



**Q&A REPORT**

# **Facing Facts**

# **DSA Webinar**

---

**Digital Services Act**

**What is the Digital Services Act (DSA) and what does it mean for hate speech monitoring and responses?**

03.09.2024

# QUESTION 1

Are there any specific legal measures in place to address hate content within live-stream formats, such as those on TikTok and Twitch, which seem much harder to monitor, document, and regulate compared to other online content?

→ Under the DSA, platforms, including those offering live-streaming services like TikTok, Twitch, and Stripchat are subject to legal obligations to tackle illegal content such as hate speech. Although live-streaming is not specifically singled out, the DSA covers all types of user-generated content.

Platforms must establish mechanisms for users to report illegal content in real-time, which must be handled swiftly and diligently (Article 16). Trusted flaggers, recognized for their expertise in identifying illegal content, receive priority in flagging harmful material, ensuring rapid response, also during live streams (Article 22). Additionally, very large online platforms (VLOPs) with over 45 million users are subject to enhanced duties, including annual risk assessments to assess and mitigate (Article 34 and 35) the spread of illegal content, such as hate speech, that must be specific to the respective service.

While platforms are not required to perform general monitoring (Article 8), they are expected to implement necessary tools, including automated systems, to swiftly remove harmful content once detected or reported. The DSA also incorporates protections from the Audiovisual Media Services Directive, requiring platforms to take steps to prevent incitement to violence or hatred in live-streamed content (Article 28b). Overall, the DSA creates a framework to ensure real-time content, including live streams, is subject to the same rigorous standards for addressing hate speech as other online formats.

# QUESTION 2

Does the DSA foresee providing trusted flaggers and monitors with more adequate funding for their activities? Additionally, are VLOPSES required to disclose staff numbers or work hour estimates for their content moderators, particularly regarding the often AI-based automated content moderation processes used by social media platforms?

→ The Digital Services Act (DSA) does not explicitly mention increased funding for trusted flaggers or monitors, but it does require platforms, particularly Very Large Online Platforms (VLOPs), to prioritise and process reports submitted by trusted flaggers. Under Article 22, trusted flaggers must demonstrate competence, independence, and diligence in identifying illegal content. Trusted flaggers are appointed by the responsible Digital Service Coordinator in each Member State.

Articles 15, 24 and 42 require platforms to ensure transparency about their content moderation practices, including the use of automated decision-making systems such as AI. Additionally, VLOPSEs must provide transparency reports that detail the actions taken to moderate content, including the involvement of automated systems, and the number of moderators for each EU language. The Commission is working on an implementing act to streamline transparency reporting obligations, which should be adopted by the end of the year.

The Commission will work to ensure that monitoring reporters operating for the implementation of the Code of conduct on countering hate speech online will be adequately supported for such work.

# QUESTION 3

**What measures can be taken to address the unequal application of DSA standards by VLOPs across different countries within the EU?**

→ The DSA foresees that all languages and cultural specificities across the EU are taken into account and considered equal. The transparency reports give a first insight into the differences there might be. This is due to a number of factors, such as the majority of the content being in English – even if present in different Member States – so the number of English moderators might be higher than other languages.

Human moderation is just one part of the whole system of mitigation measures that are taken into consideration across the EU, also due to it being context-dependent. A lot of detection is made through automated tools and looking at the scale of content that is produced daily, it would be impossible to have only human moderators to deal and supervise every single content that is published.

The European Commission is actively working with platforms to bridge the gaps with regulated entities in this matter.

Under Article 34, VLOPs are required to conduct annual risk assessments of their platforms, which includes assessing the risk of non-compliance with local laws and content standards across different countries. This ensures that platforms cannot apply uneven standards across member states.

The DSA also obliges VLOPs to a certain transparency standard, with Articles 15, 24 and 42 mandating transparency reports. These reports must provide data about how content moderation, including automated processes, is applied across jurisdictions, ensuring that there is visibility into the consistency of enforcement. If discrepancies are found, authorities are empowered to investigate and impose corrective measures to harmonise practices across the EU.

# QUESTION 4

Could you clarify the role of Trusted Flaggers and whether the criteria for their selection are standardised across all member states?

→ The DSA provides high level requirements: 1. being having specific expertise and competence on monitoring content; 2. being independent from the service provider; and 3. work diligently, accurately and objectively. Besides these three points, the DSA does not have any additional criteria for selection of Trusted Flaggers, under Article 22.

The competence for appointing Trusted Flaggers is held by the Digital Service Coordinators at the national level.

# QUESTION 5

Is the new DSA code finalised and publicly available? If not, what is the expected timeframe for its publication and enforcement?

→ No, the new Code is not yet publicly available. It is expected to be published over the next months.

# QUESTION 6

How does the DSA specifically define "hate speech," and how does it address more subtle forms of radicalising narratives, such as dog whistles and fear speech?

→ The Digital Services Act (DSA) does not provide a specific definition of "hate speech" but refers to illegal content, which includes hate speech as defined by relevant national and EU laws. This generally covers incitement to violence or hatred based on factors such as race, ethnicity, religion, and gender. The DSA requires platforms to remove such content when notified, and it establishes robust mechanisms for identifying and addressing it, including through notice and action systems outlined in Article 16.

For more subtle forms of harmful narratives, such as dog whistles or fear speech, while the DSA does not explicitly reference these forms of speech, it addresses broader systemic risks posed by content that may harm public discourse, public security, mental health or democratic processes.

Under Article 34, platforms, particularly Very Large Online Platforms (VLOPs), must conduct risk assessments on how their services are used to spread both illegal content and content that could contribute to the four categories of systemic risks outlined in Article 34. This includes forms of speech that, while not explicitly illegal, may still have radicalising effects. Platforms are mandated to adopt measures that mitigate the dissemination of such content, ensuring they are proactive in addressing both explicit and subtle threats.

# QUESTION 7

What is the difference between trusted partners and trusted monitors under the DSA?

→ The DSA does not mention trusted partners nor trusted monitors, and neither does the Code of conduct. Trusted flaggers (article 22) under the DSA are vetted independent organisations that have a particular expertise and competence for the purpose of detecting, identifying and notifying illegal content. DSCs appoint Trusted Flaggers.

The implementation of the revised Code of conduct on countering illegal hate speech online will involve a network of organisations, “monitoring reporters”, which will support e.g. the monitoring exercises on the Code and contribute to exchange on trends and developments in the area of hate speech. Monitoring reporters can also include DSA-appointed trusted flaggers, as relevant.

# QUESTION 8

Is the Commission planning any awareness-raising measures?

→ The European Commission is working on the “DSA guardians in schools” initiative. The idea is that, ultimately, every school in the EU will have someone who can tell children what the DSA does for them, how they can report content etc. Once the revised Code of conduct on countering illegal hate speech online will be finally adopted as Code under the Digital Services Act, communication and press activities will be undertaken.

# CONTACT US



**FACING  
FACTS**  
**NETWORK**

Scan here to sign up for our  
**Facing Facts Network**  
Newsletter!



Find out more about policy, training  
and research updates on **hate**  
**crime and hate speech!**

- ➔ [facingfacts.eu](https://facingfacts.eu)
- ➔ [facingfactsonline.eu](https://facingfactsonline.eu)
- ✉ [facingfacts@ceji.org](mailto:facingfacts@ceji.org)
- ⬇ **DSA Policy Brief**  
<https://bit.ly/3zu0cQG>

