Unravelling the Digital Services Act package

IRIS *Special*

A publication
of the European Audiovisual Observatory

European Audiovisual Observatory

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

IRIS *Special* 2021-1
Unravelling the Digital Services Act package
European Audiovisual Observatory, Strasbourg 2021
ISBN 978-92-871-9152-6 (print version)

Director of publication – Susanne Nikoltchev, Executive Director
Editorial supervision – Maja Cappello, Head of Department for legal information
Editorial team – Francisco Javier Cabrera Blázquez, Léa Chochon, Sophie Valais, Legal Analysts
European Audiovisual Observatory

Authors (in alphabetical order)
Joan Barata, Oliver Budzinski, Mark Cole, Michèle Ledger, Tarlach McGonagle, Katie Pentney, Eleonora Rosati, Alexandre de Streel

Translation
Julie Mamou, Nathalie Sturlèse, Stefan Pooth, Erwin Rohwer

Proofreading
Anthony Mills, Catherine Koleda, Aurélie Courtinat, Gianna Iacino

Editorial assistant – Sabine Bouajaja
Press and Public Relations – Alison Hindhaugh, alison.hindhaugh@coe.int
European Audiovisual Observatory

Publisher
European Audiovisual Observatory
76, allée de la Robertsau
F-67000 Strasbourg, France
Tél. : +33 (0)3 90 21 60 00
Fax : +33 (0)3 90 21 60 19
iris.obs@coe.int
www.obs.coe.int

Cover layout – ALTRAN, France

Please quote this publication as:
Cappello M. (ed.), *Unravelling the Digital Services Act package*, IRIS *Special*, European Audiovisual Observatory, Strasbourg, 2021

4. From risk to reward? The DSA's risk-based approach to disinformation

Tarlach McGonagle, Institute for Information Law (IViR), Amsterdam Law School

Katie Pentney, DPhil Candidate in Law, University of Oxford

4.1. Introduction

Disinformation is an increasing phenomenon and concern in society, in regulatory and policy-making circles, and in practice for the multiplicity of actors in the information ecosystem. The Covid-19 pandemic – and accompanying 'infodemic'[114] – have accentuated the risks posed by disinformation and the harms it can occasion.

Though the problem of disinformation has been identified and acknowledged by states, online platforms and civil society, several battlegrounds of contestation remain, including how to best address it, who bears responsibility for accompanying line-drawing exercises, and what safeguards must be put in place to ensure the free exchange of information and ideas online. This has resulted, to date, in various (and sometimes divergent) approaches at the domestic and regional levels, ranging from self-regulation by online platforms, to co-regulatory approaches, to State-imposed identification and removal measures.

It is into this complex and rapidly evolving regulatory system that the Digital Services Act (DSA) proposal has been introduced.[115] The DSA represents the next generation of content moderation generally in several respects, including preventive and reactive approaches, differentiation in the obligations imposed on online platforms, and the inclusion of efforts to combat and mitigate the risks and harms of 'lawful but awful' categories of speech, such as disinformation. This chapter focuses on the DSA's risk-based approach, which implements heightened due diligence obligations for very large online platforms (VLOPs) in light of their reach, scale and impact-to-risk ratio. Section 4.2 provides an overview of the European regulatory and policy frameworks on disinformation into which the DSA proposal has arrived. Sections 4.3 to 4.5 introduce the key elements of the DSA's risk-based approach, including the identification of systemic risks, the imposition of mitigation requirements and measures to ensure oversight and accountability. Section 4.6

[114] Joint Statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse and IFRC, "Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation", 2020, https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation.

[115] The Digital Services Act package includes both the Digital Services Act and the Digital Markets Act. The focus of this chapter is on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM(2020) 825, 15 December 2020 [hereafter, DSA], https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN.

offers some preliminary reflections on aspects of the proposal which may benefit from further consideration and reflection.

4.2. The disinformation landscape

It is important to clarify at the outset that the DSA is not – and does not purport to be – centrally concerned with disinformation.[116] Nevertheless, as a piece of flagship, modernising regulation for online services in Europe, it is certainly a relevant reference point. The approach taken in this chapter is thus to position the DSA within the broader, complex regulatory and policy framework governing disinformation, before zoning in on selected features of the DSA that are likely to prove most relevant for countering online disinformation. First, though, a brief survey and analysis of the most salient evolving definitions of 'disinformation' will help to clarify the scope of the term.

4.2.1. Evolving definitions

The argument from truth is one of the most enduring rationales for the protection of freedom of expression. Popularized among others by John Milton in *Areopagitica* and John Stuart Mill in *On Liberty*, this epistemic argument articulates the age-old concern that truth will vanquish falsehood, leading to individual and societal development and enlightenment. The need to counter false news, propaganda and disinformation has been a long-standing preoccupation in international human rights law; these issues were given detailed consideration by the drafters of various international treaties.[117]

Contemporary disinformation is, however, qualitatively and quantitatively different to earlier forms.[118] Its online habitat is an utterly changed information ecosystem with unprecedented opportunities for production, dissemination and amplification. Among the game-changing factors are: the ease with which, and scale on which, disinformation is being produced; the quality and sophistication of the content and output; the speed and effectiveness with which it is being disseminated and amplified; the lasting online presence of disinformation; and the possibilities to remain anonymous while engaging in these processes.

In the online information ecosystem, intermediaries and platforms have emerged as a new generation of information and communication power-brokers. Such is the extent of

[116] A simple word-search reveals seven instances of 'disinformation', some of which are very cursory references and three of which are references to the (title of the) Code of Practice on disinformation.
[117] For a detailed overview and analysis, see: Richter A., "International Standards and Comparative National Approaches to Countering Disinformation in the Context of Freedom of the Media, Vienna: Office of the OSCE Representative on Freedom of the Media", 2019, https://www.osce.org/files/f/documents/2/1/424451.pdf.
[118] McGonagle T., "'Fake news': False fears or real concerns?", *Netherlands Quarterly of Human Rights*, 2017, pp. 203-209, https://journals.sagepub.com/doi/full/10.1177/0924051917738685.

their influence that some commentators speak of the positions of "digital dominance"[119] enjoyed by a coterie of big tech companies and, more generally, of the "platformization" of society.[120] Due to their gate-keeping functions, intermediaries and platforms can facilitate or obstruct access to the online forums in which public debate is increasingly conducted.[121] Intermediaries with search and/or recommendation functions, typically driven by algorithms, have far-reaching influence on the availability, accessibility, visibility, findability and prominence of particular content. This influence is achieved, in part, through the use of algorithmic personalization (or recommender systems).[122] The operators of social network services, for instance, "possess the technical means to remove information and suspend accounts", which makes them "uniquely positioned to delimit the topics and set the tone of public debate".[123] Search engines, for their part, aim to and are able to make information more accessible and prominent. This gives them influence over how people find information and ideas and what kinds of information and ideas they find.[124] All of this has ensured that platforms have clear "discursive significance" in society.[125]

Since 2017, there has been heightened attention to, and engagement with, online disinformation at the European and national levels. At the European level, working definitions of disinformation have been put forward and progressively revised and refined.[126]

[119] Moore M. and Tambini D. (Eds.), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*, Oxford, Oxford University Press, 2018, https://global.oup.com/academic/product/digital-dominance-9780190845124?cc=nl&lang=en&q=9780190845124#.

[120] van Dijck J., Poell T. and de Waal M., *The Platform Society: Public Values in a Connective World*, Oxford, Oxford University Press, 2018.

[121] See, for example, Kuczerawy A., *Intermediary Liability and Freedom of Expression in the EU: From Concepts to Standards*, Cambridge, Intersentia, 2018, Chapters 1 and 2.

[122] For further analysis, see: Cobbe J. and Singh J., "Regulating Recommending: Motivations, Considerations, and Principles", *European Journal of Law and Technology*, 10 (3), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3371830.

[123] Leerssen P., "Cut Out By The Middle Man: The Free Speech Implications Of Social Network Blocking and Banning In The EU", *JIPITEC* 6, 2015, pp 99-119, at 99-100, https://www.jipitec.eu/issues/jipitec-6-2-2015/4271.

[124] See generally: van Hoboken J., *Search Engine Freedom. On the Implications of the Right to Freedom of Expression for the Legal Governance of Web Search Engines*, Alphen aan den Rijn, The Netherlands, Kluwer Law International, 2012, https://dare.uva.nl/search?identifier=df2041ce-167d-4e00-9a06-c3937ec5acca.

[125] Laidlaw E.B., *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility*, Cambridge, Cambridge University Press, 2015, p. 204, https://www.cambridge.org/core/books/regulating-speech-in-cyberspace/7A1E83C71D0D67D13756594BE3726687.

[126] An analysis of national approaches is beyond the scope of this chapter, but see, for a detailed overview and analysis: European Regulators Group for Audiovisual Media Services (ERGA), "Notions of disinformation and related concepts", 2020, https://erga-online.eu/wp-content/uploads/2021/01/ERGA-SG2-Report-2020-Notions-of-disinformation-and-related-concepts.pdf.

Table 2. Working definitions of disinformation

Source	Definition	Year
Information Disorder report[127]	Information that is false and deliberately created to harm a person, social group, organisation or country.	2017
High Level Expert Group (HLEG) final report[128]	All forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit.	2018
Communication,[129] Code of Practice on Disinformation,[130] Action Plan against disinformation[131]	Verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm.	2018
European Democracy Action Plan[132]	False or misleading content that is spread with an intention to deceive or secure economic or political gain and which may cause public harm.	2020
Guidance on Strengthening the Code of Practice[133]	The different phenomena to be addressed, while clearly acknowledging the important differences between them. Disinformation in this sense includes disinformation in the narrow sense, misinformation, as well as information influence operations and foreign interference in the information space, including from foreign actors, where information manipulation is used with the effect of causing significant public harm.	2021

The above table provides a brief overview of the most salient attempts to define 'disinformation' in European regulatory and policy-making circles. The table traces the rapid evolution and progressive refinement of the definition. A first observation is that there has been a move away from the definitional criterion of intention to cause harm. This element, included in the Information Disorder report and the HLEG final report, was problematic,

[127] Wardle C. & Derakhshan H., "Information Disorder: Toward an interdisciplinary framework for research and policy making", Council of Europe DGI 9, 2017, p. 20, https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

[128] "A multi-dimensional approach to disinformation", report of the independent High Level Group on fake news and online disinformation, March 2018, p. 11, https://op.europa.eu/en/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1/language-en.

[129] European Commission Communication, "Tackling online disinformation: a European approach", COM 236 final, Brussels, 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236.

[130] EU Code of Practice on Disinformation, 2018, p. 1, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454.

[131] European Commission and High Representative of the Union for Foreign Affairs and Security Policy, Joint Communication, Action Plan against Disinformation, JOIN36 final, 2018, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52018JC0036.

[132] European Commission Communication, On the European democracy action plan, COM 790 final, Brussels, 2020, p. 18, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:790:FIN.

[133] European Commission Guidance on Strengthening the Code of Practice on Disinformation, COM 262 final, Brussels, 2021, p. 5, https://ec.europa.eu/newsroom/dae/redirection/document/76495.

because it had no etymological basis and it would have entailed evidentiary difficulties (i.e., how to prove intent to cause harm). Subsequent definitions emphasise the intention to deceive and the possibility that "public harm" will be caused. Other definitional emphases reveal prevalent concerns about economic gain (e.g. click-bait) or political motives and (foreign) interference in democratic/electoral processes.

Perhaps the most important upshot of these definitional approaches is that the term has an umbrella character. It covers a range of different types of expression, which are fuelled by different motivations, are spread through different means and have different levels of impact.[134]

Given the persistence of concerns about the possible harmful effects of disinformation, it is important to disaggregate the term and identify specific harms before calibrating appropriate responses. Not all effects are harmful and not all harms are illegal. In fact, most are not illegal and 'harmful' should accordingly not be conflated with 'illegal'. As acknowledged in the DSA Proposal: "There is a general agreement among stakeholders that 'harmful' (yet not, or at least not necessarily, illegal) content should not be defined in the Digital Services Act and should not be subject to removal obligations, as this is a delicate area with severe implications for the protection of freedom of expression."[135] In light of this acknowledgement, the broader regulatory and policy context of the DSA will now be explored to give a sense of how disinformation is governed.

4.2.2. Broader regulatory and policy frameworks

The DSA proposal does not contain a single reference to either the Council of Europe or the European Convention on Human Rights (ECHR).[136] These are striking omissions in light of the proposal's repeated references to the importance of freedom of expression (safeguards). It is, of course, logical to frame the DSA within EU law and useful to explain its consistency with existing and pending EU instruments and initiatives. Nevertheless, given the congruence between the regulatory and policy frameworks of the EU and the Council of Europe, the approaches of both organisations are clearly of mutual relevance.

[134] For a neat visualisation of disaggregation of disinformation/'fake news', see: European Association for Viewers' Interests (EAVI), "Beyond 'Fake News' - 10 Types of Misleading News", Infographic, https://eavi.eu/beyond-fake-news-10-types-misleading-info/.

[135] DSA proposal, p. 9.

[136] Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 4.XI.1950 (as amended by Protocols Nos. 11, 14 and 15, supplemented by Protocols Nos. 1, 4, 6, 7, 12, 13 and 16), https://www.echr.coe.int/Documents/Convention_ENG.pdf.

4.2.2.1. Council of Europe

Over the years, the European Court of Human Rights has developed a large body of case-law that offers robust protection for the right to freedom of expression.[137] In relation to disinformation, notable emphases[138] include a firm commitment to strengthening public debate and, specifically in relation to elections, a recognition that "free elections and freedom of expression, particularly freedom of political debate, together form the bedrock of any democratic system".[139] The quality, accuracy and reliability of information during election periods are of crucial importance for an informed electorate.[140] The Court has also held that "Article 10 of the Convention as such does not prohibit discussion or dissemination of information received even if it is strongly suspected that this information might not be truthful. To suggest otherwise would deprive persons of the right to express their views and opinions about statements made in the mass media and would thus place an unreasonable restriction on the freedom of expression set forth in Article 10 of the Convention".[141]

Building on the ECHR and the Court's case-law, the Council of Europe's Committee of Ministers has in recent years adopted recommendations to member States on topics such as: media pluralism and transparency of media ownership; the roles and responsibilities of Internet intermediaries, and the human rights impacts of algorithmic systems.[142] In May 2021, the Steering Committee for Media and Information Society (CDMSI) adopted a Guidance Note on best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation.[143]

Draft and ongoing standard-setting work for the Committee of Ministers include focuses on: ensuring a favourable environment for the practice of quality journalism in the digital age; impacts of digital technologies on freedom of expression; and election communication and media coverage of electoral campaigns.[144] Of particular relevance is the Committee of Experts on Media Environment and Reform's ongoing work on "guiding principles for media and communication governance in order to address the shift from

---

[137] Voorhoof D. et al and McGonagle T. (Ed. Sup.), *Freedom of Expression, the Media and Journalists: Case-law of the European Court of Human Rights*, 6th edition, IRIS *Themes*, European Audiovisual Observatory, Strasbourg, 2021, https://rm.coe.int/iris-themes-vol-iii-2020-edition-en-28-april-2021-/1680a24eee.

[138] For more detailed analysis of a wider range of relevant, specific references, see: van Hoboken J., Appelman N., Ó Fathaigh R., Leerssen P., McGonagle T., van Eijk N. & Helberger N., *The legal framework on the dissemination of disinformation through Internet services and the regulation of political advertising*, report for the Ministry of the Interior and Kingdom Relations, Amsterdam, Institute for Information Law (IViR), University of Amsterdam, 2019, Chapter 4, (hereafter, 'IViR disinformation and political advertising study')), https://www.ivir.nl/publicaties/download/Report_Disinformation_Dec2019-1.pdf.

[139] *Bowman v. the United Kingdom*, 19 February 1998, § 42, Reports of Judgements and Decisions 1998-I.

[140] *Orlovskaya Iskra v. Russia*, no. 42911/08, § 110, 21 February 2017. See also the discussion of *Brzeziński v. Poland*, no. 47542/07, 25 July 2019, and other relevant case-law, in IViR disinformation and political advertising study, p. 53.

[141] *Salov v. Ukraine*, no. 65518/01, ECHR 2005-VIII (extracts), para. 113.

[142] For an overview of the adopted texts, see: https://www.coe.int/en/web/freedom-expression/committee-of-ministers-adopted-texts.

[143] See: https://rm.coe.int/content-moderation-en/1680a2cc18.

[144] For details of these focuses and activities, see: https://www.coe.int/en/web/freedom-expression/committees.

established channels to social networks and of related risks (manipulation of public opinion, lack of public trust, information disorder)".[145]

The foregoing demonstrates an extensive and probing engagement with different dimensions of disinformation through a range of complementary focuses rather than frontal engagement in one single standard-setting instrument. The breadth and depth of this engagement could provide useful guidance for EU regulatory and policy initiatives on similar issues, including the DSA's stated commitment to safeguarding the right to freedom of expression.[146]

4.2.2.2. European Union

The relevant EU regulatory and policy framework comprises various instruments that address different aspects of disinformation in keeping with their respective focuses.[147] The most explicit and detailed engagement with disinformation can be found in the self-regulatory Code of Practice on Disinformation. The Code of Practice was agreed on and signed by representatives of several online platforms, social networking service operators and advertising companies (hereafter "signatories") at the end of September 2018.[148] This initiative was taken in the context of a wider range of efforts by the EU to combat online disinformation, including (around the same time) the Commission's Communication, "Tackling online disinformation: A European approach'"(April 2018),[149] and an Action Plan against Disinformation (December 2018).[150]

The main aims of the Code of Practice include:

- Ensuring transparency about sponsored content, in particular political advertising, as well as restricting targeting options for political advertising and reducing revenues for purveyors of disinformation;
- Providing greater clarity about the functioning of algorithms and enabling third-party verification;
- Making it easier for users to discover and access different news sources representing alternative viewpoints;

[145] Source: https://www.coe.int/en/web/freedom-expression/msi-ref.

[146] For further discussion, see Barata J., "The Digital Services Act and its Impact on the Right to Freedom of Expression: Special Focus on Risk Mitigation Obligations", Plataforma por la Libertad de Información, n.d., https://libertadinformacion.cc/wp-content/uploads/2021/06/DSA-AND-ITS-IMPACT-ON-FREEDOM-OF-EXPRESSION-JOAN-BARATA-PDLI.pdf.

[147] See Van Hoboken et al., "The legal framework on the dissemination of disinformation through Internet services and the regulation of political advertising, final report", Chapter 5, 2019, https://www.ivir.nl/publicaties/download/Report_Disinformation_Dec2019-1.pdf. See also Chapter 3 of this publication.

[148] EU Code of Practice on Disinformation, 2018, https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.

[149] European Commission, "Tackling online disinformation: A European approach", COM 236 final, 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236.

[150] European Commission and High Representative of the Union for Foreign Affairs and Security Policy, Joint Communication, Action Plan against Disinformation, JOIN36 final, 2018, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52018JC0036.

- Introducing measures to identify and close fake accounts and to tackle the issue of automatic bots;
- Enabling fact-checkers, researchers and public authorities to continuously monitor online disinformation.

The Code of Practice sets out a list of detailed commitments, which are structured around five main pillars: A. Scrutiny of ad placements; B. Political advertising and issue-based advertising; C. Integrity of services; D. Empowering consumers; and E. Empowering the research community. Each signatory chooses the most relevant commitments for its own company – in the light of the services it offers and actions it performs. [151]

The latest and most significant development from the DSA perspective is the process to strengthen the Code of Practice against disinformation. One key aim is to develop it into a co-regulatory instrument, for which the DSA's envisaged approach to addressing systemic risks linked to disinformation (discussed in the next section) will be important. The plans to strengthen the Code of Practice involve addressing a number of horizontal issues: reinforced commitments to achieve the Code's objectives; expanded scope; broadened participation; tailored commitments; (further support for the) European Digital Media Observatory; Rapid Alert System. Specific issues to be addressed in detailed fashion are: scrutiny of ad placements; political advertising and issue-based advertising; integrity of services; empowering users; empowering the research and fact-checking community; and monitoring of the Code.

4.3. Introducing the DSA's risk-based approach

With this backdrop in mind, we turn now to the DSA's risk-based approach to content moderation. The approach – particularised in Section 4 of the DSA – aims to address harmful, but lawful, content online. It is differentiated in application and holistic in scope. There are three elements in particular which warrant closer inspection – along the lines of who it applies to; what it requires; and why it has been included.

The DSA places heightened due diligence obligations on so-called 'very large online platforms' (VLOPs), in essence those providing services to 45 million or more monthly active recipients in the Union.[152] While this includes existing 'tech giants', such as Facebook, Twitter and YouTube (Google), the Act seems to contemplate the evolution and rapid expansion of online platforms by including an ongoing review and verification process. In particular, the Digital Services Coordinator[153] must verify online platforms' monthly active recipients at least biannually, and designate (or terminate designations) of VLOP status accordingly.[154]

[151] For an overview of the implementation of the Code of Practice, see: https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.
[152] DSA, Article 25 § 1. The Article also provides a methodology for how to calculate this figure (see Article 25 §§ 2 – 4).
[153] See also Chapter 3 of this publication.
[154] DSA, Article 25 § 4.

But what does this designation require and compel? Under Article 26(1) of the DSA, VLOPs must "identify, analyse and assess" – on at least an annual basis – "any significant risks stemming from the functioning and use made of their services in the Union".[155] The risk assessment must be specific and differentiated according to the services they provide, and must include specified 'systemic risks'.[156] In addition to the dissemination of illegal content and "any negative effects for the exercise of fundamental rights",[157] the assessment must cover:

> (c) intentional manipulation of their service, including by means of inauthentic use or automated exploitation of the service, with an actual or foreseeable negative effect on the protection of public health, minors, civic discourse, or actual or foreseeable effects related to electoral processes and public security.

This list is non-exhaustive, and additional systemic risks may specifically be assessed; however, it is this last inclusion that is most relevant to disinformation.[158] While the language appears quite broad, there are important restrictions which may limit its application. In particular, the risk assessment includes only 'intentional' manipulations of service, and would seemingly apply to coordinated disinformation networks and campaigns. In addition to the required element of intention, the risk assessment applies only to manipulations with actual or foreseeable negative effects on designated categories of harm – including public health and civic discourse – and democratic pillars such as elections and public security.[159] In this way, the risk assessment is both a product of its time and an attempt to reckon with some of the key concerns within the EU and beyond its borders, including the COVID 'infodemic', electoral tampering, and concerted disinformation campaigns targeting vulnerable groups within society.

While the risk assessment envisioned by Article 26(1) is largely outward-looking, Article 26(2) compels VLOPs to look inward. In particular, VLOPs must "take into account" how their business models and design features – such as content moderation, recommender and advertising systems – influence the systemic risks referred to in paragraph 1.[160] This includes "'the potentially rapid and wide dissemination of illegal content and of information that is incompatible with their terms and conditions".[161] While this latter inclusion is broad enough to encompass content that online platforms want to avoid on their platforms, such as nudity or spam, it may also be relevant for 'lawful but awful' speech, such as disinformation.

The final consideration is why these risk assessment obligations were imposed at all. The recitals provide some insight in this regard. The recitals note that VLOPs "are used in a way that strongly influences […] the shaping of public opinion and discourse" and

---

[155] DSA, Article 26 § 1.
[156] Ibid.
[157] DSA, Article 26 §1 (a) and (b), respectively.
[158] Disinformation may also give rise to 'negative effects' for the exercise of freedom of expression, including the public's right to be 'properly informed' (see *Sunday Times v. United Kingdom (no. 1)*, App no 6538/74 (ECHR, 26 April 1979) at § 66.
[159] DSA, Article 26 § 1 (c).
[160] DSA, Article 26 § 2.
[161] ibid.

highlight the social concerns caused by the advertising-driven business model and design choices of these platforms.[162] The need to govern the 'new governors'[163] is also mentioned: "In the absence of effective regulation and enforcement, they can set the rules of the game, without effectively identifying and mitigating the risks and the societal and economic harm they can cause".[164]

There are several takeaways from the definitional elements summarised above, particularly in relation to disinformation. First, the focus on 'very large' platforms seems to equate reach with risk. Global platforms such as Facebook and Twitter may allow for greater and swifter dissemination of disinformation. However, disinformation campaigns have also arisen on smaller and peer-to-peer networks, which are not subject to the risk assessment obligations of their larger counterparts.[165] Such disinformation campaigns can also cause a range of harms; their impact can be intense even with limited reach, for example within so-called echo chambers or filter bubbles, especially involving hardened conspiracy theorists. The DSA's prioritisation of risk assessment and mitigation for very large platforms is understandable, but it only addresses a particular type of systemic risks.

Second, the risk assessment heeds calls for greater focus on context, rather than content.[166] This is evident from the inclusion of both the intention underlying and the negative effects of online manipulation in the third systemic risk category. However, it remains to be seen how VLOPs may assess the particularised risks – and specific context – arising in different member states of the European Union, where disinformation sources and targets may vary.

Finally, disinformation campaigns have been shown to be fast-moving and adaptive,[167] which may pose a challenge for the annual – rather than ongoing or *ad hoc* – and rigid risk assessment procedure. The process simply imposes minimum thresholds, however; VLOPs are free to conduct additional risk assessments, should they see fit, and to broaden the kinds of 'systemic risks' to be assessed. Whether and to what extent they will do so remains to be seen.

---

[162] DSA, Recital 56.

[163] See Klonick K., "The New Governors: The People, Rules and Processes Governing Online Speech" *Harvard Law Review* 131 (1598), https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf.

[164] DSA, Recital 56.

[165] EU DisinfoLab, position paper: "How the Digital Services Act (DSA) Can Tackle Disinformation", 2021, p. 2, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12417-Digital-Services-Act-deepening-the-internal-market-and-clarifying-responsibilities-for-digital-services/F2164131_en.

[166] Mozilla EU Policy, Mozilla position paper on the EU Digital Services Act, 2021, p. 6, https://blog.mozilla.org/netpolicy/files/2021/05/Mozilla-DSA-position-paper-.pdf/. On the importance of context in content moderation generally, see Land M.K. & Hamilton R.J., "Beyond Takedown: Expanding the Toolkit for Responding to Online Hate", in (Dojcinovic P., ed.) *Propaganda, War Crimes Trials and International Law: From Cognition to Criminality* 143Routledge, 2020,and York J.C. and Zuckerman E., "Moderating the Public Sphere" in (Jørgensen R.F. ed.) *Human Rights in the Age of Platforms* (, MIT Press, 2019.

[167] On the use of smaller, peer-to-peer networks to spread 'lawful but awful' speech, see Bevensee E. and Rebellious Data LLC, "The Decentralized Web of Hate: White Supremacists are Starting to Use Peer-to-Peer Technology. Are we Prepared?", 2020, https://rebelliousdata.com/wp-content/uploads/2020/10/P2P-Hate-Report.pdf.

4.4. Mitigating risks

In contrast to the more rigid formulation at the stage of identifying systemic risks, the DSA takes a more flexible, co-regulatory approach to how such risks must be mitigated. The key features – including the mandatory and permissive wording, and the differentiated actors involved – are set out here before highlighting how such measures may apply to disinformation.

The mitigation measures are set out in Article 27 of the Regulation, mandating that VLOPs "put in place reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified pursuant to Article 26".[168] While the requirement to put in place mitigation measures is mandatory, the list of measures enumerated to fulfil this requirement is permissive and non-exhaustive. Moreover, the measures encompass a broad conception of 'mitigation', from adapting decision-making processes, design features like content moderation and advertising systems and service functions, to strengthening risk detection systems; from cooperating with 'trusted flaggers' and other online platforms through codes of conduct and crisis protocols, to targeted measures to limit the display and reach of advertising on the platform itself.[169]

Article 27 also envisions a role in mitigating risks for other actors, including the European Board for Digital Services (comprised of Digital Services Coordinators)[170] and the European Commission. Article 27 § 2 requires that the Board, together with the Commission, publish comprehensive reports which identify and assess the most prominent and recurrent systemic risks reported by VLOPs or identified by other means, as well as best practices for VLOPs to mitigate the systemic risks identified.[171] The reports must be published on an annual basis.[172]

In addition, the Commission "may issue general guidelines on the application of paragraph 1 in relation to specific risks, in particular to present best practices and recommend possible measures".[173] While the legal force or binding effect of these guidelines remains unclear, and there appears to be broad scope with regard to which 'specific risks' might be addressed, there are two caveats which are noteworthy. The first is the requirement that the Commission have "due regard to the possible consequences of the measures on fundamental rights enshrined in the Charter of all parties involved".[174] Whether this requirement will be of any import or consequence will depend on the level of rigour of the Commission's engagement. By contrast, the "due regard" requirement is not imposed on VLOPs and is not included in the Board's annual reporting under Article 27 § 2. The second caveat provides, "When preparing those guidelines, the Commission shall organise public consultations."[175] The precise nature of these public consultations (how many must be held, what role the public's feedback ought to play, and the like) remains unclear.

[168] DSA, Article 27 § 1.
[169] Ibid.
[170] DSA, Articles 47 and 48.
[171] DSA, Article 27 § 2.
[172] Ibid.
[173] DSA, Article 27 § 3.
[174] Ibid.
[175] Ibid.

However, this inclusion provides an additional layer of public involvement and public oversight, and may ensure that those affected by the DSA's risk assessment process are afforded an opportunity to be heard.

There are several aspects of these provisions which are relevant for disinformation. First, the DSA envisions a co-regulatory approach to risk mitigation which includes multiple stakeholders playing differentiated roles in accordance with their skills and responsibilities. However, in contrast to many of the co- and self-regulatory frameworks which came before, such as the Code of Practice (discussed above), the DSA goes further in requiring certain ends (the adoption of "reasonable, proportionate and effective mitigation measures") while leaving flexibility in the means employed to achieve them. In this sense, Article 27 reflects a shift of emphasis from conduct to result, from process to output. However, while VLOPs remain, at present, free to determine the measures they will put in place – including whether they are put in place globally or within certain regions or States – the requirements of reasonableness, proportionality and effectiveness, may circumscribe the level of flexibility and permissiveness the language would otherwise suggest.

Second, the requirement that the mitigation measures be "tailored to the specific systemic risks identified pursuant to Article 26" may compel VLOPs to explore new ways and share best practices to address systemic risks, including the spread of disinformation on their platforms, with an urgency and vigour not seen to date. Disinformation poses a complex and nuanced challenge, but the imperative of combating its effects on public discourse and democratic pillars has largely befallen (and befuddled) States. The imposition of mitigation requirements to effectively address systemic risks – including the threats to civic discourse, public health, electoral processes and public security posed by disinformation – makes this a problem to be solved for VLOPs as well, and in the process, may harness the resourcefulness and efficiencies of the private sector.

Finally, the sharing of best practices to mitigate systemic risks – in annual reports by the Board and general guidelines issued by the Commission – may further increase the impacts of the DSA's risk-based approach beyond Europe's shores. Many States are keeping a keen eye on the drafting process, and will no doubt seek to model and import the mitigation measures envisioned, the best practices shared, and the general guidelines issued. Similarly, smaller platforms and peer-to-peer networks not subject to the same due diligence requirements may also take note of and seek to implement best practices to mitigate risks on their platforms. Given the potential reach of the mitigation measures, it is critical that VLOPs, the Board and the Commission consider the potential consequences for fundamental rights of any measures contemplated.

4.5. Ensuring oversight and transparency

The DSA rounds out its risk-based approach with provisions dedicated to monitoring, oversight and transparency. Two provisions in particular warrant closer inspection: the audit and reporting requirements set out in Articles 28 and 33, respectively.[176]

By virtue of Article 28, VLOPs shall be subjected to annual audits to assess compliance with, *inter alia*, their obligations to conduct assessments of, and adopt mitigation measures to combat, systemic risks.[177] In addition, the audits will assess compliance with any commitments made by VLOPs under codes of conduct and crisis protocols.[178] To comply with the provision, several benchmarks must be met. In particular, the audits must be performed by organisations which are independent from the VLOP under scrutiny; the organisations must have proven expertise in risk management, technical competence and capabilities; and the organisations must have "proven objectivity and professional ethics, based in particular on adherence to codes of practice or appropriate standards".[179] In terms of results, the organisations must produce audit reports which include (at a minimum) descriptions of the elements audited and the methodology used, the main findings drawn, and an opinion on whether the VLOPs complied with their obligations and commitments.[180] The opinion must include a ranking of either positive, positive with comments, or negative. [181]

The issuance of a negative audit opinion has two notable implications, the first for the auditing organisation and the second for the VLOP. First, the audit report must then include "operational recommendations on specific measures to achieve compliance".[182] Second, upon receipt of such a report, the VLOPs "shall take due account of any operational measures addressed to them with a view to take the necessary measures to implement them".[183] In particular, VLOPs must – within one month of receipt of the recommendations – adopt an audit implementation report, setting out those measures or – in the event they do not implement the recommendations – justifying their reasons for not doing so and setting out any alternative measures to address non-compliance.[184]

The DSA also includes heightened "transparency reporting obligations" for VLOPs under Article 33. VLOPs must make publicly available – and transmit to the Digital Services Coordinator – the elements of the risk-based approach outlined above, including:

(a) a report setting out the results of the risk assessment under Article 26

[176] There are additional elements – including provisions specifically addressing recommender systems as well as data access and scrutiny – but we have chosen to focus on these two requirements specifically, as they flow from the identification and mitigation requirements discussed previously.
[177] DSA, Article 28 § 1 (a). These obligations, set out in Articles 26 and 27, fall within Chapter III of the DSA, to which Article 28 § 1 (a) makes explicit reference.
[178] DSA, Article 28 § 1 (b). The code of conduct is provided for in Articles 35 and 36 (code of conduct) and the crisis protocols in Article 37.
[179] DSA, Article 28 § 2.
[180] DSA, Article 28 § 3.
[181] Ibid.
[182] DSA, Article 28 § 3 (f).
[183] DSA, Article 28 § 4.
[184] Ibid.

(b) the risk mitigation measures identified and implemented under Article 27
(c) the audit report under Article 28(3)
(d) the audit implementation report under Article 28(4)[185]

These transparency reporting obligations accrue at least annually, and must be met within 30 days following the adoption of the audit implementing report.[186] Exceptions are provided where, for instance, the VLOP considers that publication of the above-noted information may disclose confidential information, cause significant vulnerabilities for the security of its service, undermine public security or harm recipients.[187] However, the VLOP may only remove this information from the published reports; the complete (unredacted) reports must be transmitted to the Digital Services Coordinator and Commission, together with a statement of reasons justifying the removal of the information from public reports.[188]

Taken together, the auditing and transparency requirements provide for external review and oversight of VLOPs' compliance with the risk assessment and mitigation measures. This is a key feature to ensure that VLOPs fulfil their obligations to identify and root out systemic risks from their platforms – including disinformation – and that they do so in a way which is measured, critiqued, and subjected to independent oversight from auditors, the Digital Services Coordinator, the Commission, and – perhaps most importantly – the public. The need for greater oversight and transparency has long been noted,[189] and these mechanisms are a significant step in this regard. In addition, the multi-layered transparency obligations compel VLOPs to not only 'do the work', but to show how they have done so in a timely and reasoned manner.

However, several uncertainties remain about how these mechanisms will operate in practice. In the first place, while independent auditing has become a mainstay in realms previously shielded from regulatory oversight – such as the financial sector and data protection[190] – it is not immediately evident that auditing organisations with the requisite level of (proven) expertise and objectivity yet exist. Given the potential weight of the audit opinion and operational recommendations, such audit organisations may have significant impacts on fundamental rights such as freedom of expression and the right to non-discrimination. In light of this preeminent role, further clarity around the requisite qualifications of auditing organisations may be warranted.

---

[185] DSA, Article 33 § 2.
[186] Ibid.
[187] DSA, Article 33 § 3.
[188] ibid.
[189] See, generally, UN Human Rights Council, Guiding "Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework", Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie, 2011, https://www.ohchr.org/Documents/Issues/Business/A-HRC-17-31_AEV.pdf. In relation to the transparency of online platforms specifically, see The Santa Clara Principles on Transparency and Accountability in Content Moderation, 2018, https://www.santaclaraprinciples.org/.
[190] The EU imposed audit requirements on public interest entities, such as banks, by Regulation in 2014 (Regulation No 537/2014), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0537&from=EN. The General Data Protection Regulation (GDPR), adopted in 2016, provides for oversight in the form of data protection audits under Article 58 (Regulation 2016/679), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN.

In addition, while the measures outlined above provide a greater field of vision into the 'black box' of online platforms, and allow for oversight of the risk assessment cycle – from the assessment itself, through the mitigation measures employed, to the implementation of audit recommendations – the timelines provided are remarkably short. This is most striking in the time provided for VLOPs to adopt audit implementation reports setting out the "necessary measures" they are taking to implement any operational recommendations received: they have a maximum of one month to do so.[191] In light of the potential breadth and scope of the audits, and any ensuing recommendations, this may affect the quality of the measures implemented and the success of the process outcomes.

Finally, despite the reporting requirements which provide greater transparency and insight into the inner workings of VLOPs, there remain few enforcement mechanisms where VLOPs fail to fulfil their obligations. For instance, while VLOPs must provide reasons for not implementing operational recommendations, there is no penalty should VLOPs choose to implement only few or none. In this sense, the DSA walks a fine line between imposing heightened due diligence obligations on VLOPs (remedying previous failings of co- and self-regulatory approaches) while operating under the (perhaps misguided) presumption that VLOPs will undertake these new responsibilities in good faith.

4.6. Risky business? The risk-based approach in action

As the foregoing sections illustrate, disinformation has proven challenging for states and social media platforms to define, prevent, mitigate and remedy. It is into this thorny landscape that the DSA's risk-based approach has been introduced. While many aspects of the risk-based approach attempt to grapple with the shortcomings of previous approaches to disinformation, the battlegrounds of contestation have been made ever more apparent following the tabling of the proposal. These lingering questions will need to be considered and addressed to ensure a unified and comprehensive approach to combating disinformation online.

The first – and perhaps most rudimentary – question relates to whether disinformation ought to be addressed by the DSA at all. The Committee on Civil Liberties, Justice and Home Affairs of the European Parliament (LIBE Committee) thinks not: In a Draft Opinion, released in May 2021, the LIBE Committee suggested a series of amendments, including the deletion of the provisions setting out the risk-based approach.[192] The LIBE Committee argued the amendments were necessary to protect freedom of expression and to ensure the DSA addresses only the dissemination of 'illegal' rather than 'harmful' content.[193] With respect to Article 26, setting out the risk-based approach, the LIBE Committee expressed concern that its requirements "go far beyond illegal content where

[191] DSA, Article 28 § 4.
[192] Committee on Civil Liberties, Justice and Home Affairs for the Committee on the Internal Market and Consumer Protection on the proposal for a regulation of the European Parliament and of the Council Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM, 0825, 2021, Amendments 21-24, 28, 29, 91-93, https://www.europarl.europa.eu/doceo/document/LIBE-PA-692898_EN.pdf.
[193] ibid. Amendment 91, 'Justification', p. 64/84.

mere vaguely described allegedly 'negative effects' are concerned".[194] Similar concerns were raised regarding the independent audit requirements set out in Article 28.[195] The amendments put forward by the LIBE Committee suggest that further consideration of the scope and aim(s) of the DSA is necessary, particularly in relation to 'lawful but awful' speech, such as disinformation.

In the event the risk-based approach survives the ongoing negotiations and debate, a further consideration arises: namely, whether appropriate balances have been struck in the approach taken and substantive requirements put forward in the proposal. Certain civil society organisations have applauded the DSA for its "attempts to strike a careful balance by restricting content removal – which could impinge on freedom of speech – only to illegal content, whilst ensuring that the full range of impacts on our fundamental rights are dealt with through a procedure of risk assessment and mitigation".[196] Beyond the general approach, however, some organisations have voiced concern over the specifics, including:

(a) the vagueness of the "systemic risks" in Article 26 and the "reasonable" and "proportionate" thresholds set out in Article 27
(b) the limitation of risk assessments to (external) "manipulations of service" to the exclusion of risks posed by platforms' (internal) design choices
(c) the amount of discretion left to online platforms (and the European Commission) to decide how to go about mitigating systemic risks.[197]

These and other tensions warrant further consideration, reflection and refinement to ensure that the risk-based approach lives up to its promise and expectations, including overcoming some of the main stumbling blocks of previous efforts to curb disinformation online.

While a risk-based approach to identify, curb and remedy systemic risks is novel in the regulation of online speech and the combating of disinformation online, useful guidance and illustrative examples of how such due diligence mechanisms operate in practice can be found in (comparable) industries like finance and data protection.[198] These industries are similarly situated, in that public oversight has been instituted – and due diligence requirements imposed – to peer into the 'black box' through human rights impact assessments and audits. These experiences may prove useful when reflecting on the scope and contours of Articles 26 and 28, as well as any pitfalls which should be avoided.[199]

---

[194] Ibid,. pp. 64-65/84.
[195] ibid., Amendment 102, pp. 69-70/84.
[196] Avaaz Position Paper on the Digital Services Act, Disinformation and Freedom of Speech, 2021, p. 1, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12417-Digital-Services-Act-deepening-the-internal-market-and-clarifying-responsibilities-for-digital-services/F2164159_en.
[197] Ibid., pp. 9-12. See also Article19, "At a glance: Does the EU Digital Services Act protect freedom of expression" 11 February, 2021, https://www.article19.org/resources/does-the-digital-services-act-protect-freedom-of-expression/.
[198] In this regard, see Mozilla Mornings, "Unpacking the DSA's risk-based approach", 2021, https://www.youtube.com/watch?v=tEDJ3nx88MM.
[199] For instance, in the financial sector, concerns have been raised about the lack of competition and the perception of conflicts of interest. See, e.g., Prettner C., Ruby-Sachs E., Lehrich J. and Palstra N., "Don't throw out the Digital Services Act's key accountability tools", Euractiv, 2021,

Further guidance on how to design and implement due diligence mechanisms can also be derived from international human rights law, policy and practice. The UN Guiding Principles on Business and Human Rights – the leading international standards on the topic – are complemented (and reflected) in a European context by Recommendation CM/Rec(2016)3 of the Council of Europe's Committee of Ministers to member States on human rights and business. Certain key concepts have also been repurposed for specific application in the online environment in Recommendation CM/Rec(2018)2 of the Council of Europe's Committee of Ministers to member States on the roles and responsibilities of Internet intermediaries. These references underscore the importance of positional awareness in this dynamic field: in particular, the need to be clear-sighted regarding the DSA's relevance for addressing disinformation within a more complex regulatory and policy environment.

4.7. Conclusion

The regulation of online speech raises significant concerns from the perspective of fundamental rights, including freedom of expression, non-discrimination and the right to an effective remedy. It has proven particularly challenging in respect of 'lawful but awful' speech, such as disinformation, which may have broader ramifications for the public or democratic pillars.

It is into this web of complexity that the DSA proposal has been introduced. This chapter focused on the DSA's risk-based approach to combating disinformation. The proposal is a leap forward in terms of its differentiated, holistic and nuanced regulatory approach. This is evident in several respects, as outlined in this chapter. First, the DSA imposes heightened requirements on VLOPs because of their scale and reach, which reflects their capacity to cause or contribute to public harm(s). Second, the approach is comprehensive in that it is preventative and reactive, prescriptive yet flexible. VLOPs are required to identify, mitigate and disclose designated systemic risks on their platforms, but they are afforded certain flexibility in doing so. Finally, the risk-based approach is nuanced in that its provisions address not only (external) systemic risks, but also the role of (internal) business models and design choices.

A series of questions remain, including whether disinformation and other harmful but legal content should be excluded from the DSA altogether. This begs a further question about how to deal with risks caused by disinformation that do not amount to 'systemic' risks in the sense of the DSA proposal. The answers to both questions necessarily have to be shaped by the right to freedom of expression. Proponents of the approach set out in the DSA argue that it will be a significant step forward in identifying and combating risks, ensuring greater transparency and oversight, and moving beyond the 'black box' of content moderation by private companies. Whether it will live up to this promise remains to be seen.

---

https://www.euractiv.com/section/digital/opinion/dont-throw-out-the-digital-services-acts-key-accountability-tools/; and Clarke A. "Reforming Audit in the Public Interest", Luminate, 2020, https://luminategroup.com/posts/blog/reforming-audit-in-the-public-interest.