

Content moderation in the age of DSA

Policy Brief

12th and 13th January 2023, the Good in tech research network in partnership with the Sciences Po Chair "Digital, Governance and Sovereignty", the McCourt Institute, and the MSH Paris-Saclay with the support of Arcom, the DE FACTO project, Institut Mines-Télécom Business School and the Sciences Po medialab organised and hosted the conference "Content moderation in the age of DSA".

This conference aimed to address all aspects of moderation and to explore the contributions of DSA.

Based on the various roundtables and sessions held by the speakers at this conference, we have drawn up the following policy brief outlining several points of concern regarding the implementation of DSA

I. Introduction

1) Introduction and objectives of the Digital Services Act

The Digital Services Act (DSA) is a legislative act proposed by the European Commission. This act is closely related to another: the Digital Markets Act (DMA). Both were drafted by Margrethe Vestager, Vice-President of the European Commission responsible for promoting a Europe fitting the digital age, and Thierry Breton, European Commissioner for the Internal Market, both members of the von der Leyen Commission.

Presented at the end of 2020 by the European Commission, DSA was definitively voted by the European Parliament in July 2022 and approved by the Council of the EU on 4 October 2022. It was published on 27 October 2022.

DSA will be applicable in February 2024, to the exception of the biggest digital platforms and search engines, which will be affected from 2023 on.

DSA has three key goals:

- **Providing a better protection to consumers and to their fundamental rights online.**
- **Establishing a powerful transparency and a clear accountability framework for online platforms.**
- **Fostering innovation, growth and competitiveness within the single market.**

For this purpose, DSA implements several new obligations and measures:

- **Measures to counter illegal goods, services or content online.**
- **New obligations on traceability of business users.**
- **Effective safeguards for users.**
- **Ban on certain type of targeted adverts on online platforms.**
- **Transparency measures for online platforms.**

2) Why is regulation a primary concern?

Online platforms have become one of the main sources of information and communication between people. In 2022, around 45% of the European population declared having recently used social networks to follow the news¹. In 2016, 75% of people who followed or participated in online debates had witnessed or experienced abuse, threat or hate speech. Almost half of them said that this discouraged them from engaging in online discussions².

In the context of the Covid-19 pandemic and the various containments, a drastic increase in online hate has been observed. Nowadays, around 30% of the European population show a lack of trust in the media (including social networks) and a consequent part of the population is exposed to fake news on a daily basis.³

While moderation policies already exist on many social networks, they are not harmonised. For example, in December 2019, Instagram removed 42% of reported content and Twitter 35.9%.⁴

The reporting process and the criteria used to qualify hateful or illegal content differ according to the systems put in place by each platform. This makes reporting content less intuitive for users; and makes it more difficult to perform an overall assessment of reports of hate content on the Internet and the response to them.

3) Inadequacies of the automatic regulation

Fact-checking and automatic detection of problematic content on online platforms is efficient in most cases. However, this kind of moderation can be inadequate when the problematics are news. For example, during covid, fake news detection has been deficient.

Furthermore, algorithmic content moderation has also many biases and flaws. This moderation can be influenced by stereotypes. This phenomenon of censorship is called The Heckler's Veto. It consists in speech suppression by private mobs. It is particularly targeting vulnerable communities (women, especially journalists; people of color, trans community, dissidents).

II. Issues related to the implementation of DSA.

1) Understanding the concept of free speech on the social media era

Digital transformation has completely altered the environment for speech and made it subject to new logics of financialization, optimization and probabilistic governance.

Delimiting and regulate free speech presents several difficulties:

- Free speech on digital platforms implies a possible anonymity and gives anonymous speakers an unprecedented influence. It facilitates unaccountable speech.

¹ 2022 Eurobarometer Survey <https://www.europarl.europa.eu/news/en/press-room/20220704IPR34401/eu-citizens-trust-traditional-media-most-new-eurobarometer-survey-finds>

² 2016 Eurobarometer Survey https://ec.europa.eu/information_society/newsroom/image/document/2016-47/sp452-summary_en_19666.pdf

³ 2022 Eurobarometer Survey <https://www.europarl.europa.eu/news/en/press-room/20220704IPR34401/eu-citizens-trust-traditional-media-most-new-eurobarometer-survey-finds>

⁴ CNDCH, Avis sur la lutte contre la haine en ligne, 8 juillet 2021

- Online speech generates a confusion between the private and the public sphere. Relevant companies that have become central players in the process of digital transformation have failed to achieve putting democratic values and fundamental rights.
- Speech suppression by private mobs particularly targets vulnerable communities (women, people of colour, trans community, dissidents).

2) Allocating regulation power between public and private actors

Due to their format, online platforms, cannot be regulated as simple medias. However there still are issues related to the required regulation:

- It is complicated to find what regulation is required by, and compatible with online free speech.
- The repartition of the regulation burden between the public and private actors is a complicated issue. Platforms take huge number of decisions for content moderation.

III. Proposals for a good implementation of DSA.

If DSA is a good place to start, it is at this point DSA is more a process than an outcome. To ensure an effective implementation of the DSA that innovates the way to regulate on online platforms, three points are essential:

- **Clearly defining the goals of DSA**

DSA is a first step in the evolution of regulation. It constitutes a toolbox in online speech regulation but its concrete implementation has to be watched.

The main goals of DSA should not consist in radically transforming everything in online moderation, it only should aim to make the platforms more accountable.

To get improvements in the field of transparency and disinformation, these concepts must be defined and prioritised. The transparency the European Union will get depends on our goal, on the chosen angle to find an angle.

In the same way, a great balance between free speech and regulation must be subject of an agreement as freedom paradigm is no longer sufficient to handle internet issues.

In the same way, proper indicators and definitions of what is free speech, what is hateful speech (including disparate impacts and experiences across different groups/communities); would be necessary to set a harmonised regulation.

The institutions and the platform must work together to determine what systemic problem are they first aiming to solve at a larger scale. The question of what kind of content should be made more transparent, must be addressed in a concrete way.

For now, DSA only targets bigger players. However, we must remain vigilant so that regulation issues do not relocate elsewhere, on underregulated or unregulated platforms.

- **Data access to researchers:**

Research is key in regulation issues. An increased access for researchers to key data of the largest platforms and search engines would allow them to understand how online risks evolve. A lot of data that are not sensitive are however currently not accessible. For example, one piece of data which is still not available and could be really useful for research is the number of persons reached by a Twitter or a Facebook publication.

DSA must create a better cooperation between platforms and researchers. Furthermore, researchers must be supported by more European funds relayed by national ones.

- **Working within a relevant institutional framework:**

Article 49 of DSA sets a duty for the platforms to collaborate with the regulators and the implementation of a digital services coordinator.

However, in each European country there are at least two or three regulatory authorities (in France CNIL, ARCOM and AMF for example). We need to create a framework to effectively articulate all these authorities.

The creation of specific institutions could be effective to guarantee the respect of this part of DSA and to generate a visible and clear cooperation environment with private actors.