Implementing risk assessment obligations under the Digital Services Act

Questions, connections, and resources

Prepared for the Action Coalition on Meaningful Transparency by Brainbox

www.meaningfultransparency.tech/

May 2023
About the Action Coalition on Meaningful Transparency

This briefing note was drafted by the Brainbox Institute, in its role as the Project Lead for the Action Coalition on Meaningful Transparency (ACT), a multi-stakeholder coalition organised under the auspices of the Tech for Democracy Initiative. The ACT’s Steering Group brings together a range of civil society organisations, with diverse geographic and subject-matter expertise:

- GNI – The Global Network Initiative (Global)
- CDT – The Center for Democracy & Technology (US & EU)
- CTS at FGV – The Center for Technology and Society, FGV University Rio (Brazil)
- CIGI – The Centre for International Governance Innovation (Canada)
- CCG, NLU Delhi – The Centre for Communications Governance (India)
- DFR Lab – The Digital Forensic Research Lab (US)
- ISD – The Institute for Strategic Dialogue (UK, Germany, US & global)
- Internet Lab (Brazil)
- RIA – Research ICT Africa (South Africa)

The ACT’s Advisory Group includes individuals from companies, governments, and international organisations. We are building a broad and globally diverse multi-stakeholder coalition led by civil society to support the design and implementation of broader transparency initiatives. The ACT is working to: clarify relevant definitions and terminology; map existing and identify missing efforts; share lessons learned; identify opportunities for alignment and coordination; encourage and support participation of majority world actors in these efforts; produce recommendations for companies, donors, researchers, and governments.

The ACT is working towards these objectives by coordinating and amplifying existing work, preparing discussion documents and written outputs and hosting discussions publicly and privately. Interested parties can find out more and join the Action Coalition at meaningfultransparency.tech

Disclaimer

This briefing document reflects information shared by the ACT’s members but should not be understood to speak for them. The ACT welcomes further opportunities to engage with regulators and others who will be involved in implementing transparency frameworks on this and other issues within our mandate.
Context for briefing note

Subject of this briefing note

Risk management is at the core of the Digital Services Act, and a key part of this risk management approach is the requirement for very large online platforms (VLOPs) and very large online search engines (VLOSEs) to conduct assessments of the “systemic risk” stemming from the services they offer.

There is a wealth of useful guidance, information and commentary on risk assessment in a specific DSA context, as well as in a broader human rights risk assessment landscape. Companies are actively preparing for conducting DSA risk assessments including with external consultancy organisations familiar with risk assessment in other contexts. Industry bodies and membership organisations are also looking to support members to implement risk assessment obligations and help to resolve any practical uncertainties through closed consultations.

While there are certain risks and considerations explicitly laid out for inclusion in these assessments in the text of the DSA, much uncertainty remains about the scope of these risk assessments, what type of information they will include, what benchmarks will be used, and how they will be conducted and reported on.

Process behind this briefing note

The Action Coalition for Meaningful Transparency has prepared this briefing note to inform processes and practices around risk assessments as required by the DSA. The Action Coalition on Meaningful Transparency’s membership includes a wide range of relevant stakeholders and experts, so we sought quick feedback and input from them through an online questionnaire. From the inputs received, we identified three key themes running through the feedback, commentary, and resources received.

Broader themes

Contextuality, uncertainty and independent expert input

Respondents highlighted a range of useful frameworks (linked below), and respondents emphasised there was some value in clearly referencing to, and building upon, existing risk assessment and risk management frameworks. However, DSA risk assessments are unique, context-dependent, and interdependent:

- There was consensus that DSA risk assessments as a whole are unique, and there is no clear unitary statement of the minimum expectations of DSA risk assessments, nor a one-size-fits-all methodology that can be adopted.
- In addition, risk assessments must be highly contextualised to individual companies: the exact nature of the risk will be dependent on the type of company, the product, the reach and scale of the product, and the context it is implemented in.
- There is also a prospect that the various components comprising a VLOP or VLOSE’s systems are so interdependent that risks in one component may interact in a chaotic or unpredictable manner with others.
Risk assessment is a dynamic exercise that responds to internal and external events over time, including in responses to changes in user behaviour, broader socio-economic contexts, or other unpredictable factors.

Because assessment of risks under the DSA is unique, context-dependent, dynamic, and interdependent, there was a heavy emphasis among respondents on the value and necessity of involving independent experts and stakeholders in the design, implementation, and review of risk assessment methods and processes. There was also value identified to independent input into the design of risk assessment reports. Respondents expect risk assessment reports to illustrate the specific ways that external input had been sought and incorporated into risk assessment implementation. In fact, the relevance and importance of stakeholder engagement processes was one of the stronger themes across all responses.

Specificity

Because of the generality, uncertainty, and context-dependence of risk assessment requirements, respondents expressed concern such assessments could become too generic. As a result, a number of experts emphasised the importance of specificity in the way that risk assessments are scoped, designed, executed, and evaluated. Respondents called for specificity in risk assessments across the following areas.

- **Criteria for prioritisation:** Risks and responses to them can be prioritised through clear criteria based on the UN Guiding Principles on Business and Human Rights. Four key criteria include: scope (number of affected people); scale (gravity of impact for affected people); remediability (the effectiveness of a remedy for affected people); and likelihood (the probability and frequency of the impact occurring).

- **Rights:** It should be clear which human rights are being considered. Article 34 lists several specific human rights, but all human rights are important and interdependent. If there is a risk to a particular human right, risk assessments should be specific about the risk to that right (for example, simply naming a generic risk to freedom of expression is inadequate).

- **Scope of risks:** Risk assessments should be structured to account for macro, society-wide interactions and risks, as well as community and user-level risks and concerns.

- **Specific products and services in context:** Service providers must engage with the specific nuances of their own products and services (e.g. how a policy on monetisation for creators might impact the preponderance of disinformation within a specific product), as well as the specific diverse contexts – national, social, and political – in which they are used in (e.g. whether policies or takedown notices are provided in local languages).

- **Thresholds for action:** The DSA requires risk assessments to be carried out on functionalities that are likely to have a “critical impact” on the identified risks. It is unclear how a “critical impact” should be understood, and so companies should be clear what thresholds they have set, and how those thresholds will be measured.

- **Emerging risks:** Service providers should consider the impact that a changing global or local context, or developing or emerging technologies – such as synthetic media – might have in their risk assessments, as well as thresholds for re-considering established risks.

- **Governance, escalation and compliance structures:** Risk assessment obligations are accompanied by compliance function obligations under article 41. Further to this, risk
assessments should include information on the role that platform governance decisions can present for systemic risk. Risk assessment reports should include elements of the organisational chart and reporting structure, outlines of decision-making authority, and descriptions of other internal governance structures and processes that could impact product, policy, and assessment outcomes.

- **Incentives and conflicts**: Risk assessments should include reference to the particular incentives at play that could impact decision-making within a company that might influence its assessment of the risks presented by its product, policy, and assessment options (e.g. an advertising revenue model or stream creating a risk of incentive to apply less stringent review or content standards for ad-based content, which should be declared and acknowledged).

- **Trade-offs and decision-making**: Designing risk assessment methods, and implementing risk assessment obligations, will require decisions to be made about trade-offs. Risk assessment reports should identify any trade-offs in monitoring and accounting for risk, as well as the processes behind balancing and justifying various risks.

- **Independent input and stakeholder engagement**: Risk assessment methods and reports should document processes, outputs, and outcomes of stakeholder engagement undertaken as part of the risk assessment process. Documenting processes, outputs and outcomes is important for avoiding the risk that consultation is meaningless “engagement theatre”. However, such documentation should also be undertaken with awareness to the sensitivities that certain stakeholders may have about their involvement in such processes.

These areas will benefit from as much specificity as possible, and stakeholder engagement on risk assessment methods and reports could draw out perspectives from independent parties on any areas where specificity is perceived to be lacking.

### Measurability

A third theme from ACT respondents focused on the importance of measurability in risk assessment. Feedback from respondents covered the following points:

- **Risk assessments should include**, insofar as possible, clear measurements, benchmarks, or estimations to quantify existing risks and harms. These measurements are important for ongoing monitoring of risk, as well as measuring variations in risk over time, including in response to changing contexts and internal decision-making.

- **There is merit to both qualitative and quantitative assessments**, but there should be justifications offered for why particular qualitative or quantitative measures and approaches have been adopted, with some accounting for tracking qualitative measurements over time.

- **Risk assessment processes and reports should include clearly recorded and auditable plans and processes for the ongoing monitoring and management of risks**, including based on measurable thresholds for action.

- **Methods of measurement incorporated into risk assessment methods and implementation can benefit from independent expert involvement.** Initial methods can be tested with external parties to suggest more concrete benchmarks for measuring change over time, including by comparison of assessment reports. The Code of Practice on Disinformation was identified as an example of a structure that lacks benchmarks, and therefore presents opportunities for improvement.
● One respondent suggested that, over the next 3-10 years, DSA risk assessment could aim for standardised “systemic risk statements”, with a similar look and feel to “enterprise risk statements” included in United States Security and Exchange Commission Form 10-K reports (under Item 1A: Risk Factors).

Resources on risk assessment

The “systemic risk” assessments required by the DSA were seen as novel, requiring comparison with and development from other areas of risk assessment – particularly business and human rights risk assessment and the emerging field of algorithmic risk assessment. Aside from very broad risk assessment frameworks (such as ISO 31000), respondents did not raise financial risk assessment frameworks as a source of learning or point of comparison, but this may reflect the areas of specific interest among the respondent group.

Primers and frameworks for human rights risk assessment more broadly

Human rights risk assessment is an established area and respondents shared a number of useful resources for framing the broad approach to human rights risk assessment, as well as some specific frameworks that have benefited from testing and implementation over time, including through multiple assessment cycles.

● The [UN Guiding Principles on Business and Human Rights](https://www.un.org/en/guiding-principles-business-human-rights/) were identified as a foundational framework for DSA risk assessments.
● BSR has published a primer on how it approaches human rights assessment available here.
● BSR has published a paper from 2021 describing its approach to defining a human rights-based approach to content governance available here.
● The Global Network Initiative (GNI) operates a framework that outlines a rights-protecting approach to protecting freedom of expression and privacy in the context of government demands, pressures and restrictions on ICT companies. This framework sets out guidance and evaluates company policy and practice related to internal governance, due diligence and risk management generally, freedom of expression and privacy protections specifically, and transparency and engagement. GNI accredits and trains assessors to review company policies and practices against this framework to identify lessons and areas for improvement. The framework consists of:
  ○ The GNI principles, available here.
  ○ The GNI implementation guidelines, available here.
  ○ The GNI assessment toolkit, available here.
● The Danish Institute for Human Rights has published an introduction on human rights impact assessment in general (published in 2020, although silent on the DSA), available here.
● Human Rights Impact Assessment frameworks were suggested as a foundation for ensuring risk assessment processes consider human rights not just in their content, but also in the approach that they take to conducting the assessment. See page 27 onwards here.
● The Digital Trust and Safety Partnership and Ofcom are co-chairsing a World Economic Forum working group on a “Digital Safety Risk Assessment Framework” (workstream 3). They will develop the framework, a bank of case studies, and reports on risk factors and solutions-based interventions. This work has not yet been published.
● The Danish Institute for Human Rights has developed a “Digital Rights Check” tool, which is a guidance and assessment tool for staff who are working on digital projects or digital
solutions in a technical development cooperation, or development finance context. Available here.

- GNI and BSR have published a tool for identifying roles and responsibilities across the tech ecosystem, referred to as the “Across the Stack” tool. It is a powerpoint that illustrates different actors and responsibilities relevant to human rights due diligence related to digital and other technologies. The tool is available here.

Platform risk assessment beyond the DSA

In addition to resources on risk assessment under the Digital Services Act, respondents also drew attention to the broader legislative landscape in Europe and in other jurisdictions. Some legislative proposals impose comparable risk assessment requirements, and entities looking to implement risk assessment frameworks will be influenced to some extent by these developments.

- BSR’s analysis of the “state of play” on human rights due diligence in the technology industry, available here.
- BSR has prepared an analysis of the DSA’s risk assessment provisions and the way they overlap with three other incoming EU regulations: the EU Corporate Sustainability Due Diligence Directive, the EU Corporate Sustainability Reporting Directive, and the AI Act, available here.
- Ofcom has published an explanation of how they are approaching online safety risk assessments, available here.
- Ofcom has published a report on their first year of regulating video sharing platforms, available here.
- The Australian eSafety commissioner’s report on industry responses to the first mandatory transparency notices, available here.

Systemic risk assessment under the DSA

Respondents drew our attention to a range of materials that included specific comments on public expectations about how analysis of systemic risks under the DSA would take place, including concerns raised about the drafting and design of those frameworks during the DSA design process. Other more recent materials focus on the implementation of DSA risk assessment more specifically since the DSA came into force.

- An analysis of “systemic electoral risk” in the 2021 German Federal Elections published by Facebook and Twitter from the Sustainable Computing Lab, available here.
- A case for the need to include a gender lens in DSA due diligence and risk assessment from the Advocacy Director for Europe, Online Expression & Civic Space at the Center for Democracy and Technology, available here.
- The Institute for Strategic Dialogue touches on risk assessments in its policy digests 2-4. It is in the process of preparing notes, summaries and issue-specific reports related to risk assessment (Policy digest 2, Policy digest 3 and Policy digest 4)
- The DSA Observatory has published a piece on “what to expect from risk assessments and audits under the DSA – and when” (30 January 2023), available here.
- Verfassungsblog has published on fundamental rights impact assessments in the DSA (November 2022), available here.
The DSA Observatory published an overview of the DSA proposal dated October 2021, but it has broad thematic comments and concerns on risk assessment that have continued relevance, available here.

Mozilla published a position paper on the DSA dated May 2021, but it includes broad positions on risk assessments of ongoing relevance, available here.

The Center for Democracy and Technology (CDT) has published a report on the intersectional nature of human rights risks in the context of DSA risk assessment, available here.

CDT and GNI held a private roundtable and a public event on “Human Rights Due Diligence and the Digital Services Act” in July 2021, video and the public event report are available here.

A piece on putting the DSA into practice was published in March 2023 by authors from the University of Amsterdam, the University of Oxford, the DSA Observatory, and Verfassungsblog dealing with risk assessment, available here.

In a TVEC risk context, Brian Fishman has argued in favour of considering the “surfaces” vulnerable to attack, rather than linking regulation to quantitative measures like revenue or users, available here.

The Forum on Information & Democracy published a report in February 2023 on pluralism of news and information in curation and indexing algorithms, which touches on risk assessment in those contexts, available here.

Algorithmic and AI risk assessment

The field of algorithmic and AI risk assessment is very broad, but ACT members identified the following resources as being useful in their own practice.

- Taraaz Research has published a GitHub repository with links to more than 30 relevant resources for AI risk assessment, including on risk assessment in machine learning-enabled systems, self-assessment, and documentation tool, available here.
- LAPIN has published an analysis of Assessment of Algorithmic Impact for the Protection of Fundamental Rights, available here.
- This 2022 book explores the human rights, ethical, and social impact of AI, including many of the systems used by VLOPs and VLOSEs, available here.
- Trilateral Research has published a survey of AI assessment methodologies, available here.
- In January 2023, Jessica Newman published “A Taxonomy of Trustworthiness for Artificial Intelligence” as a white paper for the Center for Long-Term Cybersecurity at UC Berkeley. The paper covers existing frameworks for trustworthy AI. Available here.
- The US Department of Energy has developed an AI Risk Management Playbook, available here.
- The US National Institute of Standards and Technology (NIST) has developed an AI Risk Assessment Framework, available here.
- Dedicated AI risk assessment frameworks are being developed by international standards organisations like the ISO and the IEEE.

Risk assessment frameworks related to technology generally

There are a number of international standards organisations that publish either generic risk assessment and management frameworks, or tailored frameworks for application to digital technologies. In addition, there are tools and resources available to support relevant parties to conduct assessments of digital tools:
● Respondents identified the International Standards Organisation's ISO 31000 on Risk Management as a foundational resource, available here.
● The Ford Foundation has published a framework guide to vetting public sector technology vendors, available here.
● NIST (the US National Institute of Standards and Technology) has developed generic risk management frameworks as well as frameworks for particular contexts and technologies, including cybersecurity and AI.

Notable organisations

In addition to the organisations publishing the resources above, respondents particularly called out the following individuals and organisations as sources of expertise they were following:

● Organisations comprised of practitioners with actual experience of designing and conducting human rights risk assessments of technology companies, including Business for Social Responsibility and the Global Network Initiative. The Integrity Institute, a membership organisation for platform integrity professionals, may be another body to follow.
● Civil society organisations, including Access Now, Privacy International, Global Partners Digital, the European Center for Non-for-Profit Law, and the Institute for Strategic Dialogue.
● Technical standards-setting bodies including NIST, DOE, IEEE, 3GPP, and ISO.
● National Human Rights Institutions and independent human rights bodies.
● The United Nations B-Tech project.
● EU Data Protection Authorities, given their existing experience with Data Protection Impact Assessments.
● The World Economic Forum working group on a “Digital Safety Risk Assessment Framework” (workstream 3). They will develop the framework, a bank of case studies, and reports on risk factors and solutions-based interventions. This work has not yet been published.
● Taraaz, a research organisation founded by Roya Pakzad, a former tech worker and tech and human rights researcher, who has published a GitHub with more than 30 resources on risk assessment in AI systems, including documentation tools and self assessment frameworks.
● Jenny Brennan at the Ada Lovelace Institute and Dr. Anna-Katharina Meßmer at Stiftung Neue Verantwortung were recommended as knowledgeable experts to be consulted. Dr Meßmer recently published on auditing recommender systems.

Observations and concluding comments

We offer the following observations by way of conclusion.

1. The value of external expertise and stakeholder engagement at all stages:

Independent external expertise is essential when designing, implementing, and reviewing risk assessment methods and reporting frameworks. There is a role for the European Commission to foster trusted discussions with companies, practitioners, and civil society organisations to build shared expectations and knowledge about how risk assessments will be conducted. There is substantial benefit to learning from those who have existing experience in similar assessment processes. Risk assessment reports ought to describe in detail how independent expertise and experience was incorporated.
2. The value of specificity in assessment methodologies, assessments, and reports:

Risk assessments should be as specific as possible about platforms' individual products and services and their features. We have expanded on the kinds of specifics respondents expect to see in risk assessment methods and reports in the thematic section on this point above.

3. Metrics and benchmarks are essential to monitoring risk and progress:

Metrics, benchmarks, and measurements are difficult to present, but are also desirable because they present the capacity to review progress or regressions over time, including in response to contextual events, or internal decision-making within platform companies. Risk assessments should be designed with a view to the way they will be repeated over time. Risk assessment is not a one-off practice, and assessments also require active monitoring and review in between regular assessments to account for changing circumstances. Risk assessments should disclose the names and affiliations of assessors. Risk assessments should disclose relevant mitigation measures, as well as thresholds for monitoring and escalation.

4. Encourage fulsome disclosure:

While it is expected that some information may be withheld in risk assessments, any information that is withheld should be as granular as possible, and linked closely to clear withholding grounds. Any decisions to withhold information should be capable of independent audit.

These observations are drawn together by Brainbox based on ACT member responses, and should not be taken to speak for any ACT member.

Invitation to collaborate further

The ACT will be hosting further discussions about the implementation of risk assessment frameworks and aiming to coordinate and connect others working on this topic, as well as amplifying their work.

For further information, please contact the project lead for the Action Coalition on Meaningful Transparency (Brainbox Institute) and visit meaningfultransparency.tech to stay updated on relevant events and further outputs.