



Five Critical Lessons from Three Years of DSA Risk Assessments

About ECNL

The European Center for Not-for-Profit Law (ECNL) is a civil society organisation based in the Hague, the Netherlands. Driven by a mission to collaborate with individuals and groups in taking collective action to address societal challenges, we envision a world where every person has the freedom, power, and opportunity to shape a just, inclusive, and sustainable future. We are a long-standing member of the DSA Civil Society Coordination Group as well as European Digital Rights. In 2023, in collaboration with Access Now, we developed [recommendations for conducting meaningful risk assessments](#) focused on the systemic risks of negative effects on fundamental rights, grounded in DSA requirements and best practices of conducting human rights impact assessments under the UN Guiding Principles for Business and Human Rights. In 2025, we published an updated version of the [Framework for Meaningful Engagement in AI](#), which outlines recommendations for AI developers and deployers on engaging with civil society and affected communities.

About this policy brief

To contribute to the evaluation of the implementation of Article 34 of the DSA and inform efforts to provide further guidance to platforms, ECNL undertook a **review of three rounds of risk assessments (2023-2025)** based on our recommendations, thereby testing out our guidance in practice. We focused on **five platforms** that are, in our view, the most likely to affect civic space, civic freedoms and online discourse, given their size, influence, and the nature of their products: Facebook, Instagram, TikTok, X, and YouTube. Additionally, we zoomed in on the **negative effects on fundamental rights under Article 34(1)b DSA**. Our review of the first round of risk assessments informed the initial [analysis](#) published in March 2025 by the DSA Civil Society Coordination Group. With this policy brief, **we distil our research into five lessons learned** from the risk assessment process following the first three rounds. We close this brief with conclusions and recommendations.



Key takeaways

Our analysis revealed a significant gap: although each risk assessment report spans over 100 pages, we found that they lacked sufficient specificity to answer some of the central questions about platforms' negative effects on fundamental rights under Article 34(1)b DSA. **While the reports in years 2 and 3 improved in terms of structure and readability, we did not see meaningful improvement in terms of their content.**

Under the DSA, risk assessments are intended to provide much-needed transparency regarding platforms' decision-making processes and efforts to identify and address systemic risks. They would enable civil society to conduct independent analyses, inform the European Commission's oversight mandate, and enhance accountability. Unfortunately, **in their current form, risk assessments are unlikely to provide external organisations with meaningful insights into platforms' risk mitigation measures or to contribute substantively to related discussions.**

The overarching gaps we identified are:

1. the lack of detail regarding the actual assessment, i.e. how risks were conceptualised, identified, and how severity and likelihood were assessed for each risk;
2. insufficient diligence in assessing the systemic risk of negative effects on fundamental rights as well as the fundamental rights effects of mitigation measures themselves;
3. the lack of relevant metrics, data and evidence for statements and claims presented in the reports;
4. the lack of clarity as to how stakeholder engagement informed the results of the assessment;
5. insufficient consideration of the EU's geographic and linguistic diversity.

This raises the following critical questions: How should risk assessments be improved? Can they serve as valuable tools for the Commission and civil society — whose important role in the contribution to the enforcement of the DSA is recognised by the European Board for Digital Services¹— or are they mere 'check box' exercises characterised by unsubstantiated claims?

¹ First report of the European Board for Digital Services in cooperation with the Commission on the most prominent and recurrent risks as well as mitigation measures, <https://digital-strategy.ec.europa.eu/en/news/press-statement-european-board-digital-services-following-its-16th-meeting>

Lesson 1: The assessment must be specific to individual risks, not composed of vague declarations.

All risk assessments we analysed contain broad statements about how systemic risks were identified and evaluated without providing evidence that can substantiate those statements. Platforms generally list all the systemic risk categories or taxonomies they rely on, as well as the criteria used for assessing severity, which is good practice. However, **they fail to explain how these categories were defined in the first place, what constitutes an acceptable risk threshold, or what metrics and benchmarks were used to assess severity and likelihood for each identified risk.** These are precisely the elements that would allow the Commission or civil society to evaluate whether risks were properly identified and addressed. This lack of specificity significantly undermines the ability to assess the quality of the assessments and to determine whether they comply with the EU Charter of Fundamental Rights, the DSA, and other relevant standards, including international human rights law.

For example:

- YouTube stated in their 2025 report that they **reviewed risk statements against all articles in the Charter and international human rights instruments.**² While we applaud this practice, we cannot verify this claim because YouTube **does not provide the details of this review.** The full list of risk statements in Annex A is not systematically aligned with all Charter rights; instead, only selected rights are referenced, making it unclear which rights correspond to each risk statement. It is difficult to determine from the report which rights were affected and in what way. This is partly because the report focuses on mitigation measures rather than assessing individual risks, and does not include a dedicated section on fundamental rights.
- With respect to negative effects on fundamental rights, TikTok states that it “has assessed residual risk to be Medium in Year 3.”³ However, the platform provides **no substantive information on the underlying assessment,** preventing reviewers from understanding—let alone verifying—whether the severity and likelihood of the risks were assessed appropriately.
- Similarly, X describes its risk methodology and tiering system, e.g., “inherent risk is a function of probability and severity”⁴ and considers “critical or high residual risk areas to be Tier 1 risks, medium residual risk areas to be Tier 2 risks, and low or negligible residual risk areas to be Tier 3 risks.”⁵ However, it **does not disclose quantitative or qualitative thresholds or criteria for each tier,** limiting external verification.

² YouTube’s 2025 risk assessment report, page 54.

³ TikTok’s 2025 risk assessment report, p. 81.

⁴ X’s 2024 risk assessment report, p. 16.

⁵ X’s 2024 risk assessment report, p. 18.

- Facebook states that negative effects of upload filters can “manifest on Facebook through over-enforcement of non-policy-violating content, disproportionate enforcement of policy-violating content, language/dialect limitations of human reviewers or classifiers, failure to take down policy-violating content, activity that limits or discourages a user’s freedom of expression, and the inherent difficulty in balancing freedom of expression and safety concerns.”⁶ **Yet claims that effects *might* manifest are not sufficient for the Commission or civil society to evaluate the assessment’s results.** Platforms should also report whether and to what extent the negative effects actually occurred in the assessed year, in addition to potential negative effects.

Lesson 2: Platforms must apply the same level of diligence when assessing negative effects on fundamental rights as they do when evaluating other systemic risks.

Our analysis demonstrates that platforms failed to adequately identify and assess all the potential and actual negative effects on fundamental rights of their products and services, even though this category of systemic risks (Article 34(1)b) DSA) has the same legal weight as other systemic risks under Article 34.

In the first round of risk assessments, most of the reports did not even include a dedicated section on fundamental rights (as opposed to other types of systemic risks), opting instead to distribute fragmented and vague information throughout other sections. This has improved for most examined platforms in the second and third round. Except for YouTube, all the analysed platforms now include a dedicated section addressing fundamental rights risks, which improves scrutiny and readability.

However, **none of the risk assessments we examined covered the full range of rights protected under the EU Charter, despite the DSA not limiting its requirements to specific rights.** In our view, the risk assessment should, at minimum, provide a comprehensive list of all fundamental rights that platforms have identified as potentially affected, describe the nature of these effects (including how they materialised in the reporting period and how they may arise in the upcoming year), and indicate their severity and likelihood, supported by appropriate evidence. This would enable reviewers to determine which fundamental rights are most at risk. Instead, **platforms generally assessed only a limited subset of rights**, typically focusing on the most visible or least contested ones, rather than conducting the comprehensive analysis required under the DSA. Where platforms have already carried out other assessments that might overlap with DSA requirements (e.g. under

⁶ Facebook’s 2025 risk assessment report, p. 93.

the GDPR or broader human rights due diligence processes), they should at minimum integrate the relevant findings into their DSA reports.

Examples of inadequate approach:

- **Platforms failed to assess each fundamental right individually and to prioritise negative effects based on their severity;** instead, they grouped them together. For example, in their 2025 report, TikTok assessed the residual risk for fundamental rights as “medium,” but we do not know which of the rights are most affected and how.⁷ The platform lists the five most relevant fundamental rights, but does not provide any insight into how this relevance was established, what negative effects look like for each of these rights, how severe and likely the effect on each of these rights is, and whether this restriction of rights is justified.⁸
- **No platform meaningfully assessed the effects of their systems on the right to privacy and data protection.** This is important considering that platforms collect vast amounts of personal data for advertising or content curation purposes and processing this data through algorithmic systems. YouTube, for example, limits such considerations to access to data by third parties, minimising privacy and data protection risks, rather than assessing how their own advertising systems and profiling and targeting practices affect fundamental rights.⁹
- Facebook’s 2025 report refers to the company’s broader human rights due diligence processes, without summarising the results of these assessments, or explaining which of the risks are specifically relevant in the EU and DSA context, and **how exactly the findings fed into the risk assessment and development of mitigation measures.**¹⁰

Platforms also **failed to meaningfully assess the fundamental rights effects of their mitigation measures** (e.g. automated content moderation systems). In the third risk assessment, platforms increasingly acknowledge that measures implemented to mitigate one risk (e.g. proactive, automated content moderation) might inadvertently have negative effects on fundamental rights. However, none of the risk assessment reports we examined contain a rigorous analysis of this issue, despite the clear requirement in Article 35(1) DSA.

For example:

- While Instagram and Facebook include sections on risks related to their content moderation systems, their **descriptions are vague and fall short of a**

⁷ TikTok’s 2025 risk assessment report, p. 81.

⁸ TikTok’s 2025 risk assessment report, p. 78.

⁹ “While advertising makes YouTube free of charge for everyone, we do not sell personal information to anyone. YouTube’s data practices turn a significant inherent risk into a much lower residual risk”. YouTube’s 2025 risk assessment report, p. 131.

¹⁰ Facebook’s 2025 risk assessment report p. 11 (“Meta is committed to respecting the fundamental rights of our users located in the EU as outlined in the DSA. Our third annual Human Rights Report demonstrates how we are living up to the commitments made in our Corporate Human Rights Policy”) and further on p. 24 (“These approaches inform and are complementary to our DSA regulatory compliance work.”).

proper assessment. Instagram’s 2025 report notes that content moderation risks “can manifest on Instagram through over-enforcement of non-policy-violating content, disproportionate enforcement of policy-violating content, language/dialect limitations of human reviewers or classifiers, failure to take down policy-violating content, activity that limits or discourages a users’ freedom of expression, and the inherent difficulty in balancing freedom of expression and safety concerns.”¹¹ For this assessment to be meaningful, Instagram should specify whether and to what extent these risks have materialised in the past years, reference objective and relevant metrics, and demonstrate whether mitigation measures adopted are effective, supported by concrete data.

- YouTube’s transparency report for the first half of 2025 shows that out of 483,750 user complaints, 247,357 moderation decisions were reversed.¹² This means that appealed content moderation decisions were incorrect in more than 50% of cases. TikTok’s transparency report for the same period shows that out of over 3 million complaints, around 1.3 million pieces of content were reinstated,¹³ meaning that 43% of the appealed moderation decisions were incorrect, while the platform reports over 98% accuracy of their automated content moderation systems across all EU languages. While we do not know if the content subject to appeal was originally removed automatically or not, neither of the platforms presented a genuine reflection in the risk assessment report as to why there is such a high error rate, how this compares to the reported accuracy rate of automated moderation systems, and what effects it might have on freedom of expression and other fundamental rights.

Lesson 3: For meaningful scrutiny, we need relevant information that substantiates platforms’ claims.

In most cases, platforms fail to provide relevant data or comprehensive information to support their claims regarding the scale of identified risks or the effectiveness of mitigation measures. This ranges from insufficient evidence justifying why individual risks were assigned to specific risk tiers, to providing incomplete metrics or information that do not capture all relevant aspects of the risks. In addition, platforms do not share any information about the algorithms used for purposes other than automated content moderation, even though Article 34(2) DSA requires them to consider how systems used for advertising or recommending content affect systemic risks.

¹¹ Instagram’s 2025 risk assessment report, p. 91.

¹² Google’s VLOP/VLOSE Transparency Report for January - June 2025, p. 42-43.

¹³ TikTok’s VLOP Transparency Report for January - June 2025, p. 25.

The following examples illustrate the insufficient or irrelevant metrics:

- In relation to DSA’s Article 15(1)(e) requirement to share the accuracy and the error rate of automated content moderation, Instagram presents the **automation overturn rate** as a key metric in its transparency report for the first half of 2025. They state that "the automation overturn rate only captures content removed via automation that was later restored. While not all restores are errors and not all errors are restored, the metric still is a directionally approximate indicator of accuracy."¹⁴ In our opinion, for reviewers to fully understand the effectiveness and risks of automated moderation, **Instagram should explain why “not all restores are errors” (what other reason than error could there be for restoration of content) and what is the percentage of errors not restored and why.** The automation overturn rate also focuses only on false positives, and Instagram does not expand on whether and how it accounts for false negatives (under-enforcement of content policies), as these have implications for fundamental rights and user safety, too.
- To substantiate the effectiveness of mitigation measures for preventing the dissemination of illegal content, TikTok shares the total number of proactive content removals.¹⁵ However, **this information alone does not demonstrate the effectiveness of these measures because it fails to address key questions** such as (a) the overall extent of illegal content and how the volume of proactive content removals compares to it; specifically, whether there is any assessment of the total amount of illegal content and, within that total, how much is removed proactively versus how much is removed following third-party notices; (b) whether these removals were justified in the first place and did not violate the right to freedom of expression (false positives); and (c) how much illegal or violative content went undetected (false negatives).
- X claims that their measures are “robust and proactive,” yet the only quantitative information provided relates to the accuracy of the “restricted reach” label. X states that between July 2024 and June 2025, approximately 2% of posts receiving this label were appealed and 1% of those appeals were overturned, which it interprets as indicating roughly 99% accuracy.¹⁶ However, this metric is **limited to a single labeling mechanism and does not demonstrate the overall effectiveness of systems designed to prevent the dissemination of illegal content.** Moreover, no data is provided on automated detection or upload filtering performance, broader moderation error rates, or the effectiveness of proactive prevention measures. Accordingly, the cited statistic provides only a narrow indicator and does not substantiate the broader claim of effectively mitigating the risk of disseminating illegal content.

¹⁴ Instagram’s VLOP Transparency Report for January - June 2025, p. 23.

¹⁵ “The rate of proactive removals is a positive indicator of the efficacy of TikTok’s content

moderation systems and processes.” TikTok’s 2025 risk assessment report, p. 26, 31, 41, 46, 50, 59, 64, 71, 77, 82.

¹⁶ X’s 2025 risk assessment report, p. 35-36.

Lesson 4: Risk assessments should explain how platforms engaged external stakeholders, what feedback platforms received during consultations, and how it shaped their response.

Despite the DSA's explicit reference to consultation with experts and affected groups under Recital 90, platforms failed to engage meaningfully with civil society in all the three rounds of risk assessments examined. This lack of engagement is particularly concerning given that civil society organisations and academic experts in digital rights could provide valuable insights for the risk assessments and inform platforms' understanding of the effects of their products, services, and operations.

We see four primary issues across various platforms:

- **Platforms did not adequately consult stakeholders representing diverse perspectives.** For example, X listed a small number of civil society organizations consulted, but the engagement process was largely dominated by governments and law enforcement authorities.¹⁷
- **Platforms did not provide a comprehensive account of the feedback received during consultations or how it was addressed, restricting their reporting to broad discussion topics or a few examples.** For example, Facebook only mentions areas in which feedback was collected, e.g. minors or women's safety, without specifying key comments made by external stakeholders or how the platform considered them.¹⁸
- **Platforms did not explain how their broader stakeholder engagement feeds into assessing risks in the EU and preparing DSA-specific mitigation measures.** Many platforms cited events they participated in, but did not disclose how exactly these events contributed to the risk assessment. For example, TikTok mentions participation in RightsCon but does not explain what insights it took away from this event. Many platforms also cite the GNI and DTSP European Rights & Risks Forum but do not explain how they responded to the recommendations made by civil society during the event, despite those recommendations being publicly available.¹⁹
- **Platforms failed to incorporate feedback from civil society groups and academic researchers who had previously raised criticisms, and offered no explanation for why such input was disregarded in their assessments.** For example, research and recommendations by Amnesty Tech²⁰, Panoptykon

17 X's 2025 risk assessment report, p. 17.

18 Facebook's 2025 risk assessment report, p. 16.

19 Global Network Initiative, Summary Report: 2025 European Rights and Risks Stakeholder Engagement Forum, <https://globalnetworkinitiative.org/new-report-2025-european-rights-risks-stakeholder-engagement-forum/>

20 Amnesty Tech's research on TikTok's "ForYou" feed's impact on children:

<https://www.amnesty.org/en/latest/news/2025/10/tiktok-steering-children-towards-depressive-and-suicidal-content/>

Foundation²¹, Liberties and European Partnership for Democracy,²² published throughout the respective assessment periods, has not been addressed at all by platforms in the reports, even though Recital 90 requires them to ensure that their approach “is based on the best available information and scientific insights”.

Lesson 5: Platforms should more consistently account for the different languages and geographic contexts across EU Member States.

A comprehensive risk assessment should include an explanation of how risks or the effectiveness of mitigation measures varies depending on geographic and linguistic contexts. **Platforms approach this in a fragmented manner**, focusing on selected highlights from certain EU countries without explaining why they spotlighted those cases. We have not identified any risk assessments that adequately and systematically address the EU’s diversity and the resulting variety of risks. There was also minimal attention to the perspectives of at-risk communities in Europe, notably migrants and the Roma community.

For example:

- Beyond reporting on the number of content moderators per EU language, **platforms do not sufficiently consider the EU's linguistic diversity**, e.g. complementing this data with assessments of the risks that can disproportionately affect users of certain languages. There is also no discussion of how platforms deal with languages commonly spoken in the EU that are not official EU languages, such as Catalan, Arabic, Turkish, Romani, and Ukrainian.
- Even for official EU languages, there are **questions around whether platforms dedicate enough human resources for content moderation**. This is notably the case of Maltese and Irish, for which several platforms do not provide a single content moderator. However, the overall high volume of content moderation decisions means that insufficient staffing could also affect other EU languages, considering they receive disproportionately fewer human resources than English. This is particularly serious in the case of X: out of 24 EU languages, 13 of them do not have a single moderator (native speaker or not).²³
- **We did not see sufficient recognition of geographic diversity when assessing systemic risks**. For example, TikTok states that “where the risks are localised

21 Panoptykon Foundation’s research on Facebook’s recommender system: <https://en.panoptykon.org/anxious-about-your-health-facebook-wont-let-you-forget>

22 EPD’s and Liberties’ analysis of systemic risks to civic discourse and electoral processes: <https://epd.eu/news-publications/civic-discourse-and-electoral-processes-in-the-risk-assessment-and-mitigation-measures-reports-under-the-digital-services-act-an-analysis/>

23 X’s October 2025 Transparency Report.

or there are linguistic differences, TikTok takes these into consideration when assessing risk and corresponding mitigations.”²⁴ However, the platform does not present any evidence to support this claim, e.g. by further explaining which risks are localised and how. In previous assessments,²⁵ TikTok included only two national case studies (Spain and France), which is a fragmented and insufficient approach.

Conclusions and recommendations

The first three rounds of risk assessments raise fundamental questions about their utility and future direction. While we’ve seen some improvement in the latest round of assessments— notably the fact that most assessments now have a dedicated section on fundamental rights—critical issues must be urgently addressed for these assessments to serve as meaningful accountability tools.

First, future risk assessments need substantive improvement to be truly useful. Other civil society and academic analyses²⁶ as well as the Commission’s own investigations²⁷ highlight the crucial gaps of risk assessments, both in terms of methodology and supporting evidence. We urge the Commission to **provide clear guidance on what data platforms are expected to disclose in risk assessments, along with defined standards and expectations.** Without such clarity, it is unlikely that subsequent rounds of risk assessments will show meaningful improvement and achieve desired result as stated in the DSA.

Second, access to data mechanisms, while encouraged, will not necessarily resolve the identified limitations. This is due to three primary factors: 1) not many civil society organisations have the expertise needed to access and analyse non-public data, 2) platforms render data access difficult or outright deny it, and 3) the infrastructure and experience required to request relevant data from platforms make it difficult to access it. Platforms must proactively provide evidence and substantive information to support their claims.

How these issues are addressed will determine whether risk assessments become meaningful tools for transparency and accountability or remain little more than superficial compliance exercises.

24 TikTok’s 2025 Risk Assessment Report, p. 13.

25 TikTok’s 2023 Risk Assessment Report, p. 37 and p. 72.

26 See for example: <https://cdt.org/insights/dsa-civil-society-coordination-group-publishes-an-initial-analysis-of-the-major-online-platforms-risks-analysis-reports/>, <https://www.techpolicy.press/beyond-disinformation-how-dsa-risk-assessments-ignore-democracys-real-threats/>, <https://kgi.georgetown.edu/research-and-commentary/systemic-risk-assessment-under-the-digital-services-act/>

27 For example, the investigation into TikTok: https://ec.europa.eu/commission/presscorner/detail/en/ip_26_312