Media and Communication
2025 • Volume 13 • Article 9615
https://doi.org/10.17645/mac.9615

ARTICLE

Open Access Journal

# The Regulation of Disinformation Under the Digital Services Act

Ronan Ó Fathaigh [1], Doris Buijs [2], and Joris van Hoboken [3]

Institute for Information Law, University of Amsterdam, The Netherlands

**Correspondence:** Ronan Ó Fathaigh (r.f.fahy@uva.nl)

**Submitted:** 12 November 2024 **Accepted:** 31 March 2025 **Published:** 28 May 2025

**Issue:** This article is part of the issue "Protecting Democracy From Fake News: The EU's Role in Countering Disinformation" edited by Jorge Tuñón Navarro (Universidad Carlos III de Madrid), Luis Bouza García (Universidad Autónoma de Madrid), and Alvaro Oleart (Université Libre de Bruxelles), fully open access at https://doi.org/10.17645/mac.i476

## Abstract

This article critically examines the regulation of disinformation under the EU's Digital Services Act (DSA). It begins by analysing how the DSA applies to disinformation, discussing how the DSA facilitates the removal of illegal disinformation, and on the other hand, how it can protect users' freedom of expression against the removal of certain content classified as disinformation. The article then moves to the DSA's special risk-based rules, which apply to Very Large Online Platforms in relation to mitigation of systemic risks relating to disinformation, and are to be enforced by the European Commission. We analyse recent regulatory action by the Commission in tackling disinformation within its DSA competencies, and assess these actions from a fundamental rights perspective, focusing on freedom of expression guaranteed under the EU Charter of Fundamental Rights and the European Convention on Human Rights.

## Keywords

Digital Services Act; disinformation; European Union; online platforms; freedom of expression; regulatory enforcement

## 1. Introduction

When announcing an investigation into X in late 2023, the European Commission heralded the EU's new Digital Services Act (hereafter DSA; Regulation of the European Parliament and of the Council of 19 October 2022, 2022) as setting out an "unprecedented new standard for the accountability of online platforms regarding disinformation" (European Commission, 2023c). Indeed, when opening proceedings against Meta in April 2024, the Commission explained how it "suspects" that Meta "does not comply with DSA obligations" related to "disinformation campaigns" (European Commission, 2024a). Curiously, however,

disinformation is nowhere mentioned in the DSA's actual provisions and is nowhere defined in the DSA; it is only mentioned in some recitals (Husovec, 2024). And yet, the DSA seems to be becoming the main EU legal instrument to, as the Commission's president stated, "protect European citizens from targeted disinformation" (European Commission, 2024a). This approach is further confirmed by recent enforcement activities by the Commission targeting disinformation on platforms. As such, the purpose of this article is to critically examine the regulation of disinformation under the DSA, including the recent high-profile enforcement activity by the Commission in this regard. Additionally, we aim to highlight the tensions between the DSA's approach to disinformation and the fundamental right to freedom of expression. The article begins by analysing: how the DSA applies to disinformation, including how its provisions relating to platforms' terms and conditions apply to disinformation; the role of trusted flaggers; the operation of the 2022 Strengthened Code of Practice on Disinformation within the DSA's framework (hereafter the 2022 Code); and the role of data access rules facilitating research on disinformation. The article then moves to the DSA's special risk-based rules which apply to so-called Very Large Online Platforms (VLOPs) in relation to the mitigation of systemic risks relating to disinformation, which are to be enforced by the Commission. The article continues by discussing recent regulatory actions by the Commission in tackling disinformation within its DSA competencies. Crucially, the article assesses these actions from a fundamental rights perspective, focusing on freedom of expression guaranteed under the Charter of Fundamental Rights of the EU (2012; hereafter EU Charter) and the European Convention on Human Rights (Convention for the Protection of Human Rights and Fundamental Freedoms, 1950; hereafter ECHR).

## 2. The DSA and Online Disinformation

The DSA, which became directly applicable in EU member states in February 2024, is a landmark piece of legislation that seeks to set out harmonised rules for online platforms to ensure a "safe, predictable and trusted" online environment, and where fundamental rights are "effectively protected" (DSA, Article 1(1)). Of note, and as mentioned previously, following its adoption, the DSA is being presented as the most important EU tool against disinformation by the European Commission and its president respectively, e.g., by describing the DSA as the main EU legal instrument to "protect European citizens from targeted disinformation" (European Commission, 2024a).

However, as already pointed out, what is nevertheless very important to emphasise again, is that when actually reading the DSA, it is apparent that disinformation is not mentioned in any of its actual provisions. Crucially, disinformation is also nowhere defined anywhere in the DSA and is only mentioned in recitals (Husovec, 2024). Indeed, Recital 9 DSA states that one of the purposes of the DSA is to address the "dissemination of illegal content online" and the "societal risks" that the "dissemination of disinformation" may generate (DSA, Recital 9). And, while disinformation is not mentioned in the provisions of the DSA, as will be explained below, many of the articles in the DSA can be directly applicable to disinformation.

Before continuing to discuss specific provisions of the DSA, let us first briefly set out the DSA's general operation and application to online platforms. The DSA targets a range of what are called online "intermediary services," and has a specific set of rules for "online platforms." Crucially, Article 3 DSA defines an online platform in short as a "hosting service that, at the request of a recipient of the service, stores and disseminates information to the public" (DSA, Article 3(i)). This definition captures many social media platforms, including Facebook, Instagram, TikTok, X, and YouTube. For example, the Commission considers

Instagram an online platform, because it is "a hosting service" that "stores and disseminates information to the public at the request of recipients of its service" (European Commission, 2023a).

### *2.1. Platform Obligations in Relation to Illegal Disinformation*

A second major point is that in the run-up to the DSA proposal being published, it was considered EU policy that disinformation was a category of expression that is not illegal, but is harmful (and yet, lawful). This had crucial consequences for how the DSA sought to regulate disinformation. For example, the Commission's High Level Group on Fake News and online disinformation noted that disinformation is "not necessarily illegal," but may "nonetheless be harmful for citizens and society at large" (Directorate-General for Communications Networks, Content and Technology, 2018, p. 5). The Commission itself also describes disinformation as "harmful content," which is "not, *per se*, illegal" (European Commission, 2020a, p. 3). In the intervening period, there has been research on disinformation laws in the EU, and a growing realisation that the notion of disinformation is in fact illegal in many EU member states (European Regulators Group for Audiovisual Media Services, 2020). Indeed, research points to legislation in numerous EU member states which may capture the notion of disinformation. And most worryingly, it is criminalised in many of those EU member states. As Ó Fathaigh et al. (2021) note, for example, in Malta, the Criminal Code (Article 82) criminalises the spreading of "false news," and makes it an offence to "maliciously spread false news which is likely to alarm public opinion or disturb public good order or the public peace or to create a commotion among the public or among certain classes of the public" (Criminal Code of the Republic of Malta, 1854). While, under the Criminal Code of Cyprus (Article 50), it is an offence to disseminate "false news" or "news that can potentially harm civil order or the public's trust towards the State or its authorities or cause fear or worry among the public or harm in any way the civil peace and order" (The Criminal Code Law of Cyprus, 2025, Article 50). Importantly, these laws are not anachronistic and rarely invoked but are being actively enforced across member states (Espaliú-Berdud, 2022; Koltay, 2025; Radu, 2023). Indeed, as Ó Fathaigh et al. (2021) note, the European Commission itself has raised its alarm over member state laws of a "criminal nature" related to disinformation, and has warned that such laws that are "too broad" and "with disproportionate penalties" raise "particular concerns as regards freedom of expression" (European Commission, 2020b, p. 11).

In this regard, it is important to mention that the definition of "illegal content" is given an incredibly broad definition under the DSA (Ó Fathaigh et al., 2021) as it includes "any information that, in itself or in relation to an activity…is not in compliance with Union law or the law of…any Member State," and "irrespective of the precise subject matter or nature of that law" (DSA, Article 3(h)). As such, this definition of illegal content captures all of the national criminal legislation applicable to disinformation. This would mean that platforms' obligations in relation to illegal content under the DSA would apply to disinformation that has been made illegal in some EU member states. Further, these laws define disinformation differently, and this complicating factor also makes it more difficult for platforms to conform to such diverse disinformation laws across the EU (Ó Fathaigh et al., 2021, p. 15).

It is therefore important to closely look into three main provisions of the DSA which apply to illegal content, as they may cover disinformation in some of the EU's member states. The first of these is Article 9 DSA, where national judicial or administrative authorities may order online platforms to "act against" content considered "illegal content"; while online platforms must inform the national authorities "without undue

delay" of any effect given to the order (DSA, Article 9(1)). Very importantly, it is not only courts that can order content to be taken down, but this idea of "national administrative authorities" can also include "law enforcement authorities" (DSA, Recital 31). In other words, as noted by Ó Fathaigh et al. (2021), Article 9 creates an "explicit EU law mechanism to facilitate national judicial and administrative authorities" to issue orders for online platforms to "act against" specific user content that is deemed "illegal content" (Ó Fathaigh et al., 2021, p. 17). Notably, in recent transparency reports being published by platforms under the DSA, platforms such as Meta are reporting how Article 9 DSA orders are being made against its platforms under national laws applicable to misinformation (see, for example, Meta, 2024a, 2024b).

The second article to mention is Article 16 DSA, which requires platforms to implement notice-and-action mechanisms for (allegedly) illegal content. In particular, platforms are required to "put mechanisms in place to allow any individual or entity to notify them of the presence on their service of specific items of information that the individual or entity considers to be illegal content" (DSA, Article 16(1)). Platforms must process and make a decision on these notices in a "timely, diligent, non-arbitrary and objective manner," and notify their decision "without undue delay" (DSA, Article 16(5–6)). Again, due to the very broad definition of illegal content, this notice-and-action mechanism will also be applicable to all national criminal legislation on disinformation (Ó Fathaigh et al., 2021, p. 18). As such, Article 16 obliges platforms to put notice-and-action mechanisms in place for notices to be submitted of (allegedly) illegal content considered disinformation, with platforms being required to make a decision on this content without undue delay. However, it should be noted that currently, platforms may not be fully implementing these mechanisms in line with Article 16 (Holznagel, 2024a).

Further, Article 16(3) DSA provides that properly-submitted notices of (alleged) illegal content "shall be considered to give rise to actual knowledge or awareness" for the purposes of Article 6 DSA, which protects platforms from liability (DSA, Article 16(3)). In this regard, Article 6 DSA provides that platforms "shall not be liable" for any user content, even if it is illegal, provided the platform (a) does not have "actual knowledge" of the illegal content, or (b) "upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the illegal content" (DSA, Article 6(1)). What Article 16 DSA now means is that properly submitted notices of illegal content "shall be considered to give rise to actual knowledge" on the part of platforms for the purposes of Article 6 DSA. Thus, platforms are being put in a position to decide whether flagged content should be deemed illegal content under national law provisions applicable to disinformation. Platforms are also required to make this decision with a constant threat hanging over them that the submitted notice will mean they have actual knowledge of the illegal content, making them potentially liable for the content unless they "[act] expeditiously to remove or to disable access to the illegal content" (DSA, Article 6(1)). This may arguably incentivise removal.

A third important article in relation to illegal disinformation is Article 22 DSA on trusted flaggers, a flagging mechanism to inform platforms about illegal content—a practice which already existed before its inclusion in the DSA (Appelman & Leerssen, 2022). Article 22(1) requires platforms to "take the necessary technical and organisational measures" to ensure notices submitted by "trusted flaggers" through notice and action mechanisms under Article 16 DSA, are given "priority" and are processed and decided upon "without undue delay." Notably, the status of trusted flagger is to be awarded to entities by the newly-established national Digital Services Coordinators (DSCs), which are the national regulatory authorities established to enforce the DSA at a national level (DSA, Article 49). Crucially, Recital 61 DSA gives examples of such trusted flaggers, which can be "public" bodies, and "internet referral units of national law enforcement authorities"

or the "European Union Agency for Law Enforcement Cooperation" (Europol). Again, because of the broad definition of illegal content under the DSA, Article 22, too, will be applicable to national criminal legislation concerning disinformation and will facilitate internet referral units of national law enforcement authorities submitting notices of alleged disinformation, where such notices are to be decided upon with priority and without delay. It should be noted that trusted flaggers are quite important in relation to disinformation, with many public authorities having trusted flagger status before the DSA was enacted e.g., the Dutch Ministry of the Interior and Kingdom Relations has been a trusted flagger with numerous platforms (Ministry of the Interior and Kingdom Relations, 2023; see also van de Kerkhof, 2024). Thus far, trusted flagger status has been awarded to some entities with areas of expertise such as "negative effects on civic discourse and elections," including a Greece-based organisation which aims to systematically address disinformation (European Commission, 2025c). Scholars have discussed potential reasons for the lack of applications to be awarded the status of trusted flagger, such as resource constraints (Goldberger, 2024), but also in a broader sense of what the potential impact of the trusted flagger provision may be, which may not be "groundbreaking" (Rosati, 2024).

These are the main provisions that can be utilised to have disinformation removed where it comes within the definition of illegal content under the DSA and demonstrate how the DSA can be instrumentalised for the removal of illegal disinformation on platforms. However, an important point to make about the regulation of disinformation under the DSA is that not only does the DSA facilitate the removal of illegal disinformation, it also seeks to protect individuals whose content has been removed because it is considered disinformation by platforms. In this regard, the DSA has a double-edged-sword approach to the regulation of disinformation: where on the one hand it facilitates the removal of disinformation, on the other hand, it seeks to protect users' freedom of expression when content is removed for being qualified as disinformation by platforms. To this point, we now turn.

### *2.2. Disinformation Regulation and How the DSA Can Protect Freedom of Expression*

The previous section examined how platforms can be instrumentalised to remove illegal disinformation. This section in turn details how the DSA also imposes obligations on platforms to protect users' freedom of expression where their content has been removed because it is allegedly disinformation. Platforms may remove "disinformation" because it may be illegal, but also if such content violates their terms and conditions. One of the most important DSA provisions in this regard is Article 14, which regulates platforms' terms and conditions. Preceding the enactment of the DSA, it was widely recognised that the systems used by platforms to moderate expression based on a platform's terms of service were "fundamentally broken" (Culliford & Paul, 2020; York & McSherry, 2019), and "undermine" freedom of expression online (Amnesty International, 2019). This was because of "overly vague rules of operation, inconsistent enforcement, and an overdependence on automation" (UN General Assembly, 2018).

As such, one of the purposes of Article 14 DSA was to, for the first time, impose statutory regulation on how platforms enforced their terms and conditions. The most important aspect of Article 14 is how it provides that platforms, when applying and enforcing restrictions on user content based on their terms and conditions, must have "due regard" to the "fundamental rights" of users "as enshrined" in the EU Charter (DSA, Article 14(4)). As authors such as Quintais et al. (2023), have explored, Article 14 essentially means that platforms should apply their terms and conditions with "due regard" to fundamental rights, explicitly including "freedom of

expression" under Article 11 EU Charter. This means platforms need to have due regard to fundamental rights in content moderation decisions based on a platform's rules in relation to disinformation, although it may be unclear what the practical effect is of this requirement (Galantino, 2023, p. 124).

As such, Article 14(4) DSA shifts the focus to some extent towards the fundamental rights framework concerning freedom of expression. The wording of the right to freedom of expression under Article 11 EU Charter is "broad and open-ended, and gives little concrete guidance to platforms" (Quintais et al., 2023, p. 897). As such, platforms may have regard to principles from the EU Court of Justice case law under Article 11 EU Charter, and from the European Court of Human Rights case law on freedom of expression, guaranteed under Article 10 ECHR. As Quintais et al. (2023) point out, the EU Court of Justice has confirmed that Article 11 EU Charter has the "same meaning and the same scope" as Article 10 ECHR, "as interpreted by the case-law of the European Court of Human Rights" (*Sergejs Buivids v. Datu valsts inspekcija* (2019), para. 65).

Further, Recital 47 DSA explicitly states that platforms should also have "due regard" to "relevant international standards for the protection of human rights" when applying restrictions based on their terms and conditions. Crucially, there are important and relevant freedom of expression principles that may be applicable to disinformation under both European and international human rights law. While other authors have examined in depth the application of freedom of expression principles to disinformation regulation (McGonagle, 2017; van Hoboken & Ó Fathaigh, 2021), for the purposes of this article, it is relevant to mention some of these principles that platforms may have due regard to under Article 14 DSA specifically. Ó Fathaigh et al. (2021) point towards certain specific principles. First, under international human rights law, regulations prohibiting dissemination of disinformation or "false news," which are "vague and ambiguous," are "incompatible" with human rights standards on freedom of expression "should be abolished" (UN, Organization for Security and Co-operation in Europe, Organization of American States, & African Commission on Human and People's Rights, 2017, p. 3). In particular, it has been emphasised by international human rights bodies that the concept of disinformation is an "extraordinarily elusive concept to define," and may provide executive authorities with "excessive discretion to determine what is disinformation, what is a mistake, what is truth" (UN General Assembly, 2020). As such, the penalisation of disinformation may be "disproportionate" under international human rights law. Of particular note, the European Court of Human Rights has held, in a landmark judgment, that legal proceedings over the "dissemination of false information" under national election legislation were a violation of the right to freedom of expression under Article 10 ECHR (*Salov v. Ukraine*, 2005). Crucially, the Court held as a matter of principle that Article 10 ECHR "as such does not prohibit discussion or dissemination of information received even if it is strongly suspected that this information might not be truthful" (*Salov v. Ukraine*, 2005, para. 113).

Thus, it is quite clear that under European and international human rights law, prohibiting disinformation raises fundamental questions under freedom of expression standards. Should platforms wish to indeed prohibit disinformation under their terms and conditions, Article 14 DSA may place those platforms under an obligation to take into account these fundamental rights aspects by assessing whether their own definitions of disinformation are sufficiently clear, and that certain restrictions placed on content classified as disinformation would be proportionate, and be the least restrictive measure (e.g., labelled or fact-checked, rather than removed). Further, the application of Article 14 DSA may materialise not only in the initial

decision taken by a platform applying its terms and conditions but also as users invoke the DSA's provisions to challenge decisions over content restricted under a platform's disinformation rules. So, Article 14 is the first of these DSA articles that addresses the issue of procedural fairness for users in the context of moderation of disinformation (Ó Fathaigh et al., 2021, p. 19), in contrast to the other articles discussed in the previous section which fit the perspective of the DSA as an instrument to remove disinformation.

A second article in this stream seeking to protect a user's freedom of expression in relation to disinformation is Article 17 DSA. Based on this provision, platforms shall give a statement of reasons following restrictions imposed on a user's content, due to it being considered illegal disinformation, or disinformation removed under a platform's terms and conditions. The types of considered restrictions include restrictions on the "visibility" of information, including content removal, disallowing access, or content demotion, which are all used by platforms to target disinformation (Leerssen, 2023). Crucially, the statement shall be "as precise and specific as reasonably possible" (DSA, Article 17(4)), and include "explanations as to why the information" is considered to be illegal content or incompatible with terms and conditions (DSA, Article 17(3.d–e)). Thus, under this provision platforms are required to actually explain their moderation decisions around disinformation. However, early analysis of the implementation of Article 17 would seem to suggest statements of reasons are generated automatedly and quite short, resulting in users not being able to specifically understand how certain content may have violated a platform's rules (Kaushal et al., 2024).

In addition to Article 17 DSA, the next question is what sort of redress can a user avail of if a user disagrees with the statement of reasons. In this regard, Article 20 DSA is crucial, which provides that platforms must provide users with access to "effective" internal complaint-handling systems to lodge complaints. Notably, platforms shall handle complaints in a "timely" and "diligent" manner (DSA, Article 20(4)) "under the supervision of appropriately qualified staff" (DSA, Article 20(6)), and be "free of charge" (DSA, Article 20(1)). This is beneficial from a user's procedural rights point of view; and again, it means that a platform's complaint-handling system may be required to issue decisions on disinformation having "due regard" to a user's freedom of expression under Article 14 DSA.

Further, where a user disagrees with a platform's internal complaint system's decision, users shall also "be entitled to select any out-of-court dispute settlement body" that has been certified "to resolve disputes" (DSA, Article 21(1)) relating to decisions made by platforms such as removal, disabling access, and suspension or termination of a user's account (DSA, Article 20(1.a–d)). This includes disputes that "could not be resolved in a satisfactory manner through the internal complaint-handling systems" (DSA, Recital 59) under Article 20. As such, these bodies may consider how a platform's decision related to disinformation sufficiently takes account of a user's right to freedom of expression. Platforms are required to "engage, in good faith" with the certified out-of-court dispute settlement body "with a view to resolving the dispute" (DSA, Article 21(2)). However, the same paragraph notes that this body does not have the power to "impose a binding settlement" on the platform nor the involved user. In terms of setting up these bodies, the national DSCs will certify these bodies, provided they satisfy a number of requirements, such as being "impartial and independent" and having the "necessary expertise" (DSA, Article 21(3)). Holznagel (2024b) has explored the first out-of-court dispute settlement bodies under the DSA and finds these first bodies "'serious,'" but also points towards potential questionable consequences around the financial side of this new type of alternative dispute resolution.

These are the main provisions that are applicable to disinformation on platforms from an user's procedural-rights perspective under the DSA, and demonstrate the second angle upon which the DSA regulates disinformation.

## 3. VLOPs and Disinformation

The rules discussed in the previous section generally apply to online platforms. However, a particularly landmark and important aspect of the DSA is that it also has very specific additional rules for what are called VLOPs. These rules adopt a risk-based approach to platform regulation (Nooren et al., 2018) and basically impose on VLOPs an obligation to manage any "systemic risks" stemming from their platforms. It is in this area of the DSA that the regulation of disinformation is most pronounced (Husovec, 2024).

Under Article 33 DSA, VLOPs are platforms which have a "number of average monthly active recipients of the service in the Union equal to or higher than 45 million." Recital 75 explains why VLOPs are targeted: given the "importance" of VLOPs in "facilitating public debate" and the "dissemination to the public of information, opinions and ideas," it is "necessary to impose specific obligations" on them. In April 2023, the European Commission designated the first tranche of 17 VLOPs under the DSA, which included all the major social media platforms, such as Facebook, Instagram, LinkedIn, Snapchat, TikTok, Twitter (now X), and YouTube (European Commission, 2023b).

This idea of targeting VLOPs through risk-based regulation is an innovative way of regulating platforms and imposing specific rules on them (which is part of a broader regulatory trend of adopting a risk-based approach to online platforms; Efroni, 2021). Once these platforms are designated as VLOPs, they are subject to inter alia what are called systemic-risk obligations. Thus, under Article 34 DSA, VLOPs are required to carry out "risk assessments" to identify and assess "any systemic risks in the Union stemming from the design or functioning of their service and its related systems, including algorithmic systems." This is where disinformation becomes crucial, as one of the explicit purposes of the DSA is to address the "societal risks that the dissemination of disinformation" may generate (DSA, Recital 9).

Article 34 lists a number of systemic risks that must be included in the risk assessment by VLOPs: (a) dissemination of illegal content through their services; (b) any actual or foreseeable negative effects for the "exercise of fundamental rights" enshrined in the EU Charter, including freedom of expression; (c) any actual or foreseeable negative effects on "civic discourse and electoral processes," and "public security"; and (d) any actual or foreseeable negative effects in relation to "gender-based violence, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being" (DSA, Article 34(1.a–d)).

Crucially, disinformation is not explicitly mentioned as one of the four potential systemic risks. The question is therefore: How does disinformation fit within this provision? In this regard, it is important to look at the recitals of the DSA concerning systematic risks. Notably, Recital 83 states that risks of actual or foreseeable negative effects on the protection of "public health," "minors," "serious negative consequences to a person's physical and mental well-being," or "gender-based violence" can stem from "coordinated disinformation campaigns related to public health." Based on this recital, it seems disinformation is most readily identified with risk (d) under Article 34, namely negative effects on the protection of, in short, "public health." Further, Recital 84 states that

when assessing the systemic risks, VLOPs should focus on information "which is not illegal," and pay "particular attention" to how their services are used to disseminate or amplify "misleading or deceptive content, including disinformation," and how amplification of such information "contributes" to the systemic risks. Notably, the recitals concerning Article 34 and Article 35 do not state that disinformation is a systemic risk as such; only that systemic risks can *stem* from disinformation. Thus, on a strict reading of Article 34, the dissemination of disinformation may not be a systemic risk, in and of itself. For instance, Meta lists disinformation in multiple categories of "systemic risk areas," including "civic discourse & elections," "public health," and "public security" (Meta, 2024c, p. 17). LinkedIn includes mis- and/or disinformation in risk areas as "civic discourse and electoral processes," as well as under "public health" and "public security" (LinkedIn, 2024, pp. 63–64); and X, too, mentions disinformation in the context of "democratic processes, civic discourse and electoral processes," "public security," and "public health & physical and mental well-being" (X, 2023, pp. 59–68). However, as will be seen in Section 4.1, the systemic risk currently most associated with disinformation is risk (c) concerning negative effects on "civic discourse and electoral processes." Indeed, as Mündges and Park note, although Article 34 "does not explicitly mention disinformation, it implies comprehensive coverage of the phenomenon" through Article 34(1.c) DSA (Mündges & Park, 2024, p. 5). This is also where most regulatory action is currently occurring, which is discussed in Section 4.1. This is not to say that risk (c) on civic discourse and electoral processes will remain the systemic risk most closely linked to disinformation; this may change over time, and it could even be that the list of systemic risks in the DSA will be reevaluated and adapted at some point.

Once platforms have conducted their risk assessments and identified the risks, the next step is mitigating these risks. Crucially, if VLOPs identify these risks, under Article 35 DSA, VLOPs must "put in place reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified pursuant to Article 34" (DSA, Article 35(1)). Article 35 then sets out 11 measures which VLOPs can take. These are in short (a) "adapting the design, features or functioning of their services, including their online interfaces"; (b) "adapting their terms and conditions and their enforcement"; (c) "adapting content moderation processes," as well as "adapting any relevant decision-making processes and dedicated resources for content moderation"; (d) adapting "algorithmic systems," including "recommender systems"; (e) "adapting advertising systems"; (f) "reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risk"; (g) "initiating or adjusting cooperation with trusted flaggers"; (h) "initiating or adjusting cooperation with other providers of online platforms" through "codes of conduct and the crisis protocols" as referred to in Articles 45 and 48 DSA respectively; (i) "taking awareness-raising measures" in order to give users "more information"; (j) "taking targeted measures to protect the rights of the child," and (k) ensuring that items of information that "appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful is distinguishable through prominent markings when presented on their online interfaces" (DSA, Article 35(1)).

Notably, Article 34 and Article 35 are somewhat vague, e.g., "what makes a risk 'systemic'" (Sullivan & Pielemeier, 2023). As Griffin notes and what will be further discussed in Section 4.2, "The breadth and vagueness of Articles 34–35 gives the Commission significant discretion over their interpretation and enforcement" (Griffin, 2024, p. 176). The question is: What do Articles 34 and 35 mean in practice, especially in relation to disinformation? Helpfully, the Commission may, in cooperation with the DSCs, issue guidelines on the application of Article 35 in relation to specific risks (DSA, Article 35(3)). Indeed, just before the European Parliament elections in June 2024, the Commission adopted guidelines on the mitigation of systemic risks for electoral processes, to which we now turn.

### 3.1. Guidelines on the Mitigation of Systemic Risks for Electoral Processes

The guidelines on the mitigation of systemic risks for electoral processes were designed to provide "guidance" to support VLOPs to "comply with their obligation to mitigate specific risks linked to electoral processes" (Communication from the Commission, 2024, point 3). Notably, the guidelines contain specific measures for VLOPs to implement targeting disinformation, including the following: First, in order to "prevent the spread of" disinformation "on the electoral process itself," the best practice for VLOPs is to "facilitate access to official information concerning the electoral process," based on official information from the electoral authorities (Communication from the Commission, 2024, point 27(a)). A second example of best practices is for VLOPs to apply "inoculation measures that pre-emptively build resilience against possible and expected disinformation narratives" by "informing and preparing users," for example through the use of online games, videos, and other content on the generation of disinformation, which "encourages a critical reflection on the tactics" used for disinformation (Communication from the Commission, 2024, point 27(b.ii)). Third, VLOPs should use "fact-checking labels on identified disinformation" provided by "independent fact-checkers and fact-checking teams of independent media organisations" to "provide users with more contextual information" (Communication from the Commission, 2024, point 27(c–i)). The foregoing measures are very much in the vein of providing users with more information to recognise and be resilient against disinformation. However, there are particular measures which go much further and are framed in the sense of reducing the spread of disinformation. These include that VLOPs should consider to "reduce the prominence of disinformation in the context of elections," including "deceptive content that has been fact-checked as false," or "from accounts that have been repeatedly found to spread disinformation" (Communication from the Commission, 2024, point 27(d–ii)). Further, the Commission recommends VLOPs to put "systems in place to prevent the misuse of advertising systems" to disseminate inter alia "disinformation" (in the context of political advertising, point 27(e–iv)); to engage in the "demonetisation of disinformation content" (point 27(g)); to ensure the "enforcement" of terms and conditions to "significantly decrease the reach and impact of generative AI content" depicting "disinformation on the electoral process" (Communication from the Commission, 2024, point 40(a)). Thus, these measures are a mix of providing users with more information to be resilient against disinformation; and measures to reduce the reach and monetisation of disinformation.

Notably, the Commission was quite forthright in terms of VLOPs being required to implement these measures, although the guidelines contain "best practices" and recommendations (DSA, Article 35(3)). The Commission explicitly stated that VLOPs which "do not follow" these guidelines "must prove" to the Commission that the measures undertaken are "equally effective in mitigating the risks" (European Commission, 2024b). And "should the Commission receive information casting doubt on the suitability of such measures, it can request further information or start formal proceedings under the Digital Services Act" (European Commission, 2024b). So, this was very much a warning to VLOPs to follow these guidelines. Finally, and crucially, the guidelines also explicitly state that the mitigation measures "should draw" on the 2022 Code (European Commission, 2022), and it is to this code we now turn.

### 3.2. 2022 Code

When examining the DSA's regulation of disinformation, it is crucial to examine the 2022 Code as it is inextricably linked to the DSA (Brogi & De Gregorio, 2024). The Code was first put together in 2018 by the Commission and a number of online platforms starting as a self-regulatory instrument to which platforms

could voluntarily adhere (European Commission, 2018). The Code was updated in 2022, to become a mammoth 50-page document of co-regulation. This strengthened version of the Code explicitly stated already that it "aims to become" a recognised code of conduct under the DSA (2022 Code, p. 2). Indeed, the 2022 Code has been converted into an official code of conduct under the DSA, which will take effect from 1 July 2025 (European Commission, 2025b). Further, the DSA's recitals explicitly mention the 2022 Code (DSA, Recital 106), and that compliance with a given code of conduct by a VLOP can be considered as an "appropriate risk mitigating measure" in relation to systemic risks (DSA, Recital 104). This is now officially the case, as "full adherence" may indeed be considered an "appropriate risk mitigation measure" by the Commission and the 2022 Code will become a "significant and meaningful benchmark for determining DSA compliance" (European Commission, 2025b). The "refusal without proper explanations" by a VLOP to participate in the application of a code of conduct could be "taken into account" when determining whether a platform "has infringed the obligations" laid down by the DSA (DSA, Recital 104). Article 35(1.h) explicitly mentions cooperation through codes of conduct as a risk-mitigation measure. This shows how the Code has shifted from a self-regulatory instrument towards a crucial part of assessing VLOPs' compliance with the DSA in the context of disinformation.

The 2022 Code contains specific measures designed to target disinformation (the 2022 Code was renamed the Code of Conduct on Disinformation in February 2025, following its integration into the DSA framework; European Commission, 2025a). Notably, these measures include that platforms should: (a) put in place a functionality to allow users to flag "harmful false and/or misleading information," which should lead to "appropriate, proportionate and consistent follow-up actions" (Measure 23.1); (b) provide functionalities to allow users to assess the "authenticity or accuracy" of content (Commitment 20); (c) provides users with "factual accuracy of sources through fact-checks from fact-checking organisations that have flagged potential Disinformation," and "warning labels" from "authoritative sources" (Commitment 21); and (d) commit to "defund" the dissemination of disinformation (Commitment 1). A very important aspect of the 2022 Code is Measure 18.2, where platforms commit to "enforce" policies to "limit the spread" of false information, which can include prohibiting false information.

Crucially, although verification of reporting from platforms under the 2022 Code is difficult as Mündges and Park (2024, p. 1) note, "qualitative information provided by platforms often lack detail and/or relevance" and "quantitative data is, in several cases, missing, incomplete, or not robust"), these reports (at face value) reveal content considered disinformation is actually being taken down pursuant to Measure 18.2. So, when looking at reporting by TikTok in its latest report covering the period of the European Parliament elections, TikTok removed over a quarter of a million videos in the EU before the European Parliament elections (TikTok, 2024, p. 165). Similarly, YouTube reported removing almost 19,000 videos in the run-up to the EU elections based on its disinformation policy (Google, 2024, p. 146), while LinkedIn reported removing over 20,000 pieces of content under its disinformation policy in the EU in its March 2024 report (Microsoft, 2024, p. 119). This seems a considerable amount of content that is being removed under the 2022 Code, and only during a six-month period. Galantino also explicitly mentions how the 2022 Code "fails to tightly control the scope of content removals and account sanctions" and adds that "although content removal is not explicitly envisioned by the Code, it occurs in practice" (Galantino, 2023, p. 126).

As such, while the DSA does not prohibit disinformation, the 2022 Code is where disinformation can be prohibited, and where a large amount of removal is occurring. This is a crucial issue to highlight, as the 2022

Code is now seen as a co-regulatory measure under the DSA and one of the potential risk-mitigating measures under the DSA. Of note, Mündges and Park have analysed compliance by platforms with their reporting obligations under the 2022 Code, and note that "overall, platforms are only partly compliant with the Code" (Mündges & Park, 2024, p. 1).

### *3.3. Access to Data and Disinformation*

A further crucial DSA provision relevant to disinformation, which is only applicable to VLOPs, is Article 40 on access to data. As Khan (2021) has noted, the lack of access to data continues to be a "major failing" of platforms regarding disinformation and makes independent scrutiny and accountability difficult. And a lack of access to data makes it "impossible" to actually assess the prevalence of disinformation on platforms, and assess the "effectiveness" of measures adopted by platforms to address disinformation (Khan, 2021, p. 17). Notably, Article 40 DSA is a ground-breaking provision on data access. There are two main angles. First, under Article 40(1), VLOPs are required to provide the European Commission, at its "reasoned request," access to data that are "necessary to monitor and assess compliance with this Regulation." This access to data is linked to the systemic risks provisions and may include data "necessary to assess the risks and possible harms" brought about by VLOPs' systems (DSA, Recital 96).

Second, not only does Article 40 allow access to data by the European Commission, but it also crucially allows in certain circumstances for researchers to access platform data. Thus, under Article 40(4), VLOPs "shall," in principle, provide access to data to "vetted researchers," upon a "reasoned request" from the national DSC of establishment, for the "sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks" under Article 34 DSA, and to the "assessment of the adequacy, efficiency and impacts of the risk mitigation measures" under Article 35 DSA. Notably, under Article 40(8), it is the national DSC that will grant the status of "vetted researchers," a contested notion as noted by Leerssen (2021). The provision lays down a number of criteria, including that a researcher is "independent from commercial interests," "capable of fulfilling the specific data security and confidentiality requirements," and "their application demonstrates that their access to the data and the time frames requested are necessary for, and proportionate to, the purposes of their research" (DSA, Article 40(8)). Thus, vetted researchers will be able to access data from VLOPs for the sole purpose of conducting research into the systemic risks and mitigation measures under Articles 34 and 35. Given how disinformation is considered a central contribution to systemic risks under Article 34, Article 40 DSA may allow independent scrutiny of the prevalence of, and engagement with, disinformation on VLOPs. Of note, the European Commission (2024c) published a draft delegated act on data access under Article 40 DSA in late October 2024 (Albert, 2024; Vermeulen, 2024).

## 4. Regulatory Action Under the DSA Against Disinformation and the Impact of Freedom of Expression

The foregoing section sought to set out how the DSA's specific rules for VLOPs apply to disinformation, in particular the systemic risk provisions. A major final question on the DSA and regulating disinformation on VLOPs is how it will be enforced, and to this we now turn. Notably, and as mentioned before, it is not the national DSCs (Jaursch, 2023) that have the power to regulate VLOPs, but the European Commission that has "exclusive powers" to supervise and enforce Articles 34 and 35 on systemic risks (DSA, Article 56).

In this regard, the Commission is granted extensive powers under the DSA, including the power to submit a request for information to demand information from VLOPs relating to a "suspected infringement," where VLOPs can be fined for providing "incorrect, incomplete or misleading information" (DSA, Article 67(1)–(2)). Further, the Commission can impose "interim measures" targeting VLOPs "where there is an urgency due to the risk of serious damage for the recipients of the service" (DSA, Article 70(1)). The Commission has stated that this can include measures such as "increased monitoring of specific keywords or hashtags" by VLOPs (European Commission, 2025d). Ultimately, the Commission can issue non-compliance decisions against VLOPs for violating the DSA, and issue fines up to 6% of a VLOP's total worldwide annual turnover (DSA, Articles 73–74). We will now first set out the enforcement actions undertaken by the Commission thus far, after which we will make an assessment of how the DSA is being utilised to regulate disinformation, including responses from civil society organisations (CSOs), in light of freedom of expression.

### *4.1. Regulatory Action Under the DSA Against Disinformation*

The Commission has been undertaking regulatory action under the DSA, specifically targeting disinformation. The first regulatory action taken by the Commission under the DSA concerning disinformation occurred in October 2023, following the Hamas attacks on Israel. The Commission sent high-profile public correspondence to four VLOPs, namely TikTok, Meta, X, and YouTube, over disinformation related to the Hamas–Israel conflict. The Commission stated it had "indications" these platforms were being used to disseminate inter alia "disinformation in the EU" and sought to set out the "very precise obligations" under the DSA (Breton, 2023). The Commission noted that VLOPs must "diligently" enforce their terms and conditions, and must put in place "proportionate and effective mitigation measures to tackle the risks to public security and civic discourse stemming from disinformation" (Breton, 2023). The Commission noted there were many reports of fake and manipulated images and facts circulating on these platform(s) in the EU, "such as repurposed old images of unrelated armed conflicts or military footage," which the Commission considered appearing as "manifestly false or misleading information" (Breton, 2023). Certain VLOPs were invited to "urgently ensure" their systems were "effective," and report on measures taken to the Commission. Additionally, VLOPs were invited to "be in contact with relevant law enforcement and Europol" and "ensure that [they] respond promptly to their requests" (Breton, 2023). Of particular note, the Commission requested a response within 24 hours and warned that responses would be included in an "assessment file" on "compliance with the DSA," and that "following the opening of a potential investigation and a finding of non-compliance, penalties can be imposed" (Breton, 2023).

Following this, the Commission sent requests for information under Article 67 DSA to TikTok, Meta, and X over disinformation. Notably, the Commission stated it had "indications" of the "alleged spreading" of inter alia "disinformation," and noted X as a VLOP had an obligation under the DSA to mitigate risks related to the dissemination of disinformation (European Commission, 2023c, p. 1). Again, concerning its formal request to Meta, the Commission stated it required information specifically on measures taken to comply with obligations with regard to "mitigation measures" with regard to dissemination and amplification of disinformation (European Commission, 2023d). The Commission noted for both VLOPs that it could "impose fines for incorrect, incomplete or misleading information in response to a request for information" and "failure to reply by the deadline could lead to the imposition of periodic penalty payments" (European Commission, 2023c, p. 2, 2023d). And following the request for information, the Commission in December 2023 opened formal proceedings against X. This revolved inter alia around "the effectiveness of measures

taken to combat information manipulation on the platform," notably the effectiveness of related policies "mitigating risks to civic discourse and electoral processes" (European Commission, 2023e, p. 1).

Following the requests for information towards various VLOPs and opening of formal proceedings against X over disinformation, the Commission again engaged in further high-profile regulatory activity with X in August 2024 over disinformation. The correspondence was sent in relation to riots in the UK and a live-streamed interview on X between Elon Musk and US then-presidential candidate Donald Trump. The Commission stated that as a VLOP under the DSA, X was required to ensure:

> Proportionate and effective mitigation measures are put in place regarding the amplification of harmful content in connection with relevant events, including live streaming, which, if unaddressed, might increase the risk profile of X and generate detrimental effects on civic discourse and public security. (Breton, 2024, p. 1)

The Commission also mentioned examples of "public unrest brought about by the amplification" of content including "certain instances of disinformation" (Breton, 2024). Notably, the Commission also stated that:

> We are monitoring the potential risks in the EU associated with the dissemination of content that may incite violence, hate and racism in conjunction with major political—or societal—events around the world, including debates and interviews in the context of elections. (Breton, 2024, p. 1)

Finally, the Commission would be "extremely vigilant to any evidence that points to breaches of the DSA and will not hesitate to make full use of [its] toolbox, including by adopting interim measures, should it be warranted to protect EU citizens from serious harm" (Breton, 2024, p. 1).

Now that we have an overview of the enforcement activities from the Commission revolving around disinformation, we will analyse this from a critical, fundamental rights perspective focusing on freedom of expression, including responses by CSOs.

### *4.2. DSA Regulatory Measures in Light of Freedom of Expression*

The Commission's regulatory activity as described in Section 4.1 has been quite controversial with considerable criticism from CSOs invoking freedom of expression principles. Indeed, 28 CSOs signed an open letter criticising the Commission's correspondence over disinformation related to the Hamas-Israel conflict. The organisations criticised the Commission over its "false equivalence between the DSA's treatment of illegal content and 'disinformation'", "the focus on the swift removal of content," and that "the DSA does not impose an obligation on service providers to 'consistently and diligently enforce [their] own policies'" (Access Now, ARTICLE 19, AlgorithmWatch et al., 2023, p. 2). Indeed, the organisations noted that "State pressure to remove content swiftly based on platforms' terms and conditions leads to more preventive over-blocking of entirely legal content" (Access Now, ARTICLE 19, AlgorithmWatch et al., 2023, p. 2). The criticism voiced by these organisations is reflected in the text of the DSA, which does not contain explicit deadlines on when content must be removed.

Further, following the Commission's regulatory correspondence over the UK riots and Trump's interview on X, civil society accused the Commission of using the DSA as a "pressure tool against online platforms during politically sensitive times and periods of high media attention" (Access Now, ARTICLE 19, & Electronic Frontier Foundation, 2024). CSOs also pointed towards the fact that both the interview and the riots took place outside the EU, and that while such events "may certainly lead to serious negative consequences" within the EU, they were concerned that the Commission's correspondence "does neither specify whether or how" the UK events "have reached the threshold of systemic risks within the EU nor explain why" the interview broadcast required "'effective mitigation measures' in the EU" (Access Now, ARTICLE 19, & Electronic Frontier Foundation, 2024). It was furthermore "entirely unclear what ex-ante measures a VLOP should take to address a future speech event" such as this single interview "without resorting to general monitoring and disproportionate content restrictions" (Access Now, ARTICLE 19, & Electronic Frontier Foundation, 2024). Finally, the CSOs called on the Commission to abstain from "generally demanding content-specific restrictions in the context of the systemic risk assessment and mitigation provisions," and "strongly" recommended that the Commission provide "more clarity" on its understanding of systemic risks under the DSA, including the "granularity of required evidence" VLOPs must follow when assessing if their systems and processes pose risks to "public discourse" (Access Now, ARTICLE 19, & Electronic Frontier Foundation, 2024). Indeed, the former UN Special Rapporteur on freedom of expression even went so far as to state that the conduct of the then-European Commissioner responsible for enforcement of the DSA had shown the DSA "can be abused" and "may have legitimized politicization of the DSA in ways that could be used to limit public debate" (Kaye, 2024). These critiques by civil society and international experts have also been reflected in the literature, where warnings about the dangers associated with the Commission as the enforcer of the DSA had been raised as the DSA was enacted, given how highly politicised content moderation can be in practice (Buri, 2023).

Finally, it should be recognised that the Commission's regulatory activity targeting VLOPs over disinformation may raise questions under Article 10 ECHR. In this regard, the ECHR has held that certain written warnings issued by public authorities may constitute an interference with freedom of expression, especially where it was mentioned that "a failure" to heed a warning "could result in liability" (*Karastelev and Others v. Russia*, 2020, para. 74). Crucially, the underlying legislation establishing the basis for the interference "must afford a measure of legal protection against arbitrary interferences by public authorities" with the right to freedom of expression in order for that legislation to meet the requirement of "prescribed by law" (*Karastelev and Others v. Russia*, 2020, para. 79; *Rid Novaya Gazeta and Zao Novaya Gazeta v. Russia*, 2021, para. 72). The Court has noted that, in the case of *Karastelev and Others v. Russia* (2020), "clear criteria" were absent, which lead to uncertainty. This uncertainty in turn "adversely affected the foreseeability of the regulatory framework" and led to the framework being "conducive to creating a negative chilling effect on freedom of expression," and leaving "too much discretion to the executive" (*Karastelev and Others v. Russia*, 2020, para. 90). Similarly, the Court found it "reasonable to assume" that "having recourse" to a "caution procedure" under legislation was "designed to have a non-negligible chilling effect directly affecting freedom of expression" (*Rid Novaya Gazeta and Zao Novaya Gazeta v. Russia*, 2021, para. 62). The Court found that a caution issued by a regulator "must have had a chilling effect" on the applicants' "freedom of expression because it warned them against covering certain matters (in a certain manner) or reproducing specific materials." (*Rid Novaya Gazeta v. Russia*, 2021, para. 62). Crucially, these Article 10 ECHR principles do raise a question mark over whether the DSA adequately sets boundaries on the European Commission's regulatory approach to disinformation under the systemic risks provisions.

## 5. Conclusion

This article has focused on the regulation of disinformation under the DSA, and some final points can be put forward. First, because of the DSA's notably broad definition of illegal content, the DSA's obligations placed on online platforms in relation to illegal content can be utilised to have disinformation removed when it comes within the definition of illegal content under national legislation. This demonstrates how the DSA can be instrumentalised in the removal of disinformation on platforms. Second, a crucial point about the regulation of disinformation under the DSA is that not only does the DSA facilitate the removal of disinformation, at the same time, it also seeks to protect individuals whose content has been removed because it is considered disinformation by platforms. In this regard, the DSA has a double-edged-sword approach to the regulation of disinformation: On the one hand, it facilitates the removal of disinformation, and on the other hand, it seeks to protect a user's freedom of expression when content is removed for being qualified as disinformation. Third, while disinformation is not mentioned in the DSA's provisions, the systemic risk provisions under Articles 34 and 35 are the most applicable to tackling disinformation. Notably, while these provisions do not mandate that platforms remove disinformation, most of the removal of disinformation is occurring under the 2022 Code, which is inextricably linked to the DSA and its risk-mitigation measures under Article 35. It is a crucial lesson to be learned that the DSA's regulation of disinformation cannot be read in isolation from the 2022 Code. Finally, the Commission has clearly opened a salvo against VLOPs based on tackling disinformation, but there are potential warning signs that need to be heeded to ensure that freedom of expression is adequately protected. In this regard, this article has sought to highlight how the Commission's regulatory action may need to better align with the right to freedom of expression, especially in providing sufficient guardrails for how the Commission conducts its regulatory approach to disinformation under the DSA's systemic risks provisions.

### Acknowledgments

The authors wish to thank the three anonymous peer reviewers for their very helpful comments.

### Funding

Publication of this article in open access was made possible through the institutional membership agreement between the University of Amsterdam and Cogitatio Press.

### Conflict of Interests

The authors declare no conflicts of interest.

### References

Access Now, ARTICLE 19, AlgorithmWatch, ApTI, 7amleh, Bits of Freedom, Centre for Democracy & Technology–Europe Office, Digitale Gesellschaft, Electronic Frontier Foundation, Electronic Frontier Finland, Epicenter.works, European Center for Not-for-Profit Law, European Digital Rights, Foundation The London Story, Homo Digitalis, INSM Foundation for Digital Rights, IT-Pol Denmark, Justitia/Future of Free Speech, Kandoo, . . . WHAT TO FIX. (2023, October 17). *Civil society open letter to Commissioner Breton: Precise interpretation of the DSA matters especially when people's lives are at risk in Gaza and Israel* [Open Letter]. https://www.article19.org/wp-content/uploads/2023/10/Civil-society-open-letter-to-Commissioner-Breton.pdf

Access Now, ARTICLE 19, & Electronic Frontier Foundation. (2024, August 19). *Civil society statement: Commissioner Breton needs to stop politicising the Digital Services Act* [Open Letter]. https://www.accessnow.org/press-release/commissioner-breton-stop-politicising-digital-services-act/

Albert, J. (2024, November 29). Researcher access to platform data: Experts weigh in on the Delegated Act. *DSA Observatory*. https://dsa-observatory.eu/2024/11/29/researcher-access-to-platform-data-experts-weigh-in-on-the-delegated-act

Amnesty International. (2019). *Surveillance giants: How the business model of Google and Facebook threatens human rights*. https://www.amnesty.org/en/documents/pol30/1404/2019/en

Appelman, N., & Leerssen, P. (2022). On "trusted flaggers." *Yale Journal of Law and Technology*, *24*, 452–475. https://yjolt.org/trusted-flaggers

Breton, T. [@ThierryBreton]. (2023, October 10). *Following the terrorist attacks by Hamas against* 🇮🇱*, we have indications of X/Twitter being used to disseminate illegal content & disinformation in the EU* [Post]. X. https://x.com/ThierryBreton/status/1711808891757944866

Breton, T. [@ThierryBreton]. (2024, August 12). *With great audience comes greater responsibility #DSA* [Post]. X. https://x.com/ThierryBreton/status/1823033048109367549

Brogi, E., & De Gregorio, G. (2024). From the code of practice to the code of conduct? Navigating the future challenges of disinformation regulation. *Journal of Media Law*, *16*(1), 38–46. https://doi.org/10.1080/17577632.2024.2362480

Buri, I. (2023). A regulator caught between conflicting policy objectives: Reflections on the European Commission's role as DSA enforcer. In van Hoboken, J., Quintais, J. P., Appelman, N., Fahy, R., Buri, I., & Straub, M. (Eds.), *Putting the DSA into practice: Enforcement, access to justice, and global implications* (pp. 75–90). Verfassungsbooks. https://doi.org/10.17176/20230208-093135-0

Charter of Fundamental Rights of the European Union, 2012, Pub. L. No. C 326/391.

Communication from the Commission—Commission guidelines for providers of very large online platforms and very large online search engines on the mitigation of systemic risks for electoral processes pursuant to article 35(3) of regulation (EU) 2022/2065. (2024). *Journal of the European Union*, C/2024/3014. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52024XC03014&qid=1714466886277

Convention for the Protection of Human Rights and Fundamental Freedoms, 1950, Pub. L. No. 213 U.N.T.S 221.

Criminal Code of the Republic of Malta, 1854, Chapter 9. https://legislation.mt/eli/cap/9/eng/pdf

Culliford, E., & Paul, K. (2020, November 20). Facebook offers up first-ever estimate of hate speech prevalence on its platform. *Reuters*. https://www.reuters.com/article/uk-facebook-content-idINKBN27Z2QY

Directorate-General for Communications Networks, Content and Technology. (2018). *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. https://data.europa.eu/doi/10.2759/739290

Efroni, Z. (2021). *The Digital Services Act: Risk-based regulation of online platforms*. Internet Policy Review. https://policyreview.info/articles/news/digital-services-act-risk-based-regulation-online-platforms/1606

Espaliú-Berdud, C. (2022). Legal and criminal prosecution of disinformation in Spain in the context of the European Union. *Profesional De La Información*, *31*(3) Article e310322. https://doi.org/10.3145/epi.2022.may.22

European Commission. (2018). *2018 code of practice on disinformation*. https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation

European Commission. (2020a). *Digital Services Act: Deepening the internal market and clarifying responsibilities for digital services* (Document Ares(2020)2877686). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=pi_com:Ares(2020)2877686

European Commission. (2020b). *Joint communication to the European Parliament, the European council, the*

*council, the European Economic and Social Committee and the Committee of the Regions. Tackling Covid-19 disinformation—Getting the facts right*. JOIN/2020/8 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020JC0008

European Commission. (2022). *The strengthened code of practice on disinformation 2022*. https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

European Commission. (2023a). *Designation decisions for the first set of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs)*. https://digital-strategy.ec.europa.eu/en/library/designation-decisions-first-set-very-large-online-platforms-vlops-and-very-large-online-search

European Commission. (2023b, April 25). *Digital services act: Commission designates first set of very large online platforms and search engines* [Press Release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413

European Commission. (2023c, October 12). *The Commission sends request for information to X under the digital services act** [Press Release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4953

European Commission. (2023d, October 19). *Commission sends request for information to Meta under the digital services act* [Press Release]. https://digital-strategy.ec.europa.eu/en/news/commission-sends-request-information-meta-under-digital-services-act-0

European Commission. (2023e, December 18). *Commission opens formal proceedings against X under the digital services act* [Press Release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709

European Commission. (2024a, April 30). *Commission opens formal proceedings against Facebook and Instagram under the digital services act* [Press Release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373

European Commission. (2024b, March 26). *Commission publishes guidelines under the DSA for the mitigation of systemic risks online for elections* [Press Release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1707

European Commission. (2024c). *Delegated regulation on data access provided for in the digital services act*. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en

European Commission. (2025a). *Code of conduct on disinformation*. https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation

European Commission. (2025b, February 13). *Commission endorses the integration of the voluntary code of practice on disinformation into the digital services act* [Press Release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_25_505

European Commission. (2025c). *Trusted flaggers under the digital services act (DSA)*. https://digital-strategy.ec.europa.eu/en/policies/trusted-flaggers-under-dsa

European Commission. (2025d). *The enforcement framework under the digital services act*. https://digital-strategy.ec.europa.eu/en/policies/dsa-enforcement

European Regulators Group for Audiovisual Media Services. (2020). *Notions of disinformation and related concepts (ERGA report)*. https://erga-online.eu/wp-content/uploads/2021/03/ERGA-SG2-Report-2020-Notions-of-disinformation-and-related-concepts-final.pdf

Galantino, S. (2023). How will the EU digital services act affect the regulation of disinformation? *SCRIPTed*, *20*(1), 89–129.

Goldberger, I. (2024, May 13). Europe's digital services act: Where are all the trusted flaggers? *TechPolicy Press*. https://www.techpolicy.press/europes-digital-services-act-where-are-all-the-trusted-flaggers

Google. (2024). *Code of practice on disinformation—Report of Google for the period of 1 January 2024 to 30 June 2024*. https://shorturl.at/y5xWT

Griffin, R. (2024). Codes of conduct in the digital services act. *Technology and Regulation*, *2024*, 167–187. https://doi.org/10.26116/techreg.2024.016

Holznagel, D. (2024a, May 30). Follow me to unregulated waters! Are major online platforms violating the DSA's rules on notice and action? *Verfassungsblog*. https://doi.org/10.59704/80267b8bd7a278a4

Holznagel, D. (2024b, November 5). Art. 21 DSA has come to life. *Verfassungsblog*. https://doi.org/10.59704/8d27bd5f2320bbae

Husovec, M. (2024). The digital services act's red line: What the Commission can and cannot do about disinformation. *Journal of Media Law*, *16*(1), 47–56. https://doi.org/10.1080/17577632.2024.2362483

Jaursch, J. (2023). Platform oversight: Here is what a strong digital services coordinator should look like. In van Hoboken, J., Quintais, J. P., Appelman, N., Fahy, R., Buri, I., & Straub, M. (Eds.), *Putting the digital services act into practice: Enforcement, access to justice, and global implications* (pp. 91–104). Verfassungsbooks

Karastelev and Others v. Russia, Application no. 16435/10 (2020).

Kaushal, R., Van De Kerkhof, J., Goanta, C., Spanakis, G., & Iamnitchi, A. (2024). Automated transparency: A legal and empirical analysis of the digital services act transparency database. In R. Binns, F. Calmon, A. Olteanu, & M. Veale (Eds.), *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1121–1132). ACM. https://doi.org/10.1145/3630106.3658960

Kaye, D. (2024, March 21). The risks of internet regulation. *Foreign Affairs*. https://www.foreignaffairs.com/united-states/risks-internet-regulation

Khan, I. (2021). *Disinformation and freedom of opinion and expression: Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression* (UN Doc. A/HRC/47/25). United Nations. https://documents.un.org/doc/undoc/gen/g21/085/64/pdf/g2108564.pdf

Koltay, A. (2025). Freedom of expression and the regulation of disinformation in the European Union. In R. J. Krotoszynski Jr., A. Koltay, & C. Garden (Eds.), *Disinformation, misinformation, and democracy: Legal approaches in comparative context* (pp. 133–160). Cambridge University Press. https://doi.org/10.1017/9781009373272.010

Leerssen, P. (2021, September 7). Platform research access in Article 31 of the Digital Services Act. Sword without a shield? *Verfassungsblog*. https://doi.org/10.17176/20210907-214355-0

Leerssen, P. (2023). An end to shadow banning? Transparency rights in the digital services act between content moderation and curation. *Computer Law & Security Review*, *48*, Article 105790. https://doi.org/10.1016/j.clsr.2023.105790

LinkedIn. (2024). *LinkedIn systemic risk assessment: August 2024—Assessment report*. https://content.linkedin.com/content/dam/help/tns/en/2024_LinkedIn_DSA_SRA_Report_23_Aug_24.pdf

McGonagle, T. (2017). "Fake news": False fears or real concerns? *Netherlands Quarterly of Human Rights*, *35*(4), 203–209. https://doi.org/10.1177/0924051917738685

Meta. (2024a). *Regulation (EU) 2022/2065 digital services act transparency report for Facebook*. https://transparency.meta.com/sr/dsa-transparency-report-sep2024-facebook

Meta. (2024b). *Regulation (EU) 2022/2065 digital services act transparency report for Instagram*. https://shorturl.at/WvNSj

Meta. (2024c). *Regulation (EU) 2022/2065 digital services act (DSA): Systemic risk assessment and mitigation report for Facebook—August 2024*. https://tinyurl.com/2zk43475

Microsoft. (2024). *Code of practice on disinformation—Report of Microsoft for the period 1 July–31 December 2023*. https://disinfocode.eu/reports/download/44

Ministry of the Interior and Kingdom Relations. (2023). *Inzet Trusted Flagger Status BZK Provinciale Staten en Waterschapsverkiezingen*. https://www.rijksoverheid.nl/documenten/rapporten/2023/09/25/15-bijlage-bij-brief-rapportage-inzet-trusted-flagger-status-bzk

Mündges, S., & Park, K. (2024). But did they really? Platforms' compliance with the code of practice on disinformation in review. *Internet Policy Review*, *13*(3). https://doi.org/10.14763/2024.3.1786

Nooren, P., van Gorp, N., van Eijk, N., & Ó Fathaigh, R. (2018). Should we regulate digital platforms? A new framework for evaluating policy options. *Policy & Internet*, *10*(3), 264–301. https://doi.org/10.1002/poi3.177

Ó Fathaigh, R., Helberger, N., & Appelman, N. (2021). The perils of legally defining disinformation. *Internet Policy Review*, *10*(4). https://doi.org/10.14763/2021.4.1584

Quintais, J. P., Appelman, N., & Ó Fathaigh, R. (2023). Using terms and conditions to apply fundamental rights to content moderation. *German Law Journal*, *24*(5), 881–911. https://doi.org/10.1017/glj.2023.53

Radu, B. (2023). A policy perspective on regulating disinformation in Romania during the Covid-19 pandemic. *NISPAcee Journal of Public Administration and Policy*, *16*(1), 108–137. https://doi.org/10.2478/nispa-2023-0005

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). *Journal of the European Union*, L 277/1.

Rid Novaya Gazeta and Zao Novaya Gazeta v. Russia, Application no. 44651/11 (2021).

Rosati, E. (2024, February 22). The DSA's trusted flaggers: Revolution, evolution, or mere gattapardismo? *VerfassungsBlog*. https://verfassungsblog.de/the-dsas-trusted-flaggers

Salov v. Ukraine, Application no. 65518/01 (2005).

Sergejs Buivids v. Datu valsts inspekcija, Case C-345/17, ECLI:EU:C:2019:122 (2019).

Sullivan, D., & Pielemeier, J. (2023, July 19). Unpacking "systemic risk" under the EU's digital service act. *Tech Policy Press*. https://www.techpolicy.press/unpacking-systemic-risk-under-the-eus-digital-service-act

The Criminal Code Law of Cyprus, 2025, Chapter 154. http://www.cylaw.org/nomoi/enop/non-ind/0_154/full.html

TikTok. (2024). *Code of practice on disinformation—Report of TikTok for the period 1 January 2024–30 June 2024*. https://disinfocode.eu/reports/download/64

UN General Assembly. (2018). *Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression* (UN Doc. A/HRC/38/35). United Nations. https://www.ohchr.org/en/documents/thematic-reports/ahrc3835-report-special-rapporteur-promotion-andprotection-right-freedom

UN General Assembly. (2020). *Disease pandemics and the freedom of opinion and expression: Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression* (UN Doc. A/HRC/44/49). United Nations. https://www.ohchr.org/en/documents/thematic-reports/ahrc4449-disease-pandemics-and-freedom-opinion-and-expression-report

UN, Organization for Security and Co-operation in Europe, Organization of American States, & African Commission on Human and People's Rights. (2017). *Joint declaration on freedom of expression and "fake news", disinformation and propaganda* (FOM.GAL/3/17). Organization for Security and Co-operation in Europe. https://www.osce.org/fom/302796

van de Kerkhof, J. (2024). *Constitutional aspects of trusted flaggers in the Netherlands*. SSRN. https://doi.org/10.2139/ssrn.4943851

van Hoboken, J., & Ó Fathaigh, R. (2021). Regulating disinformation in Europe: Implications for speech and privacy. *UC Irvine Journal of International, Transnational, and Comparative Law*, *6*(1), 9–36, https://escholarship.org/uc/item/87m4t6vf

Vermeulen, M. (2024, October 29). Reading the European Commission's proposed implementation of

DSA Article 40: Six initial observations on a new framework for research data access. *Tech Policy Press*. https://www.techpolicy.press/reading-the-european-commissions-proposed-implementation-of-dsa-article-40-six-initial-observations-on-a-new-framework-for-research-data-access

X. (2023). *Report setting out the results of Twitter international unlimited company risk assessment pursuant to Article 34 EU digital services act: September 2023*. https://transparency.x.com/content/dam/transparency-twitter/dsa/dsa-sra/TIUC-DSA-SRA-Report-2023.pdf

York, J., & McSherry, C. (2019, April 29). Content moderation is broken: Let us count the ways. *Electronic Frontier Foundation*. https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways

## About the Authors

**Ronan Ó Fathaigh** is an assistant professor at the Institute for Information Law (IViR), Faculty of Law, University of Amsterdam.

**Doris Buijs** is a former junior researcher at the Institute for Information Law (IViR), Faculty of Law, University of Amsterdam.

**Joris van Hoboken** is a full professor at the Institute for Information Law (IViR), Faculty of Law, University of Amsterdam. At IViR, he is appointed to the chair information law, with special emphasis on law and digital infrastructure.