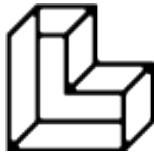European Partnership for Democracy

CIVIL LIBERTIES UNION FOR EUROPE

Civic Discourse and Electoral Processes

in the Risk Assessment and Mitigation Measures Reports under the Digital Services Act

An Analysis

Orsolya Reich and Sofia Calabrese

MARCH 2025

Digital Services Act

Table of Contents

Introduction

The EU's <u>Digital Services Act</u> (DSA) mandates Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) to conduct thorough **assessments and implement mitigation measures for systemic risks** that the use of their services pose, among others, to **civic discourse and electoral processes** (Articles 34 and 35).

The first risk assessment and mitigation measures **reports for 2023 and 2024 have been made public in November 2024**. This provides an opportunity for civil society and other experts to obtain more information on the measures adopted by VLOPs and VLOSEs and to be involved in improving the effectiveness of such measures in future reports.

This paper aims to **analyse the risk assessments to identify relevant gaps** and suggest steps on how to ensure more robust protection of civic discourse and electoral processes under the DSA. We will provide first a general analysis of the risk assessments and then focus on individual reports, namely those of Facebook, Instagram, Google Search, YouTube, TikTok, and X.

1. General observations on the risk assessments

We welcome the effort that platforms have put in place to **develop methodologies to identify risks** and to map relevant mitigation measures that have been adopted in the past years. We also welcome the **transparency of this exercise** as we understand that the information shared in the public reports is very close to what was shared with the European Commission. This will be very helpful for civil society and researchers to give relevant input to the reports to make improvements and adopt more effective mitigation measures.

At the same time, there are specific areas where we would encourage more action in future reports and possibly intervention from the European Commission with specific guidelines. In particular:

- **Format and consistency among the reports**: Each report follows a very different logic and structure, which is understandable given that they were developed separately and without specific guidance from the Commission. The absence of predefined guidelines may have helped uncover issues that a rigid structure could have overlooked. However, now that a diverse range of risk assessments has been conducted, the Commission has the understanding and expertise

to determine the structure, logic, and data requirements that would best serve its needs. Establishing a more consistent framework would enhance comparability thereby providing significant benefits not only to the Commission but also to the research community and watchdogs.

- **Lack of data on the effectiveness of risk mitigation measures**: Article 34 of the DSA requires that platforms put in place 'reasonable, proportionate and effective mitigation measures'. The reports, however, provide very little detail on the nature and impact of the mitigation measures, which makes it virtually impossible to assess whether they are actually reasonable, proportionate and effective. To evaluate compliance, the Commission would then have to send requests for information, hence defeating the purpose of the reports.

- **Details and composition of the teams involved**: While the teams involved are always mentioned, often including specific teams focused on civic space and elections, there is little detail about their size, roles and responsibilities. Without this information, it is difficult to determine whether these teams have adequate resources and expertise to fulfil their mandate.

- **Stakeholder consultation**: The reports mention that stakeholders have been consulted extensively, but the actions mentioned that have been taken to ensure consultation seem relatively narrow and are often limited to a few workshops in which both platforms and civil society were present. In fact, very few civil society organisations working on the area were consulted individually, even when reaching out proactively, especially in the area of elections and civic discourse.

- **Recommender systems**: Recommender systems play a crucial role in shaping civic discourse online by selecting the content that will be displayed to users and amplifying specific content over others. The functioning of recommender systems and criteria behind the choice of content to be displayed and amplification of content are not sufficiently explained. While there were expectations that these reports would clarify the role of recommender systems, they contain minimal information on risks related to their functioning and on their role in shaping civic discourse online, and even less information on related mitigation measures.

When it comes specifically to **civic discourse and electoral processes**, the understanding of such issues is often superficial, with a narrow focus on elections rather than fostering healthy political discourse online. While there is significant attention on disinformation, other critical aspects which we had pointed out in our report on **Identifying, Analysing, Assessing and Mitigating Potential Negative Effects On Civic Discourse And Electoral Processes**, are largely overlooked, such as the identification of political ads, the over-removal or shadow banning of political content,

the role of influencers, inclusivity, incivility, echo chambers, polarization, and the exacerbation of conflict situations. Additionally, issues such as organized campaigns against civil society, asymmetric amplification of political content from different electoral contenders, and the spread of false information regarding voting processes are insufficiently addressed, leaving significant gaps in the analysis.

2. Analysis of individual platforms' assessments

In this section, we analyse the reports published by the six platforms that are the most relevant for civic discourse and electoral processes in the EU: Facebook, Instagram, Google Search, YouTube TikTok and X.

a. Meta

Meta's systemic risk assessments for [Facebook](#) and [Instagram](#) show a **highly structured but often hard-to-navigate approach to evaluating risks**. While the company appears to have committed substantial resources to this reporting process, the overall impression is that the reports reflect a reluctance to go beyond a box-ticking compliance exercise.

Rather than taking a fresh perspective on systemic risks, the framework used in the reports seems largely anchored in Meta's existing internal policies. The risk identification and mitigation strategies rely heavily on Meta's Community Standards, leading to a **narrow focus on policy violations rather than broader, platform-driven harms**.

One particularly striking issue is the **lack of differentiation between the Facebook and Instagram assessments**. Despite significant differences in user demographics, engagement patterns, and types of content shared on the two platforms, the reports are nearly identical. This raises concerns about whether Instagram's distinct risk profile has been adequately addressed.

i. Facebook
Facebook's systemic risk assessment is structured around Meta's internal policy framework, identifying the seven systemic risks mandated by the DSA while adding an additional category related to deceptive and misleading content. Within this structure, the report outlines 19 problem areas, generating 122 distinct risks. The assessment also incorporates five influencing factors, as required under Article 34.2 of the DSA.

Systemic risks to Civic Discourse and Elections are identified across 11 problem areas. Like other companies, Meta has made **no effort to facilitate oversight of its adherence to the Commission's guidelines** or its specific measures related to the 2024 elections. Although Section 6.2.2 (Detailed Risk Observations and Mitigating Measures) frequently references the 2024 European Parliament elections, its analysis follows Meta's own internal logic, making it difficult to assess compliance in a meaningful way.

While the report outlines certain risks and corresponding mitigation efforts, most **descriptions remain vague, presenting actions in broad, generalised terms** (e.g., "the majority of full-time employees are required to take a Civil Rights and Meta Technologies training to help identify civil rights risks"). Crucially, it **lacks concrete details**, such as data or figures, that would enable external stakeholders—including the Commission, Digital Services Coordinators (DSCs), watchdogs, and researchers—to evaluate the effectiveness of these measures.

We acknowledge that not all risks and mitigation efforts can be easily quantified and that reports must strike a balance between conciseness and substance. However, in its current form, the report offers **little meaningful transparency** or compelling evidence that Facebook is taking substantive action to mitigate systemic risks and safeguard societies from significant harms.

Instead, much of what is presented appears to be a formalisation of existing risk management practices rather than a proactive response to the systemic challenges outlined by the DSA.

One of the most striking omissions is the **limited discussion of recommender systems** and their role in amplifying or suppressing content. Despite well-documented concerns about engagement-driven ranking fuelling polarisation, misinformation, and harmful civic discourse, the report provides little meaningful analysis beyond stating that certain algorithmic adjustments have been made.

The discussion on political discourse and election integrity is similarly narrow, primarily **addressing disinformation while overlooking broader issues**. Additionally, while the report acknowledges risks associated with policy enforcement—such as over-moderation or inconsistent rule application—it lacks transparency on how moderation decisions are

made at scale or how biases in automated enforcement are addressed. **Engagement with civil society appears minimal**, with no substantial evidence that external stakeholders were meaningfully consulted beyond a few workshops.

There is also **little information on benchmarks** used to determine the severity of risks, how mitigation measures were selected and tested, or whether they have been effective.
Overall, while the document may **offer useful insights for identifying key questions and justifying data access requests for researchers**, it provides little more. It falls short of delivering meaningful transparency or assurance that Facebook is genuinely committed to mitigating systemic risks.

ii. Instagram
Despite the fundamental differences between Instagram and Facebook in terms of user base, content format, and engagement models, Instagram's risk assessment is **largely a replica of Facebook's, with only minor platform-specific adjustments**. Unlike Facebook, which revolves around public posts, groups, and link-based content, Instagram's risks stem from the dominance of visual content, the influence of creators and celebrities, and the amplification of aesthetics-driven engagement. The platform also skews toward a younger audience, introducing unique vulnerabilities. These distinct features, however, receive little attention in the assessment. The failure to meaningfully differentiate Instagram's risk assessment from Facebook's represents a **missed opportunity to critically examine how platform-specific dynamics** contribute to systemic online harms.

b. Google

Google's assessment report puts forward a general part in which it explains some common best practices that are used across services, the **six-steps methodology** used to conduct the systemic risk assessment and then dives into its various specific platforms and search engines, including [Google Search](#) and [YouTube](#), assessing risks individually and outlining related mitigation measures.

Overall the report "found that risks associated with highly motivated bad actors seeking to misuse our services remain a cause of concern" including risks for civic space, but it seems to identify risks mostly when related to **misuse of the services rather than systemic** to the services themselves. Furthermore, Google focuses exclusively on elections and

not civic discourse despite the obligation rooted in the DSA.

The report also shows a **lack of meaningful stakeholder engagement**, especially when it comes to civil society, where very few actions other than the Global Network Initiative (GNI) are mentioned that relate to engagement with civil society on the development of the methodology.

On the **teams involved**, the report explains that there are specific functions in each of the services provided, in addition to cross-service teams such as Trust & Safety, Human Rights, and Civics[1], the latter in charge of initiatives to safeguard the integrity of elections. Having a specific Civics Team goes further than the efforts outlined in other reports, but it still only focuses on elections and fails to specify the role, composition or contributions of the civics team in detail.

i. Google Search
Google Search's section, when it comes to electoral processes, **focuses narrowly on disinformation**, particularly election-related disinformation, with initiatives such as the Code of Practice on Disinformation being highlighted. The report also mentions as part of mitigation measures that Google put in place "features that allow reliable information," but it does not provide specifics about which features these are or how they function.

The report also completely misses other issues linked to civic discourse which we had pointed out in our report on Identifying, Analysing, Assessing and Mitigating Potential Negative Effects On Civic Discourse And Electoral Processes.

Finally, as highlighted for the reports in general, there is **little exploration of how the search algorithm selects information** to display, leaving critical details about its mechanisms and criteria unaddressed.

ii. YouTube
As for Google Search, the approach to YouTube also shows a **narrow focus on civic discourse**, primarily centered on disinformation and election disinformation, including voter suppression. Measures such as demonetization are mentioned as mitigation as well as tuning

---

[1] *"Our Civics team works across our services, addressing threats to democratic participation in partnership with Trust and Safety specialists. The Civics team oversees products, initiatives, and promotional efforts that aim to safeguard the integrity of elections-related information and provide users with candidate information from authoritative sources. These teams also provide 24/7 support to triage emergent issues during elections."*

recommender systems so that they raise visibility of authoritative content. The report also takes into account risks regarding addiction, but addiction-related measures appear to focus exclusively on children.

Finally, as for many other reports, there is very **little information provided about recommender systems** or the mitigation measures in place to address their potential risks to civic discourse and electoral processes. Recommender systems are only mentioned as related to civic discourse when explaining that they tend to emphasize authority in results in situations in which there is breaking news. There is no discussion of related issues either, such as shadow banning or its implications for political content.

c. TikTok

TikTok's report outlines the risk assessment methodology (in Annex), an overview of the risk environment with a summary of the risk assessment results and then dives into specific risk areas, including civic discourse and electoral processes. As for other reports, however, there is **insufficient detail** regarding the methodology, teams involved, and other critical processes such as stakeholder consultation.

In the initial summary, risks to civic discourse and electoral processes are categorized as Tier 1, meaning that they are at the top priority according to their assessment having taken account of existing policies, systems and procedures for mitigating the risk. Their impact, however, is then only described as moderate, creating ambiguity about prioritization. Furthermore, for electoral processes, **only risks linked to election disinformation are identified** (as opposed to 6 different risks identified in our report, on top of disinformation in general), while the report completely **fails to adequately consider additional risks to civic discourse**. Additionally, there has been limited action on identifying political ads, with efforts mainly focused on fact-checking and labeling rather than comprehensive measures to enforce the prohibition of political ads.

TikTok's report does take into account **coordinated inauthentic behaviours** as a risk and puts forward mitigation measures to address them, such as monitoring cross platform threats, leveraging internal detection signals, and manually investigating reports from users and trusted flaggers. TikTok also offers political account verification and labels for state-affiliated media. The report, however, fails to recognise the role of **influencers** in political campaigns and risks associated with that.

In general there seems to be a lack of thorough investigation into specific categories, appearing instead to repurpose existing work—for example, using disinformation-related actions to address both electoral processes and public security.

d. X

X's systemic risk assessment stands out for its **explicit focus on content moderation risks, particularly to freedom of expression**. X frames content restrictions themselves as systemic risks rather than necessary safeguards against harm.

A notable strength of the report is its **clarity and empirical grounding**. It references research to justify decisions, even if the quality of the studies cited is open to debate. For instance, while X's reliance on Community Notes as the primary tool to counter misinformation is questionable, the report **at least attempts to provide a rationale** for this approach.

Nevertheless, X's assessment of **risks related to civic discourse and electoral processes** is **strikingly limited.** The report downplays the platform's potential influence on elections, arguing that causation is difficult to establish. It also neglects many of the risks highlighted in the Commission's election guidelines and independent research, **focusing narrowly on misinformation, political advertising, and inauthentic behaviour.** Similarly, X's response to concerns about content moderation bias is vague, stating that conclusive research is lacking but that further studies are planned. This non-committal stance suggests a reluctance to engage in meaningful reforms unless compelled by external regulatory pressure.

Despite these shortcomings, the report's structure is refreshingly clear. By adhering to the DSA framework and consolidating risk assessments and mitigation measures related to electoral integrity and civic discourse into a single section, it provides an **accessible overview of X's efforts** in these areas. This structured approach improves readability and facilitates a more straightforward evaluation of the platform's stated commitments and actions. Another positive aspect is the inclusion of **numerical data**; however, the lack of access to these figures for watchdogs raises concerns about transparency.

While the report employs a quantitative approach—assigning severity

and probability scores to risks—it **lacks transparency** on how mitigation efforts reduce these risks in practice. The **methodology behind the scoring remains unclear**, with no indication of independent validation. X does introduce a breakdown of risk severity into scope, scale, and remediability, which offers some insight into prioritisation. However, it is uncertain whether these assessments genuinely reflect platform realities or merely reinforce existing policy positions.

A key weakness of the report is its **reluctance to engage with the broader societal consequences of platform risks**. While acknowledging algorithmic risks, including recommender systems' potential to create echo chambers, the report does not meaningfully examine how these mechanisms shape civic discourse, political engagement, or the spread of harmful content. Instead, X argues that the effects of its algorithms vary across contexts, making biases difficult to assess. Given the well-documented role of engagement-driven ranking in polarisation, this lack of substantive analysis is concerning.

3. Conclusions and recommendations

1. There is a need for more guidance from the European Commission to harmonise the reports and facilitate comparative studies.

The lack of methodological consistency across platform risk assessments highlights the need for clearer regulatory guidance from the European Commission. While the DSA mandates systemic risk evaluations, each platform has adopted different frameworks, metrics, and interpretations, making it difficult to compare risks and assess compliance effectively.

Recommendations

The Commission should provide detailed guidance on risk assessment methodologies, including:

- standardised risk categories (while recognising that risks differ based on audience demographics, engagement patterns, and platform functionalities);
- minimum transparency requirements (including numerical data), preventing excessive redactions;
- specific requirements for assessing the role of recommender systems in amplifying harmful content.

2. Platforms should increase transparency on size of the teams involved and resources dedicated to risk assessment and mitigation under the DSA.

The lack of details regarding the size and expertise of internal teams responsible for risk evaluation and mitigation, and decision-making processes regarding the mitigation measures raises accountability concerns. Without a clear understanding of who is responsible for implementing risk mitigation efforts, it is difficult to evaluate whether platforms are dedicating sufficient resources to systemic risk management.

Recommendations

Platforms should disclose more details about the teams responsible for risk assessment, including:

- number of personnel assigned to systemic risk evaluation;

- breakdown of expertise (e.g., policy specialists, data scientists, legal experts, external consultants, etc.);

- governance structures and oversight mechanisms.

3. More meaningful and systematic stakeholder engagement is needed to strengthen the risk assessments.

Many platforms claim to have engaged with stakeholders, yet consultations appear superficial, often limited to a few workshops attended by both platforms and civil society representatives. There is little evidence that independent researchers, advocacy groups, or marginalised communities have been consulted meaningfully on issues related to political content moderation and civic discourse.

Recommendations

VLOPs and VLOSEs should increase the depth and frequency of stakeholder consultations, ensuring that:

- civil society organisations, independent researchers, and affected communities are meaningfully engaged and not just in broad workshops;

- consultation findings are published transparently, detailing how feedback has been integrated into risk mitigation strategies.

The European Commission should establish minimum stakeholder consultation standards.

4. Risk assessments should have more details on the role of recommender systems and algorithmic amplification in affecting civic discourse and elections.

Recommender systems play a crucial role in shaping online discourse, influencing the visibility of political content, disinformation, and extremist narratives. However, most platform risk assessments fail to provide meaningful insights into how these systems operate, what risks they introduce, and what mitigation measures are in place.

Recommendations

Platforms should be required to:

- explain how recommender systems prioritise, demote, and filter content;

- provide data on the impact of algorithmic changes on political discourse.

5. Risk assessments should expand the scope and go beyond elections and disinformation.

Most platform risk assessments focus narrowly on election integrity and disinformation, neglecting broader risks to civic discourse, political participation, and social cohesion. While disinformation is an important issue, other systemic harms—including the suppression of political content, echo chambers, influencer-driven disinformation, and the weaponisation of online platforms against civil society—are insufficiently

addressed.

Recommendations

Platforms should broaden their risk assessments to include all aspects of civic discourse, not just elections and disinformation.

European
Partnership for
Democracy

CIVIL
LIBERTIES
UNION FOR
EUROPE