

WHAT IS THE DIGITAL SERVICES ACT & WHAT DOES IT MEAN FOR HATE SPEECH MONITORING & RESPONSES?

KEY CHALLENGES IN THE IMPLEMENTATION OF THE DSA

Written by Daniel Heller.

Edited by Joanna Perry.

Publication produced by
CEJI – A Jewish Contribution to an Inclusive
Europe on behalf of the Facing Facts
Network.

We are very grateful to the respondents
who took part in our survey.

Cover design, layout, graphics and
illustrations:
Jonathan Brennan
www.aptalops.com / info@aptalops.com

© 2024 CEJI

www.facingfacts.eu

All rights reserved. The contents of this
publication may be freely used and copied
for educational and other non-commercial
purposes, provided that any such
reproduction or referencing is accompanied
by an acknowledgment stating:

‘Source: Facing Facts Network © 2024 CEJI’

Since 2022, the Facing Facts Secretariat
manages the Facing Facts Network.

The network consists of members from
various EU countries who act as change
makers to make the impact of hate visible
and provide understandings and responses
to hate crime and hate speech in Europe
and beyond.

www.facingfacts.eu/members




The Facing Facts Network is an initiative
coordinated by CEJI – A Jewish Contribution
to an Inclusive Europe.

www.ceji.org



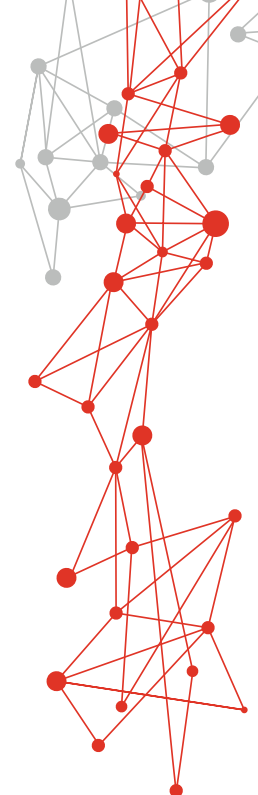
**Funded by
the European Union**

Views and opinions expressed are however
those of the author(s) only and do not
necessarily reflect those of the European
Union or European Education and Culture
Executive Agency (EACEA). Neither the
European Union nor the granting authority
can be held responsible for them.



FACING FACTS POLICY BRIEF: WHAT IS THE DIGITAL SERVICES ACT AND WHAT DOES IT MEAN FOR HATE SPEECH MONITORING AND RESPONSES?

KEY CHALLENGES IN THE IMPLEMENTATION OF THE DSA



Contents

Introduction and overview	2
Roles, responsibilities and cooperation under the DSA	3
<i>Figure one: roles, powers and relationships</i>	3
Challenges and opportunities in hate speech monitoring and responses under the DSA	7
<i>Trusted flaggers</i>	8
<i>Hate speech decision-making</i>	9
<i>Conclusions and ways forward</i>	10
Recommendations	11
<i>Facing Facts</i>	11
<i>European Commission</i>	12
<i>National Digital Services Coordinator</i>	12
<i>Governments</i>	13
<i>Trusted Flaggers</i>	13
<i>Regulated Entities</i>	13

Introduction and overview

The Digital Services Act (DSA) entered into force in February 2024. Marking a pivotal moment in EU leadership on online content regulation, the European Commission explains, ‘The Digital Services Act (DSA) and the Digital Market Act (DMA) form a single set of rules that apply across the whole EU. They have two main goals: To create a safer digital space in which the fundamental rights of all users of digital services are protected, [and] to establish a level playing field to foster innovation, growth, and competitiveness, both in the European Single Market and globally.’¹

The DSA shifts EU strategy on combating hate speech towards a proactive paradigm by giving the Commission oversight over cooperation between social media platforms and other key actors alongside powers to fine platforms and scrutinise Social Media algorithms. New responsibilities for online platforms have been imposed, and specific roles including National Digital Services Coordinators (DSCs) and Trusted Flaggers have been created.² Building on its predecessors, the Code of Conduct on countering illegal hate speech online and E-Commerce Directive,³ the DSA holds platforms accountable for user-generated content if they fail to take action against illegal hate speech after being notified of it.

The Commission is already starting to use its powers, launching formal proceedings against several regulated entities⁴ and Tiktok.⁵ The recent ECtHR Grand Chamber judgement *Sanchez v. France*⁶ establishes that in some specific situations individual users of social media platforms may be held accountable for failing to address hate speech.

In 2023, Facing Facts Network members raised questions about the impact the DSA would have on their monitoring and response activities regarding online hate speech. In response, the Facing Facts Secretariat developed and disseminated a survey during January-February 2024 to better understand the experiences and concerns of Facing Facts Network members and other organisations working on this field. This briefing aims to reflect their perspective and offer contextualised analysis and recommendations.

Our analysis shows a lack of information on DSA implementation at the national level, with implications for hate speech monitoring and responses. The apparent lack of industry knowledge of the Digital Services Coordinators appointed so far, combined with the process of appointing official Trusted Flaggers under the DSA, risks undermining the contribution, knowledge and network of civil society organisations that developed under the Code of Conduct of the previous regime.

The Commission and its partners are not starting in a vacuum in terms of legislation and networks. The key will be to build on the hate speech response system that has strengthened under the Code of Conduct and other measures, paying particular attention to the unique

¹ <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>

² See the 2022 Facing Facts report on Current Activities & Gaps in hate speech responses: A mapping report for the Facing Facts Network, p. 12: <https://www.facingfacts.eu/wp-content/uploads/sites/4/2023/04/Facing-Facts-Network-Mapping-Report-v8.pdf> and The European Commission against Racism and Intolerance (ECRI) General Policy Recommendation N°15 on combating hate speech (GPR 15): <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01>

³ Directive 2000/31/EC, Directive of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, section 230.

⁴ ‘formal proceedings to assess whether TikTok may have breached the Digital Services Act (DSA) in areas linked to the protection of minors, advertising transparency, data access for researchers, as well as the risk management of addictive design and harmful content.’ https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926, the Commission reports, ‘formal proceedings to assess whether X may have breached the Digital Services Act (DSA) in areas linked to risk management, content moderation, dark patterns, advertising transparency and data access for researchers’ https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709. See also links to proceedings against META, https://ec.europa.eu/commission/presscorner/detail/en/IP_24_2664, and Aliepress https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485

⁵ The Commission reports, ‘formal proceedings to assess whether X may have breached the Digital Services Act (DSA) in areas linked to risk management, content moderation, dark patterns, advertising transparency and data access for researchers’ https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709

⁶ *Sanchez v. France*, no. 45581/15, Grand Chamber judgement of 15 May 2023.

contribution of expert CSOs that understand the nature and impact of hate speech on the most vulnerable communities in Europe. The final section of this briefing makes a number of recommendations that aim to achieve this goal.

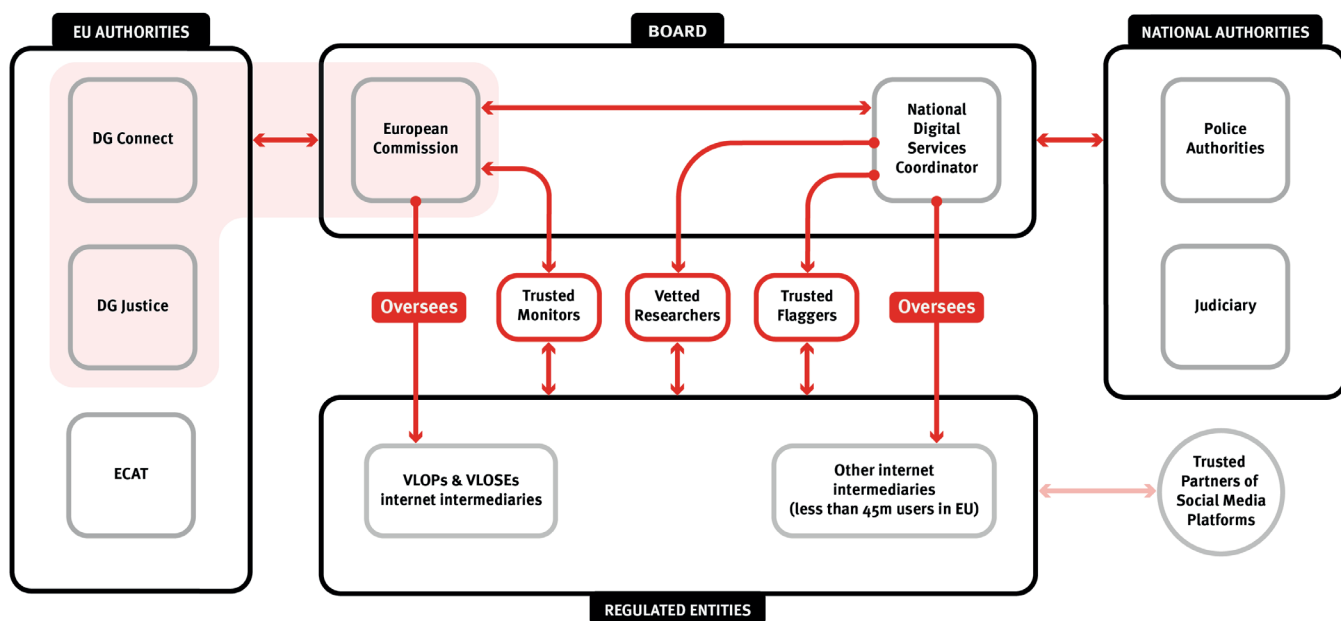
The implementation of the DSA takes place in a context of rapid development in policy and technology with the launch of the Code of Practice on Disinformation⁷, and the AI Act⁸, which aims to ‘harmonise compliance and tackle cross-border obstacles efficiently’. It will be important to closely monitor these developments and to support key stakeholders to understand their ongoing implications for their monitoring, response and victim support efforts.

Roles, responsibilities and cooperation under the DSA

The Commission explains that the DSA ‘provides a framework for cooperation between the Commission, EU and national authorities to ensure platforms meet its obligations’⁹. The graphic below sets out the key players and their powers.

Figure one: roles, powers and relationships

This graphic shows the key entities and relationships established by the Digital Services Act. If you would like to access the interactive graphic online, scan the QR code below or visit: <https://www.facingfacts.eu/facing-facts-policy-brief>



OR <https://www.facingfacts.eu/facing-facts-policy-brief>

⁷ <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>

⁸ DG CONNECT coordinates the implementation of both the AI Act and DSA.

⁹ See more: [https://digital-strategy.ec.europa.eu/en/policies/dsa-cooperation#:~:text=The%20Digital%20Services%20Act%20\(DSA,ensure%20platforms%20meet%20its%20obligations.](https://digital-strategy.ec.europa.eu/en/policies/dsa-cooperation#:~:text=The%20Digital%20Services%20Act%20(DSA,ensure%20platforms%20meet%20its%20obligations.)

The European Board for Digital Services

The Board is composed of the Member States' Digital Services Coordinators and chaired by the European Commission.

The Board advises the Digital Services Coordinators and the European Commission in accordance with the Digital Services Act to achieve the following objectives:

1. contributing to the consistent application of the Digital Services Act and effective cooperation of the Digital Services Coordinators and the European Commission with regard to matters covered by the Digital Services Act;
2. coordinating and contributing to guidelines and analysis of the European Commission and Digital Services Coordinators and other competent authorities on emerging issues across the internal market with regard to matters covered by the Digital Services Act; and
3. assisting the Digital Services Coordinators and the European Commission in the supervision of very large online platforms and very large online search engines.

National Authorities

National authorities including the police, prosecutors and judicial authorities are responsible for the investigation and prosecution of cases of criminal hate speech according to domestic law. Their role requires strong cooperation with Digital Services Coordinators, which are appointed by national authorities. DSCs carry out their functions—such as imposing fines or conducting investigations—in close cooperation with public authorities.

European Commission (EC)

The Commission is granted three main powers under the DSA: oversight, compliance and fines.

Oversight: The Commission determines whether VLOPs & VLOSEs – platforms that have more than 45 million active users per month in the EU¹⁰— are complying with their specific obligations under the DSA and lead on setting up auditing procedures including with competent national authorities to monitor compliance (See: <https://digital-strategy.ec.europa.eu/en/policies/dsa-cooperation>).

Regulated entities must grant access to the Commission and provide explanations regarding their databases and algorithms.

Compliance: The Commission may order the regulated entities to take necessary measures to mitigate and address systemic risks in the platform, eg: to draft an action plan to ensure compliance within an agreed timeframe and to inform the Commission about the intended compliance measures. The Commission will monitor the implementation of the action plan, requiring the provider to submit an audit report and keep the Commission updated on progress.

Fines: The Commission may impose fines of up to 6% of the provider's total worldwide annual turnover for 'intentional' or 'negligent' infringement; and fines up to 1% of the total annual income or worldwide turnover may be imposed for providing incorrect, incomplete, or misleading information, or failing to comply with Commission measures.

Together with the Board, in cases of “crisis situations” (ie in the case of extraordinary circumstances that lead to a serious threat to public security or public health), the Commission may require providers to analyse whether their services contribute to that

¹⁰ These guidelines explain how to determine the number of users of a given platform: <https://digital-strategy.ec.europa.eu/en/library/dsa-guidance-requirement-publish-user-numbers>

threat as established under Article 48 of the DSA [For example, this mechanism has been used in relation to the Hamas-Israel war.]

The European Commission is responsible for publishing the information on trusted flaggers in a publicly available database in an easily accessible and machine-readable format, ensuring the database is kept up to date.

Digital Services Coordinator

Digital Service Coordinators (DSCs) are designated by Member States. They coordinate the implementation of the DSA at the national level. They must be independent from any public authority and private party.

See the current list: <https://digital-strategy.ec.europa.eu/en/policies/dsa-dscs>

DSCs have several specific roles:

1. oversight of the smaller regulated entities that do not fall under the categorisation of VLOPs/VLOSEs. DSCs can access data, conduct inspections, and impose fines on these national entities. This will be done based on guidelines to be shared at the EU level.
2. designate Trusted Flaggers and can revoke the accreditation of Trusted Flaggers should the entity no longer meet the conditions set out in Article 22 (2), for example, where a Trusted Flagger consistently submits notices to Social Media Platforms that are either imprecise, inaccurate, or lack sufficient evidence. The DSCs need to communicate names, addresses, and email addresses of entities awarded the status of trusted flagger.
3. contribute to the sharing of information/expertise and understand the implications for the EU/National level through exchanges at the Board level.
4. Assess researchers applications for data access and act as a link between researchers and VLOPs/VLOSEs

Trusted Monitors

As part of the update of the EC Code of Conduct to the 'DSA Code', as specified in Article 45 Codes of Conduct of the DSA, organisations will be appointed by the European Commission and by Regulated entities as Trusted Monitor. The role of the Trusted Monitor is to conduct diligent, accurate, and objective monitoring of online platforms. This involves evaluating the platforms' compliance with legal obligations, voluntary codes of conduct, and terms of service, particularly in relation to identifying and addressing illegal content, including hate speech.

Vetted researchers

Under article 40 of the Digital Services Act (DSA), vetted researchers will be able to request data from very large online platforms (VLOPs) and search engines (VLOSEs) to conduct research on systemic risks in the EU. Systemic risks are set out in Article 34(1) of the DSA and include the dissemination of illegal content, and negative effects for the exercise of fundamental rights, including the rights to non-discrimination (among others). There are several conditions to becoming a vetted researcher, which can be found [HERE](#)

Trusted Flagger

Trusted Flaggers are appointed by Digital Services Coordinators based on their particular expertise and competence in detecting, identifying, and notifying illegal content in the national context. They must be independent from any provider of online platforms.

The role of the Trusted Flagger is to ensure diligent, accurate, and objective action in detecting, identifying, and notifying illegal content.

At least once a year, Trusted flaggers must publish detailed reports on the notices they have submitted to online platforms. These reports must include the number of notices, the type of allegedly illegal content, and the action taken by the provider. The reports should also explain the procedures in place to ensure the trusted flagger's independence. These reports must be sent to the DSC and made publicly available, ensuring that they do not contain personal data.

European Centre for Algorithmic Transparency (ECAT)

ECAT was established in April 2023. Its role is to support the Commission and national authorities in their duties under the DSA. ECAT conducts technical assessments—for example on the functioning of social media platforms' algorithms—analyses transparency reports from online platforms, and facilitates cooperation with various national and international bodies.

Regulated Entities – VLOPs/VLOSEs Internet Intermediaries

Very large online platforms and search engines are those with over 45 million users in the EU. They must comply with all of the rules listed here. Smaller platforms only need to comply with points 1-4.

Under the DSA, all platforms, including VLOPs & VLOSEs must:

1. establish a point of contact for authorities and users.
2. report criminal offences.
3. have user-friendly terms and conditions.
4. be transparent regarding advertising, recommender systems or content moderation decisions.

Specific for VLOPs/VLOSEs

5. identify, analyse, and assess systemic risks that are linked to their services. The definition of systemic risk is very broad and includes illegal content, hate speech, privacy violations, election manipulation, etc. Once the risks are identified, VLOPs and VLOSEs must put measures in place that aim to mitigate them.
6. Undergo an annual independent audit into their compliance with the DSA, overseen by the EC and paid for by the platform. The provider then has one month to either adopt the recommendations in the audit report or justify the reasons for not doing so and take alternative measures. The platforms can also carry out internal audits regarding compliance with their obligations on their own initiative.
7. Share their data with the Commission and national authorities so that they can monitor and assess compliance with the DSA.
8. Provide access to the data necessary to monitor their compliance with the DSA to the competent authorities and researches vetted by the authorities.
9. Establish a public repository on the online advertisements they displayed in the past year.
10. Provide at least one recommender system that is not based on profiling.
11. Provide a machine-readable summary of the terms and conditions, including available remedies and redress mechanisms.
12. Publish terms and conditions in each of the official languages of the countries they offer their services.
13. Establish a reporting mechanism for users.
14. report on the profile and size of human resources dedicated to content moderation every six months. The DSA requires that these staff are trained and language-diverse.

Online platforms are required to take the necessary measures to ensure that notices submitted by trusted flaggers are given priority and are processed and decided upon without delay.

Other regulated entities - under 45 million users in Europe

Under the DSA, all platforms, including those with under 45 million users in Europe must:

1. establish a point of contact for authorities and users
2. report criminal offences
3. have user-friendly terms and conditions
4. be transparent regarding advertising, recommender systems or content moderation decisions

Digital Services Coordinators (DSCs) have oversight over these platforms at the national level.

Trusted Partners of social media companies

Between 2016 and 2022 Trusted Partners of social media companies carried out monitoring exercises under the voluntary EC Code of Conduct. They are mostly composed of Civil Society Organisations. This group includes Priority Flaggers with YouTube, Trusted Partners with X, Community Partners Channel with TikTok, and Trusted Partners for the META platforms.

Challenges and opportunities in hate speech monitoring and responses under the DSA

15 organisations from 11 countries responded to the Facing Facts questionnaire, including 12 CSOs and three public authorities. Thirteen respondents are members of the Facing Facts Network. Based on an analysis of survey responses, publicly available information on the DSA, and the Network secretariat's ongoing policy and advocacy work, several challenges facing the effective implementation of the DSA can be identified.

Available information and the role of the Digital Services Coordinators

While survey respondents reported a high level of awareness about the DSA and its provisions,¹¹ several were critical of the lack of information on the implementation of the DSA at the national level:

‘There was absolutely no information on the implementation of the DSA and appointing of the Digital Services Coordinator.’¹²

Several respondents to the survey pointed to a vacuum of monitoring at the national level because national DSCs and Trusted Flaggers have yet to be appointed, and monitoring exercises established under the EU Code of Conduct have been stalled. A further strong theme of concern expressed in the survey was that proposed DSCs do not have the knowledge, background or network of relationships to effectively carry out their role.

A respondent explained, ‘[the] soon-to-be-appointed National Digital Services Coordinator is an expert body that does not have any previous contact [with] civil society and isn't aware of the work done in this area re: hate speech (primarily the implementation of The EU Code of conduct on countering illegal hate speech online).’ Indeed a recent list of appointed DSCs reveals that those appointed to date are agencies for competition and media services with little obvious experience of hate speech monitoring.¹³

¹¹ For example, 60% agreed with the statement ‘I am aware of the impact the Digital Services Act will have on the activities of my organisation’; 80% agreed with the statement ‘I am aware of the reporting needs under the Digital Services Act’.

¹² Survey respondents also reported a general lack of information about the appointment of Trusted Flaggers at the national level.

¹³ See the list of appointed Digital Services Coordinators here: <https://digital-strategy.ec.europa.eu/en/policies/dsa-dscs>

Other responses underline these concerns and point to the central importance of building and maintaining strong relationships with civil society experts and equality bodies, effective coordination with the Commission and undertaking training.

‘I would like to stress [...] the fact that authorities that will be the coordinators under DSA [do] not have any, or had very limited, previous contact with civil society. These institutions need to start consider[ing] their stakeholders differently (because they are usually market oriented) and steps need to be made in order to develop normal relationships and horizontal cooperation.’

‘While some of these institutions focused on online hate speech before, the majority of them have not yet had such an agenda within their scope and capacities. Therefore, their new responsibilities should encompass comprehensive internal training for a specialised team of employees, as well as international coordination between DSCs and EC.’

‘... [DSAs] are all agencies that worked on tele-communications and competence regulations. Nothing related to hate speech, and mostly the non-inclusion of [an] Equality body in any ‘table’ I think...is also particularly problematic. They are creating this Board (all the DSCs from all EU MS), but they are not consulting or working building up from the good practices of Equality bodies or CSO-institutions working together.’

Trusted flaggers

The burden to identify hate speech falls overwhelmingly on Trusted Flaggers which have a significantly lower status and resources than the platforms/regulated entities they are tasked to monitor. As such they should not replace the role of these entities as primary duty-bearers for identifying and reporting hate speech. Although platforms must take down content which is not in compliance with their terms and conditions, the risk is that most hate speech goes unaddressed unless reported by Trusted Flaggers – or platforms’ Trusted Partners. Relying on incidents submitted by third parties is not a feasible strategy for effectively addressing hate speech. Sufficient funding needs to be secured for Trusted Flaggers to meet the DSA’s expectation, especially due to the requirement of independence from any provider of online platforms.¹⁴

Transparent accreditation and removal processes of Trusted Flaggers are among the most sensitive areas of DSCs’ responsibility. However one survey respondent expressed the following concern, ‘There is a risk that the Digital Services Coordinator will not be neutral when awarding the Trusted Flagger status.’ DSCs will also need to pay particular attention to not exclude communities from being appointed as Trusted Flaggers, and to ensure that all communities are represented and consulted whether or not they are covered by national legislation on ‘hate speech’. For example, effective hate speech monitoring often requires specific knowledge about linguistic and historical contexts. Diverse pools of Trusted Flaggers are more likely to have this knowledge as well as relationships with various communities. Comprehensive training programs and continuous evaluation mechanisms are also essential for an effective Trusted Flagger system. While there has been progress on developing application processes and guidelines at the national level,¹⁵ as of June 2024, the necessary methodologies and training mechanisms have yet to be fully implemented by the Board for DSCs.¹⁶

¹⁴ See Article 22 (2) (b) of the DSA.

¹⁵ The Board meets regularly and developments are ongoing. National guidelines have been shared by Ireland DSC. Please see https://www.cnam.ie/wp-content/uploads/2024/02/20240216_Article22_GuidanceForm_Branded_vF_KW.pdf

¹⁶ The Board meets regularly and developments are ongoing. National guidelines have been shared by Ireland DSC. Please see https://www.cnam.ie/wp-content/uploads/2024/02/20240216_Article22_GuidanceForm_Branded_vF_KW.pdf

Referring back to Figure one, there is a lack of clarity about the position and role of those organisations previously appointed as ‘trusted partners/flaggers’ for Social Media Platforms. Between 2016-2022 these organisations were a central feature of the previous regime, carrying out monitoring exercises under the EC Code of Conduct. The upcoming update to the Code of Conduct is expected to offer greater clarity on their role. The majority of respondents to the survey were ‘trusted partners/flaggers’, however the pathway to becoming a Trusted Flagger under the DSA regime is not without obstacles. Eight respondents to the survey reported that they will apply for the trusted flagger status, six still need to finalise the decision, and one respondent explained that they will not due to budgetary issues.

Overall, for some organisations, applying for the official status of a Trusted Flagger under the DSA is not possible due to bureaucratic and financial constraints or lack of clarity about the terms of independence of their role. There is a concern that social media platforms and Member States will treat only official Trusted Flaggers as their partners and not sufficiently take into account the perspectives and knowledge of other key stakeholders, including the Platforms’ Trusted Partners, and the soon to be appointed Trusted Monitors. Indeed, while there is some anecdotal evidence that social media platforms will continue to engage with Trusted Partners appointed under the Code of Conduct without a formal role, a loss of knowledge and networks developed under the previous regime is a risk to the system. The relationship between Trusted Flaggers appointed under the DSA, Trusted Reporters, social media platforms and the Commission should be further clarified. One respondent explained,

‘For us the biggest risk is being made irrelevant or redundant by official Trusted Flaggers. For many of our members, applying for official TF (Trusted Flagger) status is a no go. With the appointment of other organisations, platforms and Member States could argue that only official TFs are relevant and non-official TFs are redundant or maybe even politically motivated [...] On the other side, becoming official TFs also has its risks. There is a lot of bureaucratic weight put on official TFs and no foreseeable funding line, which makes applying for the status not too desirable.’

Survey respondents also had questions on how the European Centre for Algorithm Transparency (ECAT) will monitor social media algorithms while also drawing on data from audits prescribed under the DSA, and how this will feed into the oversight process.

In a signal of good practice, one respondent to the survey reported that France is planning to pass its own digital services laws, clarifying certain DSA rules, specifying the role of the supervisory authority, and providing for criminal penalties for the authors of unlawful comments. Other Member States are likely to follow suit.

Hate speech decision-making

EU law has not so far provided a specific definition of hate speech. However, the 2008 Framework Decision on Combating Certain Forms and Expressions of Racism and Xenophobia (Framework Decision) identified the types of hate speech acts which should be criminalised across the EU.¹⁷ The Framework Decision is constructed around a closed list of protected grounds: ‘race,’ colour, religion, descent or national or ethnic origin. In 2021, the Commission proposed an initiative to extend the list of EU crimes, as per Article 83(1) of the Treaty on the Functioning of the EU (TFEU), with hate crime and hate speech regardless of a protected ground, based on sex, sexual orientation, gender identity or disability.¹⁸ The

¹⁷ Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law, adopted on 28 November 2008 by the Council of the EU. See See Article 1(1)(a) and (c) which require Member States to punish incitement to violence or hatred or directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin, as well as acts of denialism.

¹⁸ See Communication from the Commission to the European Parliament and the Council “A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime”, COM/2021/777 final.

unanimous consensus among all Member States in the European Council necessary to adopt this decision has yet to be sought.¹⁹

The DSA does not establish a substantive legal framework by, for example, providing a definition of illegal hate speech applicable across all Member States. According to the DSA, the term ‘illegal hate speech’ is to be interpreted based on ‘Union law or the law of any Member State’²⁰. Both the Framework Decision and national legislation lack some of the key protected grounds such as sexual orientation, gender identity and disability. If an expression is deemed illegal in one Member State, there is a question about how the Commission might take this on board with regard to its own powers in relation to VLOSEs and VLOPs in the context of another Member State in which a particular characteristic is not legally protected.

It is not clear whether the term ‘illegal’ refers only to criminalised hate speech or also speech that violates national civil or administrative law. As explained by one respondent,

‘The general public do not understand what constitutes criminal behaviour versus online (non criminal) hate speech. Clarity is required on what should be reported and to whom – the social media company, the national authority or the police.’

There is a need for clarity on the process of reporting. For example, if all reports are to go through the DSCs or the police, the burden could be overwhelming. If complaints go through social media platforms, there are questions regarding their legitimacy in deciding what is legal and what is not. If there is a legal direction that content needs to be removed, the user or the organisation, or any third party, can flag it to the social media platform, which is then required to remove the content. However, there are no proactive steps required to be taken by the social media platform. In this regard, the DSA has yet to represent progress from the 2016 Code of Conduct. For some respondents who have been involved in monitoring hate speech, there is a sense that the DSA might not lead to improvements, ‘the risk that, despite DSA rules, platforms will not improve their moderation of online hate content. Even today, on some platforms like X, we can’t see any improvement.’

Conclusions and ways forward

The DSA should increase access to justice, safety and support for victims and groups targeted by hate speech. Close and effective cooperation among the key actors described in Figure one as well as effective victim support will be essential to achieve this. There should be clarity on how national or administrative authorities’ orders or requests for information follow the DSA’s criteria and are complied with.²¹ To achieve this, there should be a unified methodology for hate speech monitoring, responses, and referrals across the EU, prioritising consistency to mitigate the risk of divergent practices among Member States. Article 35 of the DSA could support this approach:

‘Under Article 35 of the DSA the Commission, together with the Digital Services Coordinators of the Member States, may issue guidelines in relation to specific risks, to present best practices and recommend possible mitigation measures.’

Hate speech is a specific and systemic risk across the European Union. Article 35 guidelines could bring legacy knowledge and networks from the previous regime and help deliver the potential of the DSA as a step-change in EU regulation on hate speech. Guidance should also cover decision-making processes related to the assessment of what constitutes illegal hate speech and Trusted Flagger accreditation criteria and withdrawal procedures.²²

¹⁹ It is important to note that there is a high degree of divergence in legislation at the national level. The FF secretariat are compiling a list of legislation as a resource: <https://www.facingfacts.eu/hate-speech-legislation/>

²⁰ As specified in Article 3, letter H) of the DSA.

²¹ See Articles 9 and 10 of the DSA.

²² In a recent positive development the EC called on VLOPs & VLOSEs to create guidelines to protect election integrity.

Regulated entities need to take full responsibility to address hate speech on their platform, scrutinise and respond to any content that is incompatible with their terms and conditions regardless of whether they are notified by a third party. Article 9 and 10 of the DSA oblige platforms to cooperate with requests for information from national authorities and to appoint a single point of contact to support this process. These are key provisions to ensure victims can access their right to an effective investigation and an effective remedy. In this context the work and role of ECAT²³ could be further clarified.

Respondents made several suggestions such as to include an indicator for cooperation with civil society in the reports, oversight and analysis of the DSA implementation. Another suggestion is that Digital Services Coordinators should be obliged to organise an annual event with civil society on DSA implementation.

The role and position of DSCs in relation to established regulators, such as press and election regulators that have an existing power or duty with regard to hate speech responses should be clear. Existing relationships and mechanisms should be strengthened not unnecessarily replaced or made weaker. To support this outcome, DSCs should map their national context of regulation and related key stakeholders as a resource of responses, legal and non-legal, criminal and non-criminal. The [Facing Facts system mapping methodology](#) could be a useful resource for this exercise.

Equality bodies could play a key role by supporting efforts to effectively investigate and support prosecution of illegal hate speech, as well as systematically monitor and collect data on such cases.²⁴ For example, equality bodies could have powers to ensure that DSCs appoint Trusted Flaggers that represent all targeted communities, including those who are not currently covered by national hate speech laws.²⁵

Recommendations

This section draws on our findings and analysis to present detailed recommendations for all stakeholders with responsibilities within the ‘hate speech response system’ at the EU and national levels including the Facing Facts Network, the Commission, DSCs, Governments, Trusted Flaggers, and intermediaries/online platforms.

Facing Facts

- *Training:* Develop regular and up-to-date training programmes on hate speech monitoring and the regulatory framework, including the DSA. Ensure that training programmes incorporate the importance of mental health, and that they are trauma-informed.
- *Dissemination:* Share work of members, for example current projects involving ‘massive monitoring’ such as SafeNet.
- *Raise public awareness:* Share activities and information on how to identify and report criminal and non-illegal HS.

See: <https://digital-strategy.ec.europa.eu/en/news/commission-gathering-views-draft-dsa-guidelines-election-integrity>

²³ ECAT should support the regulator in assessing whether the functioning of algorithmic systems is in line with the risk management obligations that the DSA establishes for VLOPs and VLOSEs. It should become a gravity centre for the international community of researchers and expert auditors in the field, acting as a knowledge hub for research based on platforms’ data. See more at: https://ai-watch.ec.europa.eu/collaborations/european-centre-algorithmic-transparency-ecat_en or https://algorithmic-transparency.ec.europa.eu/index_en

²⁴ 2022 Facing Facts report on Current Activities & Gaps in hate speech responses: A mapping report for the Facing Facts Network, cited above, p. 11.

²⁵ 2022 Facing Facts report on Current Activities & Gaps in hate speech responses: A mapping report for the Facing Facts Network, cited above, p. 11.

- *Monitor implementation:* Work with members to monitor how the DSA is implemented at the national level, and bring key themes and gaps to the attention of partners at the EU level, including the High Level Group on Combating Hate Speech and Hate Crime.

European Commission

- *Transparency Guidelines:* Develop and publish comprehensive transparency guidelines under Article 35 of the DSA outlining the operationalisation of the DSA framework.²⁶ Ensure clarity on decision-making processes, accreditation criteria, and the process for withdrawing Trusted Flagger status under Article 22(2). Include civil society in the development phase in an inclusive and effective way, e.g. via focus groups or consultation meetings²⁷.
- *Unified Methodology:* Establish a unified methodology for hate speech decision-making across all EU Member States. This will guarantee consistency and avoid discrepancies in the interpretation and application of hate speech regulations by various actors such as Trusted Flaggers, platforms and also, subsidiarily, police, prosecutors and judicial or administrative authorities.
- *Coordination mechanism:* Establish an effective coordination mechanism, focusing on/drilling down to the national level, involving the Commission, DSCs, platforms and Trusted Flaggers as well as civil society organisations working on hate speech allowing for an open and continuous exchange of knowledge and information with the aim of effectively implementing the DSA.
- *Hate speech as an EU crime:* continue pursuing the effort to establish hate speech as an EU crime under Article 83(1) TFEU.
- *Support for Trusted Flaggers:* Provide adequate funding for Trusted Flaggers who may face challenges in meeting the independence requirements while, at the same time, fulfilling their role as those reporting illegal hate speech to platforms. Without their proactive involvement, the DSA mechanism loses much of its relevance. Tangible financial support provided to Trusted Flaggers is crucial for effective implementation of the DSA.

National Digital Services Coordinator

- *Continuous Training:* Institute ongoing training programs for DSC officers dealing with accreditations of trusted flaggers. This will enhance their capacity to make informed decisions, mitigating the risk of errors or biased evaluations. Participate in and organise training in collaboration with CSOs to understand their perspectives and challenges, understanding the (potential) Trusted Flaggers landscape and diversity.
- *Proactive collaboration with CSOs working in the area of discrimination, hate speech or hate crimes, as well as national Equality Bodies:* the DSCs should seek collaboration with the experts at the national level, including through consultation or training, to better understand the legal and non/legal problems of addressing hate speech and the needs of various victims and targeted communities.

²⁶ 'Under Article 35 of the DSA the Commission, together with the Digital Services Coordinators of the Member States, may issue guidelines in relation to specific risks, to present best practices and recommend possible mitigation measures.' Hate speech is a specific and systemic risk and guidelines would be a way to bring legacy knowledge from the previous regime.

²⁷ Following the Roundtable with Civil Society Organisations on the implementation of the Digital Services Act.

Governments

- *Legislative Alignment:* Align national legislation with relevant human rights standards, e.g. the Council of Europe Committee of Ministers Recommendation CM/Rec(2022)16 on hate speech, as well as with the DSA framework to ensure a cohesive approach. This will contribute to harmonising definitions of hate speech and streamlining legal responses across Member States.
- *Support for DSCs:* Provide adequate staffing and financing for DSCs. This support is crucial for effective implementation, preventing potential resource gaps or delays.
- *Support for Trusted Flaggers:* Provide adequate funding for Trusted Flaggers who may face challenges in meeting the independence requirements while, at the same time, fulfilling their role as those reporting illegal hate speech to platforms. Without their proactive involvement, the DSA mechanism loses much of its relevance. Tangible financial support provided to Trusted Flaggers is crucial for effective implementation of the DSA.
- *National cooperation mechanism:* Establish or support cooperation mechanisms to regularly exchange views between the DSCs, other relevant national authorities and civil society.
- Support the Commission's initiative to make hate crime and hate speech EU crimes in the European Council.

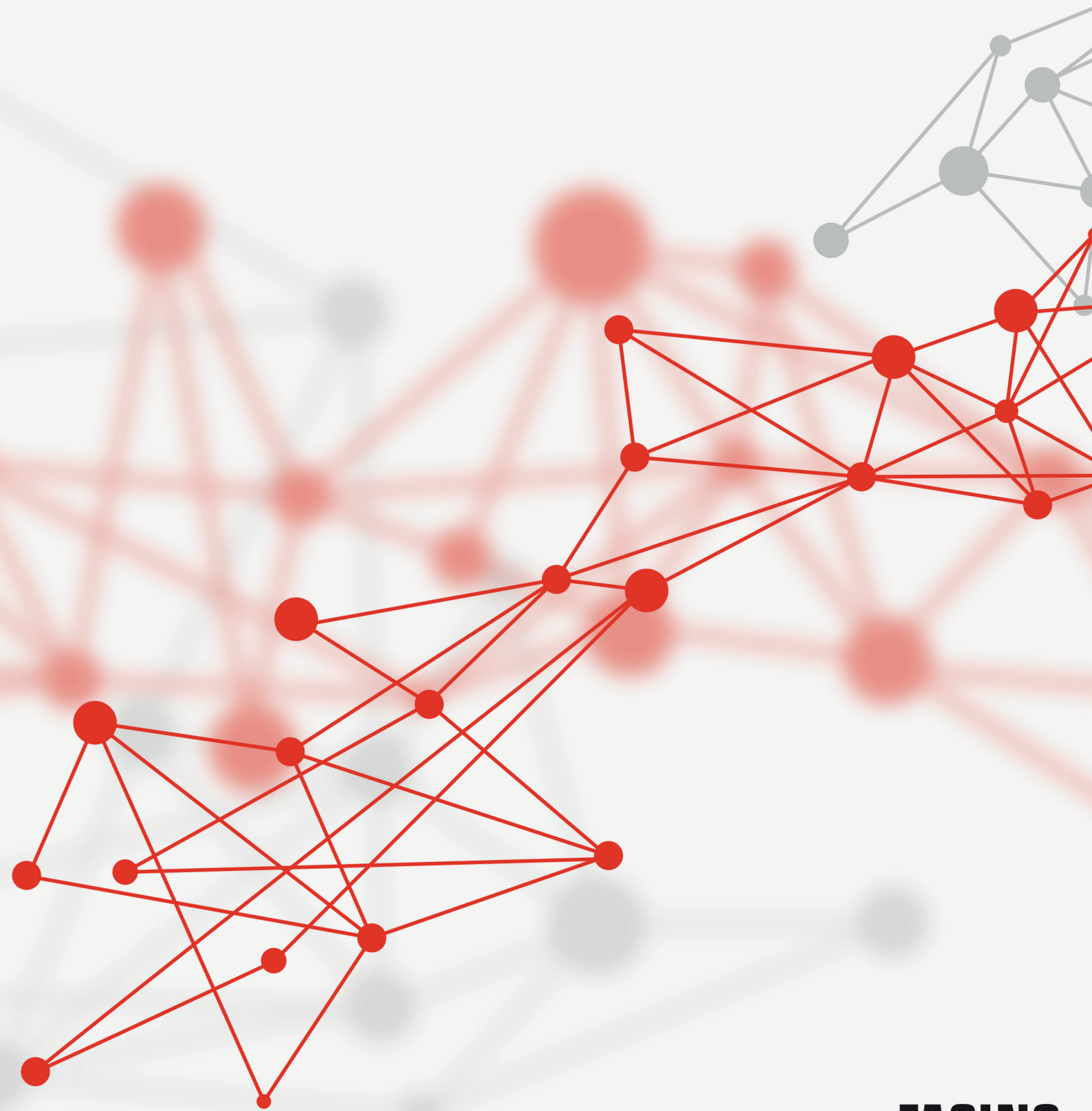
Trusted Flaggers

- *Diverse Accreditation Criteria:* Advocate the EC, DSCs and the Board for appropriate accreditation criteria to ensure representation from various organisations and communities working on different topics. This will foster a broad perspective in addressing hate speech, allowing the possibility of tackling diverse types of hate.
- *Transparent Methodology:* Call for a transparent methodology in the accreditation process, emphasising diversity, impartiality and adherence to predefined standards. This will ensure the reliability and fairness of the Trusted Flaggers system.
- *Training:* Ensure that staff is trained in the area of recognising illegal hate speech in the EU and national context.

Regulated Entities

- *Clear Reporting Mechanisms:* Implement clear reporting mechanisms in line with DSA requirements, ensuring accessibility and user-friendliness. This will facilitate efficient reporting of hate speech while upholding users' fundamental rights.
- *Regular Communication with DSCs, coordinated by the EC:* Establish regular communication channels with DSCs and civil society organisations to facilitate the exchange of information and support problem-solving. This collaboration will contribute to a more effective and transparent implementation of the DSA.
- *Increase linguistic expertise of content moderation team:* Ensure that the composition of moderation teams have a diversity of linguistic expertise, in addition to existing AI functionality.

- *Consistent application of own policies and regulations:* Besides complying with the DSA, ensure consistent and proactive application of own policies and regulations to address hate speech, seek improving these regulations in the light of the existing best practices, while respecting freedom of expression across the Member States. Ensure that sufficient resources are allocated to combating hate speech online by developing technological solutions grounded in human rights principles. This includes investing in AI-driven tools for automated moderation and enhancing the capacity of moderators and decision-makers. In collaboration with CSOs and monitoring organisations, regulated entities in the development and funding of this AI tool will support the effective implementation of the Digital Services Act (DSA).
- *Training:* Ensure that moderators together with VLOPs and VLOSEs staff receive training from civil society organisations on the impact of hate speech for the different communities and on its linguistic updates.



**FACING
FACTS**
NETWORK