Civil society statement on meaningful transparency of risk assessments under the Digital Services Act

Meaningful transparency of risk assessments and audits enables external stakeholders, including civil society organisations, researchers, journalists, and people impacted by systemic risks, to scrutinise the assessment and ensure it is more than merely a “tick box” exercise. Transparency is crucial for explaining how exactly the risk assessment may have influenced VLOPs and VLOSEs’ design and development of their services, including algorithmic systems.

As per Article 42(4) of the DSA, VLOPs and VLOSEs must publish extensive documentation of the risk assessment and auditing process, including a report on the results of the risk assessment, the specific mitigation measures implemented, and audit documentation, including the audit report and implementation report. While the DSA does not explicitly require the publication of the entire risk assessment, Recital 100 emphasises the importance of comprehensively reporting on risk assessments, considering the heightened risks related to the functioning of VLOPs and VLOSEs.

VLOPs and VLOSEs should therefore publish all information about the risk assessment and auditing processes that makes it possible for stakeholders to meaningfully evaluate the results of the assessment (i.e., the completeness and accuracy of the identified risks and corresponding mitigation measures) as well as the methods used. This will ensure that multiple

perspectives can identify potential gaps and unaddressed risks, assess whether mitigation measures are appropriate and uphold fundamental rights, as well as learning and sharing best practices to raise the general standards of risk assessment and mitigation. It is widely accepted that implementation of the DSA must build on joint learning and expert collaboration. These annual risk and audit reports are one of the key ways by which external actors can do that. It is therefore vital that we have the information we need to make meaningful contributions.

Specifically, we expect VLOPs and VLOSEs to publish, as a minimum, the following information regarding risk assessments:

1. **A detailed methodology of the risk assessment**, including:
    - The definition of the assessment scope, identifying relevant boundaries of assessment, addressing applicable services to evaluate, and assessment frequency.
    - The definition and the threshold adopted by platforms for classifying a risk as "systemic", and what criteria were used to assess this;
    - How systemic risks listed in Article 34(1)(c) and (d) were defined and what approach was taken to assess them;
    - Technical approaches, metrics, and benchmarks adopted, and whether there were key technical implementation challenges for DSA risk assessments. This should include descriptions of the following factors:
        i. Data Collection, e.g., establishing appropriate and relevant metrics and data, including how on- and off-platform data was used. Where relevant data was not available, this should be stated.
        ii. Strategies and systems to identify false positives and false negatives in identifying risks, and how the impacts of false positives versus false negatives were weighed and the process adjusted accordingly.
        iii. How risk scoring processes and metrics were developed, including whether these were used consistently and/or differed depending on context.
    - Whether any additional assessments were conducted besides the annual one and if so, what were the grounds for making this decision.
    - How the VLOP or VLOSE's previous, or other in-use, risk or impact assessment processes and methodologies were drawn upon.

    - Any real or hypothetical case studies used.
2. A detailed **mapping and description of the assessed service and its overall structure**, including for any algorithmic or advertising systems, so the public can understand the platform's architecture, its functioning, and the interdependence between its different systems or elements.
    - **For algorithmic systems the description should include, the purposes and key design specifications of each assessed system**, including the system's general logic, the rationale and assumptions made with regard to potentially-affected persons or groups of persons, the main classification choices (including if these are deterministic or use a statistical model, and if so information about the functioning of these), what the system is designed to optimise for and the relevance of the different parameters, and the decisions about any possible trade-offs;
3. **Identified risks**, including an explanation of which risks were prioritised in the assessment and mitigation and why, a classification of the risks, and a clear indication of which part(s) of the service they stem from;
    - Also a clear explanation of any decisions made to prioritise or weight certain risks over others.
4. **Measures or actions already taken, or to be taken,** following the assessment, including:
    - Any **mitigation measures** adopted for each risk, as per Article 42(4)(b), including those adopted following the audit, and how their impacts are being monitored and assessed (including potential negative impacts on fundamental rights);
    - Any decision to abandon the launch of new products or features because it was considered that their risk to fundamental rights could not be effectively mitigated;
    - How the threshold of unacceptability was defined for other systemic risks and which risks were assessed as unacceptably high;
    - For risks assessed as "unacceptably high", an explanation of steps taken by the platform to prevent them materialising, e.g., changing the design of the service, discontinuing a specific system or policy, refraining from introducing a new system or policy, or hiring/training staff.
5. **Which internal departments, teams, experts were involved** in the assessment, including the level of seniority responsible for approving the assessment and

implementing the audit and what forms of expertise were decided to be the most relevant for conducting different areas of the assessment.

6. **Which external stakeholders were consulted** as part of the risk assessment, when and how they were engaged with, and any outcomes of these consultations, including any details of how the VLOP or VLOSE responded to external stakeholders' input and integrated it in their assessment and mitigation measures, or more generally in their services and activities.
    - Platforms should indicate if they conducted stakeholder consultations specifically for the risk assessment in question and what this process looked like including indicating the size and type of consultations and the types of participants, e.g., users, potentially impacted groups, civil society representatives and independent experts (as outlined in recital 90). If they engage with stakeholders on a continuous basis, they should explain how exactly these consultations influenced the risk assessment and mitigation processes and how they informed stakeholders about the measures implemented as a result of their input.
7. Plans or proposals for changes to the process of the next risk assessment(s), including the annual assessments and any upcoming new functionality assessments, with the grounds for these changes as well as broader lessons learned from this first round of assessments.

Audit Reports

When it comes to audit reports and audit implementation reports, we expect VLOPs and VLOSEs to publish:

- the full audit report, or at least the part relating to the evaluation of identified risks and the appropriateness of mitigation measures, and relevant recommendations;
- the full implementation report, especially explanation why the respective VLOPs and VLOSEs chose not to implement certain recommendations and how they chose to address the identified issues instead.

We also expect that in audit reports auditors will take note of the requirement in the template under the delegated act on audits that "the information provided should be complete and detailed such that a third party with no previous connection with the audit is **able to understand the description of the findings**," drawing on but not limited to the specifications in Recitals

19-36 of the delegated act. In addition to understanding the descriptions, complete and detailed information is vital to build trust and the opportunity to provide constructive input. This information should include, at minimum:

- how relevant expertise, including in the field of fundamental rights, was employed by the auditing firm (including via subcontracting);
- comments on availability, or not, of relevant data and evidence;
- any issues in communication or working with audited platforms during the audit process, and how (if at all) these were addressed;
- comments on organisational aspects of VLOPs and VLOSEs which are relevant to the effectiveness of risk assessment and mitigations, such as reporting structures and staffing;
- where external resources were used, how these were searched for / located and deemed relevant;
- plans or proposals for changes to the process of the next audit(s), with the grounds for these changes as well as broader lessons learned from this first round of audits.

Signatories:

- Access Now
- AI Forensics
- Algorithm Watch
- ARTICLE 19
- Centre for Democracy and Technology Europe
- Civil Liberties Union of Europe
- Das NETTZ
- Eticas Foundation
- European Centre for Non-for-profit Law
- European Partnership for Democracy
- Future of Free Speech
- Global Disinformation Index
- Global Witness
- Institute for Strategic Dialogue (ISD)
- International Media Support
- Mnemonic
- Panoptykon Foundation
- People vs Big Tech