.

[43] Regulation on the transparency and targeting of political advertising (n 18) see, inter alia, recitals 7 and 74.

[44] Nikiforov (n 31. For more on the need for contextual moderation decisions, see Therese Enarsson, Lena Enqvist, and Markus Naarttijärvi, 'Approaching the Human in the Loop – Legal Perspectives on Hybrid Human/Algorithmic Decision-Making in Three Contexts' (2022) 31(1) Information & Communications Technology Law 123; Enarsson (n 22).

[45] Artificial Intelligence Act (n 13) recitals 110, 120 and 136.

[46] ibid, recital 120.

clearly illegal content, disinformation should be framed as conduct – such as disinformation campaigns – involving rapid dissemination of content. More precisely, it refers to content situated within or concerning certain contexts, particularly elections and democratic or political processes, which may in various ways pose risks to society or result in societal harm. With that being said, the AI Act and the Regulation on the Transparency and Targeting of Political Advertising are two examples of actions or contexts that have been singled out and provide more narrow and more precise aims for what kind of disinformation is to be targeted – with corresponding mentions in the DSA.

On top of this, the DSA also takes a rather wide approach to disinformation and encourages certain actions, such as awareness-raising efforts in relation to the risks of disinformation campaigns. Another such vital aspect highlighted in the DSA is joining self- and co-regulatory instruments, mainly codes of conduct, to strengthen cooperation in countering harmful content that is spread across several platforms online.[47]

This brings us to the most relevant framework within the EU on countering disinformation on VLOPs, alongside the DSA – the strengthened Code of Practice on Disinformation. The Code of Practice was first signed in 2018 but has since been strengthened in 2022. When developed and strengthened, the Code was updated in accordance with suggestions by the Commission, and the preamble already included an intention to become a code of conduct in the, then future, DSA. As mentioned, the Code became a formal code under the DSA in 2025. Yet, The Code of Practice is not originally a product of the European Commission but was shaped by the Signatories. Initially, it was therefore best understood as a self-regulatory instrument, where private companies voluntarily developed their content, rules and more detailed demands for their practices.[48] However, its current and future status is best understood as a co-regulatory instrument between the Commission and the Signatories of the code.[49]

The use of self- and co-regulatory instruments is part of the network of regulatory instruments targeting the online sphere, and other highly important codes of conduct have emerged, such as the EU Code of Conduct on Countering Illegal Hate Speech Online.[50] One major difference is that the Code of Practice on Disinformation also targets mainly *lawful* speech. This brings us back to defining or understanding what disinformation *is* under the emerging regulatory framework. In the Code of Practice on Disinformation, the scope is very broad, but there are clear definitions of what kinds of *actions* are included, and, to some extent, which *actors*.

Here, the Signatories to the Code, including VLOPs like Meta and TikTok, are required to fight against disinformation, including misinformation, meaning that spreading misleading information in a harmful way, with (or even without) intent to cause harm and to a relatively small circle, is included in this definition, as well as the more obvious understanding of disinformation as a tool to mislead and deceive, often with a specific goal in mind, such as

[47] Digital Services Act (n 12) recital 88.

[48] Tony Porter and Karsten Ronit, 'Self-Regulation as Policy Process: The Multiple and Criss-Crossing Stages of Private Rule-Making' (2006) 39 Policy Sciences 41; Carl Vander Maelen, 'Hardly Law or Hard Law? Investigating the Dimensions of Functionality and Legalisation of Codes of Conduct in Recent EU Legislation and the Normative Repercussions Thereof' (2022) 47(6) European Law Review 752.

[49] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20); Griffin (n 32) 174.

[50] European Commission, 'The EU Code of Conduct on Countering Illegal Hate Speech Online' (n 16).

political gain. Furthermore, both foreign and domestic campaigns to influence a certain audience for a specific purpose are included in the Code of Practice, as well as efforts to influence or disrupt the free political will of individuals by foreign state actors.[51] This followed after the Commission encouraged such an inclusion.[52]

The broad definitions of disinformation in EU regulatory instruments can be, and have been, criticized for not providing legal certainty and effectiveness, and this vagueness could be problematic when it comes to protecting freedom of expression.[53] At the same time, the European Commission does not wish to censor content, and trying to clarify every action or type of content that could constitute disinformation is of course highly complex, if not impossible. However, how disinformation is presented in these regulatory instruments still provides insight into their purpose.

The focus must be understood as acts of manipulation and the potential risks posed by certain types of content – particularly in contexts such as public health (as will be shown below) and democratic security – rather than on the precise content of the information itself. Perhaps this shows a potential misconception: maybe VLOPs are not meant to identify and analyse disinformation per se, but rather to assess and mitigate risks.

3 RESPONDING TO DISINFORMATION ON VERY LARGE ONLINE PLATFORMS

3.1 DEMANDS TO ASSESS AND MITIGATE *RISKS* OF LAWFUL DISINFORMATIVE ACTIONS

The turn to risk as a parameter is a common trait in regulations affecting the digital domain in recent years, from the GDPR to the DSA and the AI Act. The matters of risk and societal harm are prominent in the DSA and the AI Act alike. Risk assessments and risk mitigation are at the core of the DSA, and the understanding of these concepts is largely connected to risks for society or individuals. Building on the understanding above, where targeting disinformation goes beyond targeting clearly illegal content, the question becomes 'how do obligations to conduct risk assessment and risk mitigation under the DSA apply to lawful contexts and actions of disinformation as presented above?'

Article 34 DSA, regulating risk assessment, and Article 35, regulating risk mitigation, are very broad. The scope of Article 34 and the risk assessments to be made, include the spread of illegal content on the platforms, actual or potential future infringements of fundamental rights, negative effects on civic discourse and electoral processes, as well as public security. Beyond this, VLOPs are to consider actual or foreseeable negative effects regarding gender-based violence, as well as the protection of public health. VLOPs must also

[51] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20) 1 (a) with footnotes.

[52] Sharon Galantino, 'How Will the EU Digital Services Act Affect the Regulation of Disinformation?' (2023) 20(1) SCRIPTed 89, 126; European Commission, 'Guidance on Strengthening the Code of Practice on Disinformation' (last updated 26 January 2023) <https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation> accessed 1 September 2025.

[53] Ronan Ó Fathaigh, Natali Helberger, and Naomi Appelman, 'The Perils of Legally Defining Disinformation' (2021) 10(4) Internet Policy Review, 18-19 <https://policyreview.info/articles/analysis/perils-legally-defining-disinformation> accessed 1 September 2025.

consider serious negative consequences for a person's physical and mental well-being and protect minors. To describe this as very broad might even be understated. Nonetheless, this broad responsibility shows that VLOPs must conduct risk assessment and mitigation in a wide range of situations that can pose a risk or cause harm to individuals and society, and are therefore risk-based in relation to specific contexts and not explicitly focused on specific content or expressions.[54]

However, to further understand how risk assessment and mitigation in the DSA can apply to disinformation, it is necessary to determine whether disinformation is included in the four categories presented in the DSA that are to be under specific assessment by VLOPs. The categories are: (a) the dissemination of illegal content; (b) any actual or predictable impact of the service on the exercise of fundamental rights; (c) any foreseeable or actual negative effects on democratic processes, civic discourse and electoral processes, or public security; and lastly (d) any actual or foreseeable negative effects relating to the protection of minors, public health or gender-based violence.[55] In different ways, these can all apply to disinformation. As mentioned, disinformation is not automatically defined as illegal but can consist of illegal content or be part of illegal activities, such as spreading terrorist propaganda or illegal hate speech.[56] What is illegal offline should, according to Article 3(h) and recital 12 DSA, be illegal online. As also mentioned, even though disinformation can be classified as illegal content, that is left out of the scope of this article.

Disinformation can also affect the exercise of fundamental rights, for example through aspects mentioned under (c) that can impact free elections. As has been shown so far, this is a clear and expressed aim and focus in the Regulation on the Transparency and Targeting of Political Advertising and the AI Act as well, highlighting that the importance of protecting free elections is part of the reason for countering disinformation.[57] It is also prominently targeted in the Code of Practice on Disinformation. However, it is not expressly referenced in relation to disinformation within the DSA. The European Commission did, however, clarify the need for risk assessment and mitigation leading up to the European Parliament elections in June 2024, though, and highlighted that VLOPs should cooperate with a number of actors, including the EU and national authorities, civil society organizations and other experts to counter disinformation and other manipulation or interference with the election.[58] That VLOPs must guard electoral freedom and identify, analyse and assess interferences with elections is therefore still evident.

The only one of these four categories requiring risk assessment in the DSA that *is* mentioned explicitly in relation to disinformation is the fourth, relating to Article 34(1)(d). In recital 83 it is stated that platforms should assess concerns of manipulation of the

[54] Husovec, 'The Digital Services Act's Red Line' (n 36) 47. This is supposed to make the DSA stand the test of time by being content- and technology-neutral.

[55] Digital Services Act (n 12) Articles 34(1)(a)–(d), recitals 80–83. With regard to embedded AI systems or AI models in designated VLOPS, it is worth noting that since VLOPs must follow the risk-management framework provided in the DSA, the same obligations under the AI Act are to be presumed to be fulfilled if taken such measures, if not significant systemic risks stemming from the use of AI is not covered in the DSA are identified (Artificial Intelligence Act (n 13) recital 188).

[56] Enarsson (n 22).

[57] Regulation on the transparency and targeting of political advertising (n 18), see, e.g., recital 4; Artificial Intelligence Act (n 13) recital 120 for example.

[58] European Commission, 'Commission publishes guidelines under the DSA' (press release, 26 March 2024) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1707> accessed 1 September 2025.

platform, including 'coordinated disinformation campaigns related to public health'. In the following recital 84, it is further stated that VLOPs should not only focus on illegal content, but also on that which is not illegal but contributes to systemic risks, proclaiming that '[…] [p]roviders should therefore pay particular attention on how their services are used to disseminate or amplify misleading or deceptive content, including disinformation'. Recital 83, as well as recital 84, clarify that of risks stemming from the use or design of the service, also relate to inauthentic use of the platforms' services. Here, fake accounts and bots are mentioned as risk factors and ways to rapidly spread either illegal content, or such content that contributes to disinformation campaigns. Acting on conceivable risks is at the core of the AI Act as well, focusing on risks stemming from the use of AI. Here, the AI Act and the DSA are again intertwined.[59]

In recital 136 of the AI Act, it is clarified that the obligations stated in the regulation are crucial for an effective implementation of the DSA, specifying the obligations to enact risk assessment and mitigation of, for instance, disinformation. However, in the AI Act, most of the responsibilities for AI generated content are placed on the deployers[60] or providers[61] of AI systems or models. The AI Act also identifies and classifies AI according to the level of risk a certain system poses to individuals or society, from unacceptable and high risk to limited risk or minimal risk. On this four-step scale, deepfakes generated by AI are classified as a *limited* risk. This means that deployers of, for instance, a deep fake, only have to clarify that the content has been manipulated, 'in an appropriate manner that does not hamper the display or enjoyment of the work'.[62] However, concerns have been raised about this classification for deepfakes due to the transnational nature and potential harms of deepfakes, advocating for a higher level of risk classification.[63] Of course, deepfakes could also be part of widely spread disinformation campaigns that pose a risk to society and individuals, thereby requiring VLOPs to act under the DSA.[64]

Ongoing risk analysis, with corresponding responses from VLOPs, is therefore demanded, even regarding 'lawful but awful' disinformation. It is not the content, its illegality

[59] See Artificial Intelligence Act (n 13) recital 136.

[60] 'Deployer' as defined in Article 3(4) and recital 13 of the Artificial Intelligence Act (n 13), is to be interpreted as natural or legal persons *using* an AI system, except for when an AI system is used for strictly personal reasons.

[61] A 'provider' is, according to Article 3(3), a 'natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge'.

[62] ibid Article 50(4).

[63] Cristina Vanberghen, 'The AI Act vs. Deepfakes: A Step Forward, but Is It Enough?' (*Euractiv*, 26 February 2024) <https://www.euractiv.com/section/artificial-intelligence/opinion/the-ai-act-vs-deepfakes-a-step-forward-but-is-it-enough/> accessed 1 September 2025. If a VLOP develops their own AI systems, the obligations of providers would be applicable to it. According to Article 50 and recital 134 in the AI Act a transparency obligation is placed on providers of AI systems, and deployers who manipulate content to 'clearly and distinguishably' disclose that it has been manipulated and is of artificial origin (see the AI Act, recital 134). This includes AI generated text as well, if it aims at informing the public of matters of public interest. But if it has undergone '[…] human review or editorial control and a natural or legal person holds editorial responsibility for the publication of the content' it can be excluded from such demands. This indicates that it is not the actual content, the (dis)information, that is considered in this recital – only whether it was created by an AI, or if a human was involved in the publication process. If VLOPs use such systems embedded into their services, they are to assess them in accordance with Articles 34 and 35 in the DSA (see the AI Act, recital 118).

[64] Digital Services Act (n 12) recital 104.

or lawfulness, or necessarily how widely disinformation is spread, but the *risk* it poses that must be assessed and subsequently mitigated.

Building on this, there are also demands for VLOPs to manage urgent action in times of crisis. In light of Russia's full-scale invasion of and war of aggression against Ukraine in 2022, an article on crisis was added to the DSA in haste.[65] A crisis is defined here as an occurrence '[…] where extraordinary circumstances lead to a serious threat to public security or public health in the Union or in significant parts of it'.[66] If a crisis occurs, and if it is strictly necessary and proportionate, the Commission can demand that VLOPs activate crisis response mechanisms for a limited time, to urgently answer to that specific ongoing crisis.[67] On top of this, in Article 48 DSA, there are provisions encouraging voluntary crisis protocols for addressing crisis situations. The purpose of this would be to rapidly respond to ongoing disinformation campaigns in a coordinated manner.[68]

Still, much is unclear as to the exact demands placed on VLOPs to counter harmful disinformation on their platforms. Due to the open phrasing in articles 34 and 35 of the DSA – enabling platforms to adopt different systems and designs to enforce these requirements – different platforms can choose different focal points in their assessments, or abide by different understandings of, for instance, disinformation.[69] In Article 34 DSA, and the DSA in general, disinformation is not targeted on the basis of its potential illegality, but through its (non-) compliance with a platform's terms and conditions or terms of service.[70] Adapting and adjusting moderation systems, the use of terms of service, and the design and the systems of the platforms, are factors to consider in these risk assessments and mitigations. However, allowing platforms to find best practices benefits flexibility but could also make it more difficult to evaluate the risk assessment regime of VLOPs, to compare and to find a common standard. How much leeway or flexibility is awarded to VLOPs in practice still remains to be seen. The European Commission will, over time, have to clarify whether platforms have effective risk mitigation measures in place for disinformation, and, if not, what exactly is required.

The VLOP 'X' was the first VLOP against which the Commission initiated formal proceedings, only months after the DSA came into force. The European Commission sent requests to X after alleged dissemination of illegal content and disinformation at scale on the platform.[71] The Commission has since opened a formal investigation into several areas of

[65] Doris Buijs and Ilaria Buri, 'The DSA's Crisis Approach: Crisis Response Mechanism and Crisis Protocols - DSA Observatory' (*DSA Observatory*, 21 February 2023) <https://dsa-observatory.eu/2023/02/21/the-dsas-crisis-approach-crisis-response-mechanism-and-crisis-protocols/> accessed 1 September 2025.

[66] Digital Services Act (n 12) Article 36(2).

[67] ibid Article 36, recital 91. This first after a recommendation from the European Board's for Digital Services (Article 36(1)).

[68] ibid recitals 91 and 108; Baskaran Balasingham and Sofia Minichová, 'The DSA's Crisis Response Mechanism and the Indispensability of Social Media Networks' (2024) 17(30) Yearbook of Antitrust and Regulatory Studies 127, 129–130; Buijs and Buri (n 65).

[69] Balasingham and Minichová (n 68) 128-130.

[70] João Pedro Quintais, Naomi Appelman, and Ronan Ó Fathaigh, 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation' (2023) 24(5) German Law Journal 881, 883. Understood in relation to Article 14 in the DSA, stating that VLOPs must take due regard to fundamental rights in their T&Cs, also show the new and groundbreaking way to delegate responsibilities to VLOPs, which in turn, should delegate responsibilities to users by accepting the T&C of the service.

[71] European Commission, 'The Commission Sends Request for Information to X under DSA' (press release, 12 October 2023) <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4953> accessed 1 September 2025.

content moderation and risk mitigation on the platform. This has not yet been resolved and therefore does not fully clarify the consequences of not moderating disinformation on the platform. The investigation concerns whether X has effective risk mitigation measures in place, such as sufficient moderation resources and adequate moderation systems.[72]

No matter the outcome of such a decision from the Commission, which will add to the understanding of the overarching regulatory measures against disinformation, it is important to recognize the relevance placed on *reach* in relation to risk. Most of the responsibilities imposed on platforms to conduct risk assessment and implement mitigation measures only apply to VLOPs, as in platforms with at least 45 million monthly users. They must comply with the most rigorous rules. This means that other platforms could still have a significant number of users yet not have the same responsibilities to mitigate risks in relation to disinformation. This could indicate a perceived correlation between the reach of certain content and the potential risk for society. The more users that could potentially engage with certain content, the higher the risk.

3.2 PUSHES FOR VOLUNTARY ACTIONS

To guide VLOPs and further clarify their responsibilities, there are explicit references to codes of conduct in the DSA to contribute to its proper application. This is emphasised through Articles 45–47.[73] In recital 104 of the DSA, codes of conduct are mentioned specifically in relation to systemic risks to society and democracy, for example through the dissemination of disinformation. From that same recital it is also clear that 'adherence to and compliance with' a code of conduct by a VLOP could be a suitable mitigating measure for an identified risk and that, if a VLOP refuses to sign or adhere to such a code of conduct, this can be taken into consideration in assessing whether their obligations under the DSA have been met. The importance of VLOPs cooperating is reiterated in both the regulation on the transparency and targeting of political advertising[74] and in the AI Act, which state that VLOPs should adapt AI systems and recommendation systems if needed to mitigate negative effects stemming from personalized recommendations to users. In particular, in order to address problems shared across platforms, VLOPs are expected to join cooperative measures such as codes of conduct.[75]

In other words, directing VLOPs to use codes of conduct for guidance plays a prominent part in the European regulatory framework targeting disinformation. This

[72] European Commission, 'Commission Opens Formal Proceedings against X under the DSA' (press release, 18 December 2023) <https://ec.europa.eu/commission/presscorner/detail/en/IP_23_6709> accessed 1 September 2025. There are some preliminary findings available though, on for example the use of dark patterns to mislead users, European Commission, 'Commission sends preliminary findings to X for breach of DSA' (press release, 12 July 2024) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761> accessed 1 September 2025.

[73] However, the use of codes of conduct is mentioned throughout the DSA, in articles as well as recitals, also highlighting their significance in the regulatory framework the DSA provides, see Griffin (n 32) 169

[74] Regulation on the transparency and targeting of political advertising (n 18) recital 20. Here it is stated that '[t]o counter information manipulation and interference in political advertising, "online platforms" as defined in Regulation (EU) 2022/2065 are encouraged, including through the Code of Practice on Disinformation, to establish and implement tailored policies and other relevant measures, including by means of their participation in wider disinformation demonetisation initiatives to prevent the placement of political advertising containing disinformation'.

[75] Artificial Intelligence Act (n 13) recital 88.

approach also allows the Signatories to go beyond the requirements set by the EU by applying their own terms of service. As mentioned, it would be crossing a red line for the Commission to enforce platforms to ban lawful content in their terms of service,[76] but by allowing and encouraging Signatories of codes of conduct to act in certain ways, the responsibilities of VLOPs to act on disinformation become clearer, nonetheless. As mentioned, in early 2025 the European Commission reiterated that 'the Code will become a relevant benchmark for determining DSA compliance regarding disinformation risks'[77] again highlighting its importance. This would also indicate that the Commission gains more influence over how disinformation is interpreted and countered, bypassing the need for detailed regulation.

As previously noted, the Code of Practice on Disinformation adopts a broad definition of the types of content, particularly in certain contexts, that may be classified as disinformation, as well as the actions associated with it, such as foreign or domestic campaigns aimed at interfering with free elections. The question arises, however, as to how Signatories are to perform their duties under the Code and what this will reveal about their obligations under EU regulation to counter lawful disinformation.

Under commitment 14, in the Code of Practice on Disinformation there are several actions and behaviours that Signatories are to target and review, such as ill-intended deepfakes, fake accounts, account takeovers and amplification using bots.[78] In this regard, the transparency obligation for generative AI (deepfakes) under the AI Act is also reiterated for Signatories developing or operating AI-systems and disseminating content on platforms.[79]

It is noteworthy, however, that the first 13 of the 44 commitments relate directly to a single – albeit broad – context: political or issue-based advertising in its various forms. This is a clear focal point of the Code, as very little else, in terms of contexts or types of content, is singled out. The Signatories to the Code[80] commit to reach a common understanding and definition of what 'political and issue advertising' is, which must align with the understanding in the Regulation on the transparency and targeting of political advertising.[81] They also commit to taking actions such as the following: preventing misuse of advertisement systems; clearly indicating if and how advertisement is allowed on their platform; clearly marking content that is sponsored as paid-for content; implementing systems for verification of providers and sponsors of content; and informing users as to why they are seeing certain sponsored political ads.[82] This is directly linked to the risks with disinformation being spread through the use of advertisement. In commitment 13 it is stated that '[r]elevant Signatories agree to engage in ongoing monitoring and research to understand and respond to risks

[76] Husovec, 'The Digital Services Act's Red Line' (n 36).
[77] European Commission, 'The Code of Conduct on Disinformation' (n 20).
[78] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20) Commitment 14.
[79] Artificial Intelligence Act (n 13) recital 134. See also European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20) Commitment 15.
[80] European Commission, 'Signatories of the 2022 Strengthened Code of Practice on Disinformation' (16 June 2022) <https://digital-strategy.ec.europa.eu/en/library/signatories-2022-strengthened-code-practice-disinformation> accessed 1 September 2025.
[81] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20) Commitment 4, measure 4.1.
[82] ibid Commitments 2, 6, 7, 8, 9 .

related to Disinformation in political or issue advertising'. For example, the use of 'blackout periods' for political or issue advertisements in relation to elections is to be discussed.

As mentioned in the introductory background to this article, the digital sphere can both enhance the spread and specificity in reach in an efficient way, but beyond these digital developments also give rise to and enable new types of behaviour – such as the creation and amplification of content by non-human actors simulating human interaction. The ever-evolving nature of online communication and manipulation is something that is addressed in the Code of Practice where Signatories are required to recognize novel and evolving disinformation risks in relation to political or issue advertising. This work is intended to be coordinated with other Signatories to the Code of Practice, but also with a *Task-force*, to '*identify novel and evolving disinformation risks*'.[83]

This Task-force is interesting in more ways than one, given that under commitment 37 all Signatories must be part of the permanent Task-force, alongside the European Commission – which chairs the Task-force – representatives from European Regulators Group for Audiovisual Media Services (ERGA), the European Digital Media Observatory (EDMO), and representatives of the European External Action Service (EEAS). Other experts can also be included when needed. The Task-force is mandated to establish more clearly focused methods for risk-assessment to be used in times of crisis or other situations requiring extra structure in order to swiftly respond to risks, such as those surrounding elections. That also includes coordination of actions between platforms. It is also to work to set up and share best practises, for instance on flagging of content as potentially disinformative and creating safe designs of platforms.[84] Such methods and collaborative efforts could further refine the work of VLOPs and other Signatories and help operationalize Articles 34 and 35 of the DSA with its demands for risk assessment and mitigation. It indicates the role of the future Code of conduct on disinformation under the DSA, as a dynamic tool against disinformation.

The Code of Practice and its Signatories also have the potential to increase knowledge on disinformation, for example its progression, contexts, prevalence, and audiences with different levels of resilience, as well as its automated amplification.[85] This should be done by developing structural indicators that are to be evaluated to access the effectiveness of the code going forward.[86] Each Signatory also commits to establishing and maintaining Transparency Centres, with publicly accessible information, including structural indicators, but also other relevant information about how their services work.[87] This can enable both users and other interested parties, such as researchers, to better understand the efforts undertaken to address disinformation and its development.

In the Code of Practice on Disinformation, much of the actual detection of suspicious content is 'delegated' to other parties that are to take some kind of action, other than the

[83] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20) Commitment 13, measures 13.1 and 13.2.

[84] ibid Commitment 37, measure 37.2.

[85] Iva Nenadić et al, 'Structural Indicators of the Code of Practice on Disinformation: The 2nd EDMO Report' (2024) <https://edmo.eu/wp-content/uploads/2024/03/SIs_-2nd-EDMO-report.pdf > accessed 1 September 2025. See especially pages 6-8.

[86] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20) IX. Pemanent Task-force, (f), and Commitment 41.

[87] ibid VIII. Transparancy Centre, (a), and Commitments 34 and 35, measure 35.6.

platforms and their moderation systems, namely fact-checkers and the users themselves. Signatories must provide the users with access to functions that enable them to verify the information, such as fact-checking tools where third parties provide verified information or some kind of warning label from reliable sources.[88] The Signatories also commit to cooperate with the fact-checking community, ensuring that such cooperation are solid in terms of funding and neutrality, and when possible, provide them with automated access to information on the platform, as swiftly as possible. This is to guarantee swift and impactful responses from the fact-checkers. Signatories are also required to integrate the findings of fact-checkers on their platform and ensure availability in all languages.[89]

In what must be seen as an important development, in early 2025 Meta changed its policy on disinformation on the platforms Facebook and Instagram, as well as their fact-checking. This shift in Meta's policies is, for now, limited to the United States and nothing is said about challenging the new EU regulations, but it is an interesting development in the approach to disinformation and may indicate something about the need for risk assessments. The spokesperson for the Commission stated that if Meta were to consider such a change in Europe as well, under the DSA a risk assessment submitted to the Commission would be necessary.[90] Considering the strong emphasis on the need for fact-checkers in the Code of Practice on Disinformation, this statement is also noteworthy from that perspective, in terms of the role the Code plays and the strength of certain requirements.

4 CONCLUSIONS: TARGETING 'LAWFUL BUT AWFUL' DISINFORMATION

These concluding remarks aim to summarise and shed light on how the EU targets 'lawful but awful' disinformation by placing responsibilities on VLOPs through these emerging regulatory frameworks. From this study, it is clear that the understanding of what disinformation is and the corresponding demands for VLOPs to counter it are both strongly linked to the assessment and mitigation of risks on platforms. This finding is not in itself all that surprising, given that a focus on platform regulation has been a way for the EU to balance different interests and rights, steering EU policies in general toward risk-based approaches.[91] However, when this is added to how disinformation is discussed in these regulations – as something separate from clearly illegal content – it contributes to a better understanding of when risks will be perceived as having the most harmful impact and the role of VLOPs in countering it.

The definitions of lawful disinformation in these regulations are broad and vague (and often sparse), strengthening the Commission's objective to 'not cross the red line' by censoring specific content on social media platforms. Even so, there is still a clear

[88] European Commission, 'The Code of Conduct on Disinformation' (n 20) Commitments 21 and 22.

[89] European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 20) Commitments 30, 31, 32

[90] Cynthia Kroet, 'Meta Needs to Analyse Risks If It Drops Fact Checkers in EU Too' (*euronews*, 8 January 2025) <https://www.euronews.com/next/2025/01/08/meta-needs-to-analyse-risks-if-it-drops-fact-checkers-in-eu-too-commission> accessed 1 September 2025.

[91] Giovanni De Gregorio and Pietro Dunn, 'The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age' (2022) 59(2) Common Market Law Review 473, 475.

relationship between the demands to assess and counter risks and certain content, or rather, *certain contexts*. Free elections and safeguarding democratic processes, broadly speaking, are emphatically and continuously highlighted in these instruments. This regulatory approach by the EU aims to increase transparency for users: clarifying who is behind messages, why they are shown, and whether the message could be false or misleading. This is intertwined with, and sometimes separate from, transparency in advertising.

From this study it is clear that the legal understanding of risk, as opposed to the intent of actors, is the driving factor behind these responsibilities placed on VLOPs. The intent of actors (state, groups/individuals) is not as relevant as the societal and democratic risk of certain actions in certain contexts. The role of social media platforms and their reach in society is contextually bound to the risk of certain content and actions on platforms – especially on VLOPs.

Another indication of this is the intended role of VLOPs as quasi-collaborators – or perhaps even spokespersons – for the European Commission during times of crisis. The crisis response mechanism in Article 36 of the DSA is a good example of both the understanding of, and intended role for, VLOPs under these emerging frameworks. Not only does the potential to cause harm on major platforms serve as proof of their prominent role, but so does the possibilities of using them to quickly disseminate reliable information. However, this is also a way for the Commission to rapidly reach the public with information it deems important during times of crisis, giving the Commission significant scope to intervene in the moderation process of VLOPs. Though this is only for a limited time (a maximum of three months),[92] it can provide the Commission with a substantial power to seek control of the dissemination of certain content.[93] This has also been criticised because of its implications for freedom of expression.[94] The crisis response mechanism was initially intended as a voluntary measure,[95] however, it has since become evident that the Commission considered the risks posed by the widespread dissemination of disinformation or illegal content during crises sufficiently serious to warrant removing such discretion from platforms.

The fact that VLOPs are clearly singled out in terms of obligations under these regulatory instruments also indicates a perceived relationship between risk and *reach*, as mentioned. The more users that could potentially engage with certain content, the higher the risk for society. However, the dissemination of disinformation across smaller platforms or communities could still cause harm, for example, through radicalisation among conspiracy theorists.[96] To some extent the risks posed on smaller platforms or forums could indirectly be said to be addressed in the DSA through provisions encouraging self- and co-regulatory instruments between and across platforms of different sizes in order to cooperate and

[92] Digital Services Act (n 12) Article 36(3)(c).

[93] Ronan Fahy, Naomi Appelman, and Natali Helberger, 'The EU's Regulatory Push against Disinformation' (*Verfassungsblog*, 5 August 2022) <https://verfassungsblog.de/voluntary-disinfo/> accessed 1 September 2025.

[94] European Digital Rights, 'Public Statement On New Crisis Response Mechanism and Other Last Minute Changes to the DSA' (EDRi, 12 April 2022) <https://edri.org/wp-content/uploads/2022/04/EDRi-statement-on-CRM.pdf> accessed 1 September 2025.

[95] Jörg Frederik Ferreau, 'Crisis? What Crisis? The Risk of Fighting Disinformation with the DSA's Crisis Response Mechanism' (2024) 16(1) Journal of Media Law 57, 58.

[96] McGonagle and Pentney (n 23) 49.

develop best practices.[97] That being said, it is reasonable to impose obligations based on platform size due to VLOPs' dominant positions and reach, as well as the need for financial resources and a well-functioning, adaptable infrastructure. This focus on VLOPs can nevertheless leave blind spots for the dissemination of disinformation.[98]

The efforts to encourage adherence to the Code of Conduct on disinformation are relevant to understanding how VLOPs and other signatories are expected to combat disinformation and, in particular, to understanding how they are supposed to work with assessing and mitigating risk. The DSA does not provide clear guidance as to the specifics of the methodology for this, which also makes the assessment of VLOPs' adherence to the demands difficult and hinders comparisons between platforms. The work of the Task-force to establish more clearly focused methods for risk-assessment could prove useful in structuring the work of VLOPs in line with regulatory demands.

Because the Code of Practice has now become a code of conduct under the DSA, more VLOPs can be expected to join as a way to fulfil their obligations. However, the many explicit efforts to encourage participation in codes of conduct would still not – in themselves – provide much incentive for compliance, since there are no sanctions for non-compliance.[99] With that being said, the VLOP 'X' left the code in 2023, prompting warnings from the Commission that the rules under the DSA still must be followed.[100] Nevertheless, complying with a code of conduct is an indication of compliance with the mentioned demands for risk assessment and risk mitigation, and non-compliance with risk assessment obligations can result in significant fines under Article 74. At the same time, participating and joining a code of conduct does not assure compliance with the DSA,[101] but is surely the strongest indicator for VLOPs of what actions are needed.

Taken together, the demands stated in the Code of Practice on Disinformation are far more precise than those found in the DSA, the AI Act and the Regulation on the Transparency and Targeting of Political Advertising. The Code builds on all three instruments, reiterating the risks of political interference through the use of fake accounts, bots and deepfakes, while highlighting the need for transparency so as to provide users with insight into who is behind an account or message. In addition, Signatories are subject to continuous reporting obligations concerning the policies and routines they apply, as well as the actions they take to moderate and detect content, thereby allowing insight into their ongoing work against disinformation.

How the Code of Practice, now a formal code of conduct under the DSA, is put into action and enforced will have a significant impact on targeting disinformation that does not, in itself, fall into the category of illegal content. This could provide greater clarity regarding the Signatories' compliance with emerging regulatory instruments. In turn, it may contribute

[97] Digital Services Act (n 12) recital 88.

[98] This is also discussed in McGonagle and Pentney (n 23) 49.

[99] Jaursch (n 42); Julian Jaursch, 'What Can DSA Codes of Conduct for Online Advertising Achieve?' (*Interface*, 16 December 2024) <https://www.interface-eu.org/publications/dsa-advertising-codes> accessed 1 September 2025.

[100] Lisa O'Carroll, 'EU Warns Elon Musk after Twitter Found to Have Highest Rate of Disinformation' (*The Guardian*, 26 September 2023) <https://www.theguardian.com/technology/2023/sep/26/eu-warns-elon-musk-that-twitter-x-must-comply-with-fake-news-laws> accessed 1 September 2025.

[101] See Cynthia Kroet, 'Online Platforms Disinformation Code Going Formal, but X Is Out' (*euronews*, 13 February 2025) <https://www.euronews.com/next/2025/02/13/online-platforms-disinformation-code-going-formal-but-x-is-out> accessed 1 September 2025.

to a better understanding of how 'lawful but awful' disinformation should be addressed and help to clarify the role of VLOPs in this regard. Or in the words of Mündges and Park:

> This will be particularly significant when the Code of Practice transitions into a code of conduct under the DSA. As it is likely to be the first one, the way this transition is managed, its impact, and its implementation by regulators and policymakers will largely determine the DSA's effectiveness in regulating harmful but mostly lawful speech.[102]

This will not only demonstrate the DSA's effectiveness but also the influence of the Commission on 'lawful but awful' disinformation. The focus on risk and the need for VLOPs to demonstrate risk mitigation measures are tied to voluntary codes where the Commission holds significant influence. This at least presents the possibility that the Commission may move increasingly toward a more specific focus on content, without formally legislating. This will perhaps not cross, but rather blur, the red line through soft law instruments and compliance mechanisms.

[102] Stephan Mündges and Kirsty Park, 'But Did They Really? Platforms' Compliance with the Code of Practice on Disinformation in Review' (2024) 13(3) Internet Policy Review, 16 <https://policyreview.info/articles/analysis/platforms-compliance-code-of-practice-on-disinformation-review> accessed 1 September 2025.

LIST OF REFERENCES

Balasingham B and Minichová S, 'The DSA's Crisis Response Mechanism and the Indispensability of Social Media Networks' (2024) 17(30) Yearbook of Antitrust and Regulatory Studies 127
DOI: https://doi.org/10.7172/1689-9024.yars.2024.17.30.5

Benjamin Farrand, 'Regulating Misleading Political Advertising on Online Platforms: An Example of Regulatory Mercantilism in Digital Policy' (2024) 45(5) Policy Studies 730
DOI: https://doi.org/10.1080/01442872.2023.2258810

Chander A, 'When the Digital Services Act Goes Global' (2023) 38(4) Berkeley Technology Law Review 1067
DOI: https://doi.org/10.15779/Z38RX93F48

De Blasio E and Selva D, 'Who Is Responsible for Disinformation? European Approaches to Social Platforms' Accountability in the Post-Truth Era' (2021) 65(6) American Behavioral Scientist 825
DOI: https://doi.org/10.1177/0002764221989784

De Gregorio G and Dunn P, 'The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age' (2022) 59(2) Common Market Law Review 473
DOI: https://doi.org/10.54648/cola2022032

De Gregorio G and Pollicino O, 'The European Approach to Disinformation: Policy Perspectives' (Institute for European Policymaking, June 2024) <https://iep.unibocconi.eu/sites/default/files/media/attach/PB19_Disinformation_%20Pollicino.pdf> accessed 1 September 2025

Diaz Ruiz C and Nilsson T, 'Disinformation and Echo Chambers: How Disinformation Circulates on Social Media Through Identity-Driven Controversies' (2023) 42(1) Journal of Public Policy & Marketing 18
DOI: https://doi.org/10.1177/07439156221103852

Enarsson T, 'Navigating Hate Speech and Content Moderation under the DSA: Insights from ECtHR Case Law' (2024) 33(3) Information & Communications Technology Law 384
DOI: https://doi.org/10.1080/13600834.2024.2395579

Enarsson T, Enqvist L, and Naarttijärvi M, 'Approaching the Human in the Loop – Legal Perspectives on Hybrid Human/Algorithmic Decision-Making in Three Contexts' (2022) 31(1) Information & Communications Technology Law 123
DOI: https://doi.org/10.1080/13600834.2021.1958860

Ferreau JF, 'Crisis? What Crisis? The Risk of Fighting Disinformation with the DSA's Crisis Response Mechanism' (2024) 16(1) Journal of Media Law 57
DOI: https://doi.org/10.1080/17577632.2024.2362481

Galantino S, 'How Will the EU Digital Services Act Affect the Regulation of Disinformation?' (2023) 20(1) SCRIPTed 89

Griffin R, 'Codes of Conduct in the Digital Services Act: Functions, Benefits & Concerns' (2024) 2024 Technology and Regulation 167

Howard PN and Kollanyi B, 'Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum' (*SSRN*, 20 June 2016), 5
<https://papers.ssrn.com/abstract=2798311> accessed 1 September 2025
DOI: https://doi.org/10.2139/ssrn.2798311

Husovec M, 'Rising Above Liability: The Digital Services Act as a Blueprint for the Second Generation of Global Internet Rules' (2023) 38(3) Berkeley Technology Law Journal 884
DOI: https://doi.org/10.15779/Z38M902431

——, 'The Digital Services Act's Red Line: What the Commission Can and Cannot Do about Disinformation' (2024) 16(1) Journal of Media Law 47
DOI: https://doi.org/10.1080/17577632.2024.2362483

Kozyreva A et al, 'Resolving Content Moderation Dilemmas between Free Speech and Harmful Misinformation' (2023) 120(7) Proceedings of the National Academy of Sciences e2210666120
DOI: https://doi.org/10.1073/pnas.2210666120

Łabuz M, 'Regulating Deep Fakes in the Artificial Intelligence Act' (2023) 2(1) Applied Cybersecurity & Internet Governance 1
DOI: https://doi.org/10.60097/acig/162856

Lewis B, 'Rabbit Hole: Creating the Concept of Algorithmic Radicalization' in Farkas J and Maloney M (eds) *Digital Media Metaphors* (Routledge 2024)
DOI: https://doi.org/10.4324/9781032674612-10

Mbioh W, 'Beyond Echo Chambers and Rabbit Holes: Algorithmic Drifts and the Limits of the Online Safety Act, Digital Services Act, and AI Act' (2024) 33(3) Griffith Law Review 189
DOI: https://doi.org/10.1080/10383441.2025.2481544

McGonagle T and Pentney K, 'From Risk to Reward? The DSA's Risk-Based Approach to Disinformation' in Capello M et al (eds), *Unravelling the Digital Services Act package* (European Audiovisual Observatory 2021)

Mündges S and Park K, 'But Did They Really? Platforms' Compliance with the Code of Practice on Disinformation in Review' (2024) 13(3) Internet Policy Review <https://policyreview.info/articles/analysis/platforms-compliance-code-of-practice-on-disinformation-review> accessed 1 September 2025
DOI: https://doi.org/10.14763/2024.3.1786

Nenadić I et al, 'Structural Indicators of the Code of Practice on Disinformation: The 2nd EDMO Report' (2024) <https://edmo.eu/wp-content/uploads/2024/03/SIs_-2nd-EDMO-report.pdf > accessed 1 September 2025

Nikiforov L, 'Transparency in Targeting of Political Advertising: Challenges Remain' (*SSRN*, 1 November 2024) <https://papers.ssrn.com/abstract=5054430> accessed 1 September 2025
DOI: https://doi.org/10.2139/ssrn.5054430

Ó Fathaigh R, Helberger N, and Appelman N, 'The Perils of Legally Defining Disinformation' (2021) 10(4) Internet Policy Review <https://policyreview.info/articles/analysis/perils-legally-defining-disinformation> accessed 1 September 2025
DOI: https://doi.org/10.14763/2021.4.1584

Pariser E, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think* (Penguin Books Ltd 2011)

Peck A, 'A Problem of Amplification: Folklore and Fake News in the Age of Social Media' (2020) 133(529) Journal of American Folklore 329
DOI: https://doi.org/10.5406/jamerfolk.133.529.0329

Porter T and Ronit K, 'Self-Regulation as Policy Process: The Multiple and Criss-Crossing Stages of Private Rule-Making' (2006) 39 Policy Sciences 41
DOI: https://doi.org/10.1007/s11077-006-9008-5

Quintais JP, Appelman N, and Ó Fathaigh R, 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation' (2023) 24(5) German Law Journal 881
DOI: https://doi.org/10.31219/osf.io/f2n7m

Rauchegger C and Kuczerawy A, 'Injunctions to Remove Illegal Online Content under the Ecommerce Directive: Glawischnig-Piesczek' (*SSRN*, 31 August 2020) <https://papers.ssrn.com/abstract=3728597> accessed 1 September 2025
DOI: https://doi.org/10.2139/ssrn.3728597

Rojszczak M, 'The Digital Services Act and the Problem of Preventive Blocking of (Clearly) Illegal Content' (2023) 3(2) Institutiones Administrationis – Journal of Administrative Sciences 44
DOI: https://doi.org/10.54201/iajas.v3i2.85

Shattock E, ‘Fake News in Strasbourg: Electoral Disinformation and Freedom of Expression in the European Court of Human Rights (ECtHR)’ (2022) 13(1) European Journal of Law and Technology 1

Steinert S and Dennis MJ, ‘Emotions and Digital Well-Being: On Social Media’s Emotional Affordances’ (2022) 35 Philosophy & Technology 36
DOI: https://doi.org/10.1007/s13347-022-00530-6

Vander Maelen C, ‘Hardly Law or Hard Law? Investigating the Dimensions of Functionality and Legalisation of Codes of Conduct in Recent EU Legislation and the Normative Repercussions Thereof’ (2022) 47(6) European Law Review 752