GLOBAL NETWORK INITIATIVE

Implementing risk assessments under the Digital Services Act

DISCUSSION SUMMARY

JUNE 2023

Digital Trust & Safety Partnership

Disclaimer

The findings, interpretations and conclusions expressed in this document are a result of a process facilitated by the Global Network Initiative, the Digital Trust and Safety Partnership and Brainbox, but do not necessarily represent the views of those organisations, nor the entirety of their members, partners, or other stakeholders. Participants attended under the Chatham House rule. The comments and observations in this document reflect our understanding of the comments made during the discussion and should not be attributed to any individual participant.

Introductory comments

The EU Digital Services Act (DSA) is built on years of conversations and experimentation around how best to identify and control risks associated with digital content and services. The DSA recognises and seeks to support the steps that many platforms have taken voluntarily over the last 15 years, including expanded transparency reporting, robust content moderation, and increasingly mature approaches to human rights due diligence. Recognising the positive potential of these and other rights-enhancing mechanisms, the DSA seeks to ensure that they are applied more broadly, consistently, and effectively across the online ecosystem. This substantive continuity is reinforced by the DSA's open architecture, which leaves room for, and in some cases explicitly encourages, opportunities for intermediaries, civil society, researchers, auditors, dispute resolution bodies, and others to engage with each other and with regulators to understand and improve digital services.

In the spirit of this openness, on 24 and 25 May 2023, the Digital Trust and Safety Partnership (DTSP) and the Global Network Initiative (GNI) invited our respective company members to join with a wide range of academic and civil society experts in conversation around the risk assessment provision of the DSA (Article 34). The conversations were conducted virtually and spread over two,150 minute workshops, in order to maximise participation across key time zones.

The purpose of the workshops was to facilitate input from civil society experts about how companies can understand and undertake risk assessment and design risk mitigation as laid out under the DSA, as well as for civil society to hear questions, concerns, and ideas from companies. The timing of the workshop was intended to help inform compliance efforts being undertaken by those participating companies that have been designated as Very Large Online Service Providers or Search Engines ("providers"). Because of this timing, company participants were limited in what they could share about risk assessments that were ongoing at the time of the workshops.

The discussions during the workshops were constructive and open-ended. Specific sessions were organised first around the concept of "stakeholder engagement," and then each of the four categories of "systemic risk" set out in Article 34(1)(a-d). Experts were chosen to give short introductions to each topic before opening the floor to discussion. The summary below attempts to capture the key points raised in each session, while protecting participant identities in line with the Chatham House Rule.

Three overarching points emerged across the various sessions. First, participants recognised that the text of the DSA leaves substantial room for interpretation. While this ambiguity was recognised as creating challenges for providers, it was also seen as affording flexibility, experimentation, and room for collaboration. Second, participants were largely in agreement that the DSA should be interpreted in line with existing international frameworks and good practices related to business and human rights. This approach not only allows for continuity within and across company practice, but helps foster shared understandings and expectations across stakeholders. Finally, participants were unanimous in their endorsement of the need for on-going openness and collaboration across companies and other stakeholders. To this end, DTSP and GNI consider the workshops to have been a success and are committed to continuing to support and convene conversations across stakeholders, including relevant regulators at the EU and member state levels.

GNI has worked over two decades to build collaboration on digital rights across an increasingly diverse membership of academics, civil society organisations, information and communication technology (ICT) companies, and investors. The GNI Principles and Implementation Guidelines draw on international human rights principles to guide ICT companies' efforts to identify and mitigate human rights risks. GNI's unique assessment process relies upon GNI-accredited assessors to review member companies' relevant systems, processes, and experiences and to share detailed, sensitive, and often non-public information with GNI's non-company Board representatives, who are charged with determining whether companies are implementing the Principles and Guidelines in good faith, with improvement over time.

DTSP was established in 2021 to promote a safer and more trustworthy internet. DTSP's partner companies are committed to developing, using and promoting industry best practices, reviewed through internal and independent third-party assessments, to ensure consumer trust and safety when using digital services.

Jason Pielemeier
Executive Director
Global Network Initiative

David Sullivan
Executive Director
Digital Trust and Safety Partnership

Scene-setting: the importance of stakeholder engagement

Discussions began with a concise comparative overview of the role of stakeholder engagement in risk assessments under the DSA, by reference to broader human rights frameworks.

Participants noted that while Recital 90 emphasises the importance of stakeholder engagement, the DSA does not stipulate a method of identifying relevant stakeholders, nor any process requirements around how they should be consulted. Relatedly, participants noted that there were a range of existing frameworks for stakeholder engagement in the context of tech company human rights impacts, and also expressed some concern at the prospect that stakeholder engagement processes might diverge unnecessarily in the way they're conducted under the DSA by comparison with other frameworks, such as the UN Guiding Principles on Business and Human Rights (UNGPs).

Stakeholder engagement serves multiple important purposes. The obligation to engage with stakeholders in the DSA is positioned alongside an obligation to ensure that risk assessments take advantage of best available scientific information and insights, as well as accurate information. In this regard, stakeholder engagement plays an important role in ensuring that risk assessments are accurate and reliable, including whether risk mitigations are effective. Stakeholder engagement also enables risks to human rights to be more comprehensively assessed, as well as anticipation of potential remediations that may be required where human rights are undermined or infringed for particular individuals or groups.

Participants emphasised that stakeholder engagement must be *effective* in order for it to play its proper role. Participants shared extensive prior work to outline the requirements on stakeholder engagement processes to make them meaningful. By way of summary:

- Engagements must be based on a shared purpose and shared interest. There must be real buy-in from the outset by all parties, and vigilance is required to ensure engagements are not performative.
- Engagements must be founded on a transparent and trustworthy process. The process must be inclusive, open, fair and respectful. It must also account for the diverse needs of different groups, including through reasonable accommodations.
- Importantly, engagements must also "close the loop" and have a visible impact that is reported back to stakeholders. Companies and others leading engagements must report back to consulted stakeholders to inform them how their concerns were addressed, and what has been done to act on stakeholder feedback.

A significant challenge faced by providers of very large online platforms and of very large online search engines ("providers") conducting DSA risk assessments will be defining a necessary, justifiable and proportionate scope for who constitutes an affected stakeholder. There are two key dimensions of complexity here: the extent to which non-European stakeholders are engaged; and the inclusion of non-users.

The DSA will have a global impact, both because of the "Brussels effect", and because platforms tend to have global content policies (with some limited jurisdictional variation). Equally, some participants expected that companies will, as a matter of practical necessity, be engaging with "systemic risks" that extend beyond Europe as a necessary part of DSA compliance, since it is very difficult to "contain" the jurisdictional scope of digital risks. While processes put in place to implement DSA risk assessment and mitigations may have protective effects in other jurisdictions (positive spillover), there is also a risk that those measures may have unintended negative consequences elsewhere (negative spillover). Among these, it was noted, is the possibility that the resources required for implementing thorough risk assessment and mitigation in Europe may leave fewer resources available to address risks in other jurisdictions. As a result, participants noted the importance of avoiding overly narrow approaches to stakeholder engagement.

It is also clear that DSA risk assessments – as a matter of existing human rights practice and observed experience in contexts such as Myanmar and Ethiopia – will also have to include the impacts of products and services on people and groups who are not users of the services themselves. The need to consider the impact on non-users is also underscored by the expectation that products and services will not only refrain from interfering in human rights, but will also protect and promote human rights too. It was noted that non-users can be more difficult for providers to identify and engage directly, which underscores the role that credible civil society organisations can play in helping providers understand, identify, and mitigate the risks to these individuals.

Participants shared a number of resources that could be used by companies and affected stakeholders to build and shape expectations around DSA stakeholder engagement, in the context of Article 34 and systemic risk.

What is systemic risk?

Substantial work is underway to arrive at an operable definition of "systemic risk" as a concept under the DSA. A shared definition of systemic risk, as understood in the context of the DSA, is critical, given the way it is threaded through a range of relevant provisions, including Article 40 as well as Articles 34 and 35.

What makes a risk "systemic"?
Given the lack of clarity as to what "systemic risk" means in the digital services context, several participants pointed to other frameworks that have been widely used to identify and prioritise business-related risks. It was noted that the UNGPs and the OECD Guidelines on Multinational Enterprises (Guidelines) recognise the need for "prioritisation" in the context of assessing and mitigating risks. Under these frameworks, prioritisation is tied to the "severity" of risk, which can be identified by examining its scope (gravity), scale (how widespread it is), and remediability (extent to which it can be remedied).

The discussion identified a range of features that could be used to understand when a risk is "systemic", which broadly fell into two categories: a) risks that have broad impact on "systems"; and b) risks that are caused or exacerbated by systems.

Examples of risks that are systemic because of their impact on systems include the following:

- Where content and conduct occur across various digital products and services, including those operated by multiple providers. When this is the case, the scale of impact is often amplified, while mitigation efforts are also made more complex. This is a dimension that relates to the "scale" criteria used in the UNGPs.
- Where risks impact a range of interdependent fundamental rights. This dimension can also correlate to scope and remediability considerations.
- Because of the size and scale of a specific provider's digital services, as indicated by the overall numbers of users as a classification mechanism. While the DSA's risk assessment provisions apply by definition to "very large" providers, there are still significant relative differences in size among them.
- Because of the nature of one or many digital services as a public space, where and/or when activities of significance for fundamental rights are conducted.
- Because they have impacts on systems and societies, at a macro level. This was thought to be significant by contrast with risks created for individuals, individual enterprises, or specific businesses.

Examples of risks that may be systemic because they are caused or exacerbated by systems or systemic factors include the following:

- A notable line of discussion related to participant suggestions that risks are systemic where they are caused to a material extent by platform systems. For example, systemic risks may not flow from content itself, but may flow from the way that platform infrastructure or systems expose large numbers of users to that content without users intentionally seeking it out.
- Systemic risk can also occur as a result of the way several distinct but related systems interact with each other, potentially amplifying human rights risks. This is especially complex where some relevant systems may be excluded from the DSA's scope – for instance, a small chat forum or a large messaging service may not fall within scope of the DSA's systemic risk assessment and mitigation requirements, but can nevertheless impact content distribution, as well as potential attempts at intentional manipulation. Advertising intermediary services were also mentioned as relevant.
- Others suggested that risks are systemic where they cannot be dealt with in the ordinary course of events by usual content moderation and content governance mechanisms. Risks are systemic where they require dedicated risk mitigation systems to protect fundamental rights. For example, risks may be systemic where there are deliberate attempts to thwart content moderation and risk mitigation systems, including by intentional manipulation. Participants noted that intentional manipulation of systems featured strongly in the negotiation of the DSA, even if not reflected prominently in its final text.
- Participants observed that systemic risk can be created by situations where systems for mitigating risk fail. It was also noted that content moderation itself may also create systemic risk by undermining freedom of expression or other rights, if it is not properly designed, calibrated, or applied.

Other methods of defining systemic risk
While some participants pointed to ambiguities in the definition of "systemic risk", others emphasised the view that systemic risk was a legal concept defined comprehensively by a closed list in Article 34. It was noted that, in terms of the paragraphs in Article 34(1), there is established jurisprudence on the concepts included in paras (a) and (b), while there is greater legal uncertainty regarding paras (c) and (d).

In this regard, several participants emphasised the relevance of existing good practice and tools for prioritising risks under the UNGPs and OECD Guidelines.

Isolating causative impacts in a defined system
Regardless of how "systemic risk" is defined as a concept, it is also important that companies, stakeholders and regulators are able to set a theoretical scope for the relevant "system" being assessed. Participants noted it was difficult to assess whether a system had been impacted without a clear understanding of the boundaries and interdependencies of the relevant system.

Relatedly, in order to assess the relative impact of various factors within a system, it is necessary to be able to isolate and understand those factors. For example, causative assessments of the risks created by or to systems, or to fundamental rights, may require methods that can examine the relative impact of platform design, specific content types, coordinated intentional manipulation, user behaviour and sensitivities, and other society-wide factors. In this regard, to assess "systemic risk" on a platform-by-platform basis may be inadequate for assessing "system risks" at a society-wide level.

Importance of stakeholder engagement
Participants expressed a view that definitions of systemic risk and methods for its assessment will require refinement over time. Specifically, comparisons with financial services regulation have limitations: it was noted that failure in the banking system is easy to define, but systemic failures in media pluralism or public discourse may be more difficult to describe. Participants emphasised that effective and comprehensive stakeholder engagement would play a crucial role in systemic risk definition. Given the complexity of the systemic risk concept, meaningful and informed engagement with diverse communities and a range of perspectives is necessary.

Priority fundamental rights

Article 34(1)(b) requires companies to assess systemic risk to "any actual or foreseeable negative effects for the exercise of fundamental rights". It then goes on to list a specific set of rights "in particular". Participants noted that, while this is sometimes seen as "the human rights paragraph", human rights are woven throughout the DSA and the other paragraphs of Article 34.

Participants described the negotiation process that informed the drafting of the DSA. The DSA was initially framed narrowly by reference to rights to privacy and freedom of expression, before being expanded to include specific interests (such as media pluralism, for example), and approaching a general human rights impact assessment obligation. Ultimately, a number of specific rights were included against a broad implied background obligation to consider all fundamental human rights in the European Charter of Fundamental Rights, including the fundamental right to human dignity in Article 1. An explicit goal of the DSA is to ensure people can effectively exercise all of the rights set out in the Charter. Participants observed that some rights (such as the right to own a business, or rights to consumer protection) were included in the Charter, but not in the Universal Declaration and associated instruments, which might mean that Human Rights Impact Assessment methodologies developed and refined over the last two decades under the UNGPs framework might need to be adjusted to the European context.

Balancing rights
Fundamental rights in the Charter cover a broad range of potential impacts, and they are interdependent and interconnected. While no rights are explicitly more important than others, in practice, some balancing and prioritisation is required. This prioritisation exercise can draw on stakeholder engagement frameworks as discussed above, as well as human rights impact assessment frameworks. The kinds of rights to be considered will also need to follow the specific features of providers' products and services, as well as the context in which they are deployed. Participants noted that baseline "human rights saliency" assessments were necessary to identify relevant human rights, and that some companies already conduct these assessments. Other additional rights of relevance identified by participants included rights to life, to liberty and security, to non-discrimination, to freedom of assembly and association, and to an effective remedy, in addition to other categories of rights, such as environmental and labour rights.

It was also noted that providers' products and services generate complex impacts on fundamental rights: for example, the same features of the same product or service (such as content moderation in social networks) can both protect and promote human rights, as well as hinder them. Further complexities can result from potential cross-border impacts: for example, content may be illegal in one member state, but permitted in others. Participants discussed how tensions among fundamental human rights should be resolved in such instances. Additionally, some participants noted that the breadth of potential human rights considerations may present a temptation for political leaders and other interest groups to use risk assessment processes as leverage to drive greater attention and resources to specific priorities, such as copyright infringement.

Other points raised by participants related to the foreseeability of risks to fundamental rights. In this regard, it was noted that proportionality is an important consideration

for mitigating the risk that companies are held to an impossible standard of identifying and preventing risks, which even careful stakeholder engagement and risk assessment could not have helped them anticipate.

Designing for rights
Given the complexity of this exercise, and the need to generate confidence in how it is conducted, participants emphasised the importance of careful system and process design. This includes documentation of key decisions, processes for identifying and escalating potential problems, and associated transparency around these processes and their outcomes. Decision-making processes around risk management must capture relevant information and include relevant people with specific expertise in the matter at hand. It was noted that various decision-making processes may already exist, but in a voluntary compliance context, and that as a result of the transition to a legal compliance context, aspects of these processes are facing significant change. More specifically, these documentation processes are now compulsory under Article 41, with fines for non-compliance, and they will be subject to independent audit and assurance.

Proportionality and prioritisation exercises
Identifying and prioritising among rights and risks requires consideration of nuanced issues such as proportionality, necessity, and legality, and documenting these assessments enables independent scrutiny by external stakeholders of how relevant decisions were reached. Different parties may have reasonable disagreements about how these exercises should be conducted, and this may pose particular challenges for auditors, the Commission, and external stakeholders seeking to objectively and comparatively evaluate these approaches. Similar decisions may also have differential impacts depending on the specific service and where a company or provider sits in the overall "stack". There was a strong sense among the participants that tremendous value could come from further collective discussion across stakeholders about the methodologies used for assessing risks, the strategies deployed for addressing them, and the standards and indicators against which those efforts should be measured.

Transparency and disclosure
Transparency was a strong theme in the discussion. Participants suggested that platform compliance burdens could be mitigated by greater disclosure of information, including through Article 40 researcher access mechanisms. This would enable researchers to flag relevant systemic risks and participate in monitoring and mitigation. Wide-ranging, proactive transparency was suggested as one method of mitigating the need to perform complicated compliance actions. Equally, participants suggested that broad transparency would help platforms demonstrate compliance with a range of interconnected frameworks being developed across the world, and in the Union itself.

Timeliness of published risk assessments
Another notable concern identified in the discussion related to the way that risk assessment reports would be published. The DSA provides little guidance on the time period between companies submitting risk assessments to the Commission, and the publication of those reports to the public. This left participants wondering whether risk assessment reports may be already out of date by the time they are published.

Market assessments create difficulties
Among the identified rights in Article 34(1)(b), the obligation to consider systemic risks on media pluralism received particular attention. To assess systemic impact on media pluralism requires platforms to assess a whole market, including what voices are represented in the public sphere, relevant balances of power, the diversity of various entities, and the ownership of media channels. Issues such as market concentration were raised in the negotiation of the DSA, but, more importantly, have become prominent in the context of obligations among gatekeepers under the Digital Markets Act (DMA). Participants suggested these kinds of assessments were particularly complex for companies given their own roles as market actors and commercial entities, and it was also a specific example of a situation where reasonable minds may differ on "what good looks like", or how market failure should be defined.

Links with wider risk frameworks
Along with links to the DMA, participants observed that the DSA is one part of a broad network of related laws that will require providers to perform and disclose information about human rights due diligence and risk assessment processes and results, including the AI Act, the Corporate Sustainability Reporting Directive, and the Corporate Sustainability Due Diligence Directive. Outside the European Union, there are forthcoming compliance regimes being developed in many countries, including Australia, Brazil, India and the United Kingdom. These frameworks also include obligations of independent audit, human rights risk assessment, and stakeholder engagement to varying degrees.

Participants expressed concern about the interaction between the DSA and other European and global risk assessment frameworks in two important respects. First, an emphasis on compliance-based approaches could make it difficult for companies to engage in open and trusted discussions with relevant stakeholders about issues of systemic risk and compliance, out of concerns for unauthorised or unintended disclosures, or legal and compliance risk. Second, given the burden of compliance with various frameworks, companies may choose to treat these legal frameworks as a kind of maximum, or "ceiling", for their human rights assessment obligations, rather than a minimum, or "floor", for their human rights practices.

Given the legal risks created by diverging from established compliance processes, companies may also be reluctant to explore creative or innovative methods of protecting fundamental rights that go beyond their minimum obligations. The discussion concluded by noting that the framework for conducting DSA risk assessments around fundamental human rights will lead to a new, long-term, institutional-level framework which will need to be developed in an iterative fashion over time. Participants expressed a desire for credible sources of guidance, recommendations, or specifications for appropriate methodologies and systems, as well as renewed commitment to effective stakeholder engagement.

Illegal content

Participants acknowledged that it will never be possible to prohibit or eliminate all illegal content (defined broadly by Recital 12 as "information relating to illegal content, products, services and activities"), just as it is impossible to 100% enforce content moderation rules. Some emphasised that this understanding needed to be built into risk assessment frameworks, as did the potential risks associated with "over-compliance" of efforts to detect and address illegal content. It was also noted that, while providers may have engaged in assessment in this area previously, new allocation of resources and infrastructure may be necessary to meet their obligations under the DSA. Some participants reminded the audience that while specific laws will create particular requirements and risks, the overarching framework through which the DSA encourages VLOPs/VLOSEs to address these is by identifying and focusing on "systemic risks".

Challenges to effectively assessing risks associated with illegal content that were discussed included:

- Measuring the "impact" of illegal content, given that current tools available to platforms have largely not been designed for this purpose.
- Defining how potentially illegal content is dealt with, given that content moderation decisions are generally made with reference to platform guidelines, not the law directly.
- Determining when illegal content risks become "systemic", and differentiating between external influences or trends and structural design risks.
- Accounting for cross-platform proliferation, networks, and campaigns. While this can be complicated by the differing policies, practices, and emphases of different platforms, participants emphasised its importance.
- The differences in the ways that different types of illegal content are disseminated. For instance, hate speech and terrorist and violent extremist content (TVEC) are often publicly disseminated (usually most damagingly by high-profile "influencers"), while child sexual abuse material (CSAM) tends to be shared through private communications, which are not covered by the risk assessment provisions of the DSA.
- Identifying and adapting to novel actors or risks. Many participants noted the benefits of addressing novel crimes and harms with existing measures, while acknowledging that this may not be possible in all instances.

Political conflicts

Some participants pointed out that illegal content has been and likely will be used in the future as a tool for political actors to achieve goals that may not always be consistent with fundamental rights, and there was agreement that cross-stakeholder and cross-border pressure and negotiation were inevitable. While it was agreed that there is increasing harmonisation of laws and definitions on some issues, such as TVEC, divergences remain, especially around issues like hate speech and disinformation.

There were concerns that legal divergence would continue or even intensify as individual jurisdictions create new categories of illegal content. In the interim, uncertainty about how to enforce the DSA was seen as likely. The DSA specifies Union law *and national law that's in line with Union law*, and it was noted that there may be legal disputes about what is in line with Union law (in this context the Conseil Constitutionnel's decision on the Avia Law in France was referenced). Questions were raised about how providers should address such uncertainty in their risk assessments.

Participants also saw a clear distinction between "content" and "conduct", with some reading the DSA as primarily focused on the former, while others thought that conduct was incorporated implicitly and by reference. There were also concerns about how this would apply to the "very politicised" child protection space. Resolving these contradictions and tensions was agreed to be a significant challenge, and a risk to platforms that some participants felt should be examined and reflected in their risk assessments and associated reporting.

Development and coordination
Participants agreed that platforms cannot conduct individual risk assessments for all countries in the EU and all content types, and work would be necessary to align requirements and streamline processes across countries in the EU as well as across providers and requirements that might exist or emerge in other jurisdictions. They also agreed that these processes and assessments were likely to grow more sophisticated and comprehensive over time. Comparisons were drawn to reporting under the EU Disinformation Code of Practice, which was seen as already having matured significantly in its short lifespan. Case studies from other industries were seen as a valuable tool to accelerate this evolution.

Mechanisms and forums that allow the providers to come together with each other and other stakeholders to work on shared challenges were agreed to be crucial. This includes the efforts like the Christchurch Call, DTSP, GIFCT, and GNI, as well as collaborative efforts aimed at addressing CSAM. There was also interest in exploring possible forums or "early warning systems" to discuss illegal content risks, jurisdictional conflicts, and legal challenges more broadly, and that "crisis protocols" and "codes of conduct" in Articles 45 and 48 could be relevant.

Civic discourse and electoral processes, and public security

This category was seen as more diffuse and vague than "illegal content", posing challenges to assessing the issue consistently and at scale. However, there was a sense that risk assessments regarding these issues could nonetheless be performed effectively, especially if it is focused on "systemic risks". Given perceptions of growing societal and political polarisation, including the endorsement of violence as a political tool, demonisation of vulnerable minority groups, and the record number of elections coming up in 2024 (65 elections across 54 countries), this was seen as a particularly urgent area to address.

Measuring risks to civil discourse
Participants noted that risk assessment was complicated by the fact that high engagement audiences, even if relatively small, can generate a lot of risk. For instance, only a tiny fraction of Americans or Brasilians took part in the January 6th and January 8 riots, respectively. If a small fraction of users generate a high prevalence of extremist/hateful content, participants agreed that this can pose a real risk to society.

Some participants argued for the importance of relative metrics and trend lines. More and more radicalising rhetoric could be a signal that risk is high or growing, while a reduction could indicate that mitigations are being effective. Some proposed utilising the UNGPs-based methodology of considering scope, scale, and remediability to determine severity of the risk in tandem with likelihood/probability. Engineering systems were seen as easier to build than human systems that were considered equally necessary, and all participants emphasised the necessity of taking a robust and holistic approach to outcomes and systemic risks.

Article 40 (data access) was seen as an important complement to and check on the self-assessments and self-audits required by the DSA, though participants also noted the need to avoid creating significant expectations or burdens for researchers and civil society without also granting them the necessary resources.

It was also noted that the structure of different online platforms individually, as well as collectively, can allow for both the acceleration and mitigation of risks. In this sense, it will be important for providers to be able to accurately and quickly take the "temperature" of public debate and enable mechanisms that can introduce friction or facilitate positive interventions to address scenarios that become overheated.

Cross-border concerns
Some participants raised concerns of a "dark Brussels effect" – the DSA having negative unintended consequences for non-EU nations. The increase in provider responsibilities in the EU will require them to invest a lot of resources in Europe, which has the potential to limit the amount of resources available for risk assessment and mitigation elsewhere, including countries at significant risk for authoritarianism and political violence. Others offered a more optimistic view: that responses to the DSA will help define what responsible investment looks like, allowing effective practices to be replicated globally.

Interplay between Terms of Service and Public Law
There was a useful discussion about the extent to which the DSA gives providers a role in making important determinations through their terms of service that can take a public law-like form (for instance, in determining what sort of discourse is considered civil). At the same time, by

requiring providers to conduct risk assessment of those same terms and their systems for enforcement, the DSA creates opportunities for broader, collaborative, multi stakeholder discussion and determination of what levels of risk are appropriate and how best to balance the respective risks of over and under-enforcement.

Platform decision-making was acknowledged to be inherently political, having the possibility to affect political outcomes across many jurisdictions, each of which may have very different contexts (for example, some have very limited windows for sanctioned election campaigns, while in the US there is a perception that "there is no such time as a non-campaign time"). Systemic risk was seen as having to capture the risks across (and created by interactions between) all operating contexts.

Terms like "civic discourse" and "election integrity" are not defined clearly in the DSA, and while noting that this was deliberate, participants expected some degree of political conflict and a "tug of war" about their definitions and applications, with some parties trying to define them to their own benefit. There was agreement that civil society needed to make its voice heard in these discussions, and that definitions would likely evolve over time.

Civil society was also seen as a crucial participant in discussions around how much polarisation or "extremism" is necessary in political discourse. A certain level of radicalism was agreed to be necessary to allow for protest and social change. For instance, it was noted that at some point suffragists were considered "extremists". Participants agreed it should not just be up to companies to decide what the acceptable level of polarisation or "extremism" is. Participants also expressed a desire to go beyond assessing risk, to also define what "high quality" content means and consider the extent to which platforms can also promote positive actors, public debate, and responsible journalism.

Overall, there was a sense that providers should not be expected to get everything right immediately, and that we must allow room for good faith experimentation and evolution of risk assessment and mitigation. In addition, further discussions, like the ones facilitated by these workshops, are necessary and important for co-developing appropriate understandings of "what good looks like", both in terms of the substantive benchmarks of progress and the procedural mechanisms that can help achieve them.

Gender-based violence, protection of public health and minors, serious negative consequences to physical and mental well-being

For this discussion, participants focused on gender-based violence, recognising that more time would be needed to consider the full range of covered risks. Participants acknowledged that gender-based violence (GBV) is a broad and challenging category – definitions often include acts that involve sexual, physical, mental, and economic harm in online or offline spaces. This can include online harassment and abuse, stalking, threats of violence, non-consensual image sharing, and doxxing. A common outcome of being targeted with GBV is trauma, which undermines the physical and mental wellbeing of victims. Much of this behaviour is illegal, meaning that risk assessment of GBV has implications and overlaps with risk assessment of illegal content.

Connections were also drawn between GBV and civic discourse, due to the former's tendency to exclude people from public participation. Participants discussed research showing that female politicians and journalists often hold back from public engagement due to their experience of GBV. Additional research shows that witnessing this targeting makes young women more reluctant to enter these professions/spaces.

It was noted that the grammar and syntax in Article 34(1)(d) was somewhat confusing ("any actual or foreseeable negative effects in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being"). One participant noted that they read this sentence as identifying three related but distinct risk categories: GBV; protection of public health and minors; and impacts on physical and mental well-being.

Intersections

Online misogyny was seen as something of a "gateway hate" that often leads individuals to other forms of hatred and violence, such as racial or LGBTQI+ hate. Participants noted that extremists can use gender resentment and violence as a recruiting and radicalisation tool, and that this was important to be aware of when assessing the risks associated with GBV.

Participants agreed that understanding the intersection of different identities, including gender, race, sexuality, and religion, was crucial to capturing the real risks and impacts of GBV on discourse, politics, and online engagement for key sectors of the population. Research on GBV shows disproportionate impacts on people in intersectional spaces (e.g. women and girls internationally, if black or from ethnic minorities, disabled, or LGBTQI+, are more likely to suffer negative impacts).

An overlap between abuse and mis/disinformation was also noted. There is an emerging body of work on "gender disinformation", which is disinformation promoting negative narratives around women and trans individuals, and seeks to reinforce bias. This intersection between false and abusive information was of particular concern to several participants.

The potential relevance of EU Member State laws on domestic violence in both defining and assessing GBV-related risks was also noted.

Difficulties

It was agreed that studying and assessing GBV is often difficult, due to a number of factors:

- Defining "harm" and "violence" can be challenging. Distinguishing between "insults", which may be allowed by terms of service and the law, and "attacks" that would be considered inappropriate or illegal is a highly delicate and contextual exercise that depends on factors such as cultural context, public prominence, role, and history.
- Harassment for particularly targeted minority groups can be so pervasive they don't report it, simply accepting it as "the rules of the game".
- Access to information relating to GBV, and especially its impacts, can be sensitive for platforms and carry privacy risks for users. This speaks to the rights-balancing exercise platforms will need to undertake simply to assess risk.
- Important intersectional categories such as trans and non-binary people are under-researched or simply not captured in available data.
- GBV is almost by definition "systemic" in that it is reflected in and perpetuated by larger societal systems. As a result, risks and impacts on one platform will very often be related to risks and impacts embedded in broader systems and can often be linked to multi-platform abusive campaigns, making it difficult for any individual platform or research team to fully assess or mitigate them.
- Definitions, rules, and processes for protecting "health" and mitigating the harms of GBV are not always clear across – or even within – platforms.
- We all bring our personal views to gender, and therefore, how we approach GBV. It's important to include these perspectives for people moderating these kinds of views too. This reflexive component is very difficult to tackle from a risk assessment perspective.

Participants also noted that while fairly well-established cross-platform and multistakeholder initiatives exist to help address risks faced by providers regarding extremism or government demands, there were not many established initiatives active on GBV.

"Protection of public health and minors" was seen as a potentially controversial issue, due to discussions around the pandemic, arguments about the potential for platform design to be inherently harmful or addictive (especially to children), and the risk of particular groups and governments misusing components of the DSA to persecute minority groups, for instance labelling any LGBTQIA+ content – or even mentions – as "harmful to minors". Stakeholder engagement was seen as key to addressing the above challenges – in order to have the best possible understanding of the risks requires actually consulting with the groups that are facing the risks in a disproportionate way.

List of participants

Participants attended under the Chatham House Rule. None of the comments or observations in this document can or should be attributed to any participant listed below. The discussion summary reflects understandings by GNI, DTSP, and Brainbox of participant comments, but not the position of participants themselves.

Aaron Tidman, Pinterest

Abby Lawson, Integrity Institute

Adeboye Adegoke, Paradigm Initiative

Alexander Hohlfeld, Stiftung Neue Verantwortung

Alexandre de Streel, University of Namur and CERRE

Alexandria Walden, Google

Allyn Robins, Brainbox Institute

Anagha Krishnan, Apple

Andrea Calef, UEA/CCP

Anita Househam, Telenor Group

Anita Williams, Google

Apar Gupta, Internet Freedom Foundation

Arturo Carrillo, GW Law School

Ben Scott, Reset Tech

Carla Weitkamp, Microsoft

Caroline Greer, TikTok

Catharina Vilela, InternetLab

Cathrine Bloch Veiberg, Danish Institute for Human RIghts

Celina Bottino, ITS

Charlotte Yarrow KC, Apple

Chris Sheehy, GNI

Courtney Radsch, UCLA ITLP, Article 19, Fellow at CDT

Danielle Osler, Google

David Kaye, GNI

David Sullivan, DTSP

Dhanaraj Thakur, CDT

Dunstan Allison-Hope, BSR

Eduardo Bertoni, GNI-Academic Board Member

Elisabetta Stringhi, Information Society Law Center

Eliska Pirkova, Access Now

Elizabeth DeYoung, LinkedIn

Elonnai Hickok, GNI

Farzaneh Badiei, DTSP

Gabrielle Guillemin, Meta

Haakon Bratsberg, Telenor Group

Hannah Darnton, BSR

Helena Schwertheim, Institute for Strategic Dialogue

Iná Jost, InternetLab

Isabel Ebert, OHCHR B-Tech

Jana Lasser, Graz University of Technology & Complexity Science Hub Vienna

Jason Pielemeier, GNI

Jeff Allen, Integrity Institute

Joan Barata, Senior Legal Fellow, Justitia

Joris van Hoboken, Institute for Information Law, University of Amsterdam

Katherine Sandell, Google

Kathleen Stewart, Meta, Board Member DTSP

Jacqueline Rowe, Global Partners Digital (GPD)

Kathryn Doyle, Global Partners Digital (GPD)

Katie Sandell, Google

Laura Becana Ball, GFMD

Lene Wendland, Office of the UN High Commissioner for Human Rights

Lillian Nalwoga, CIPESA

Liz DeYoung, LinkedIn

Madelaine Harrington, TikTok

Maria Paz Canales, Global Partners

Marlena Wisniak, ECNL

Mathias Vermeulen, AWO

Melanie Walsh, LinkedIn

Melody Patry, Internews

Mina Narayanan, Center for Security and Emerging Technology

Minna Iveson, Pinterest

Mira Milosevic, Global Forum for Media Development (GFMD)

Olaf Steenfadt, Global Media Registry

Ollie Irwin, Google

Paddy Leerssen, University of Amsterdam, Institute for Information Law, DSA Observatory

Pamela Almaguer, Microsoft

Patrick Gage Kelley, Google

Peter Chapman, Article One

Princess Ifon, Google

Richard Gaines, Wikimedia Foundation

Robert Gorwa, WZB Berlin Center for Social Science

Roya Pakzad, Taraaz

Sally Broughton Micova, University of East Anglia and Centre on Regulation in Europe (CERRE)

Sandra Aceng, Women of Uganda Network (WOUGNET)

Sara Bundtzen, Institute for Strategic Dialogue (ISD)

Shahla Naimi, Google

Shobhit Shukla, Centre for Communication Governance, National Law University Delhi (NLUD-CCG)

Tavishi Ahluwalia, Centre for Communication Governance, National Law University Delhi (NLUD-CCG)

Thobekile Matimbe, Paradigm Initiative

Tom Barraclough, Brainbox Institute

Tommaso Venturini, University of Geneva and CNRS

Ximena Smith, Brainbox Institute

List of relevant resources

Participants shared various resources during the discussion, which we have collated here for convenience.

General resources

- The United Nations B-Tech project has published a range of useful resources on the various dimensions of human rights due diligence and risk assessment (link here).
- The Action Coalition on Meaningful Transparency has released a briefing note on implementing risk assessments under the Digital Services Act (link here).
- Susan Ness and Chris Riley have been advocating for a "modular approach" to platform regulation to promote alignment among democracies on internet governance. A one-page explanation and links to associated resources is available (link here).

Stakeholder engagement

- AccessNow and ECNL are conducting shared work on stakeholder engagement processes required for Article 34 risk assessments, with details to be published shortly.
- The United Nations OHCHR has published a paper on five practices to improve stakeholder engagement by tech companies (link here)
- ECNL has published a framework for stakeholder engagement on AI systems (link here) with a related blog post summary (link here) and a summary of key takeaways from a closed-door consultation at the UN Forum on Business and Human Rights (link here).
- The Danish Institute for Human Rights has prepared guidance, with stakeholder input, on human rights impact of digital business activities, with ten key criteria (link).

What is systemic risk?

- A report by the World Economic Forum identifies various examples of risk assessment frameworks being developed and implemented in the digital realm (link here).
- Risk assessment obligations for digital services are also being considered in jurisdictions outside the EU, such as the UK and Brazil (link here), and in other European legal frameworks, such as the proposed Artificial Intelligence Act.
- Work is underway with a view to defining systemic risk by reference to banking and financial regulatory systems, with a literature review expected from the Centre on Regulation in Europe (CERRE) in July 2023.
- The Centre on Regulation in Europe has published a discussion document examining factors that influence relative risk aside from the relative size of a digital service (link here)
- GNI and BSR have published an "across the stack" human rights due diligence tool for key nodes of relationships and issues in the ICT ecosystem (link here).
- Luke Thorburn, Jonathan Stray, and Priyanjana Bengani have published an analysis of different methods of determining cause and effect when it comes to assessment of recommender systems (link here).

Priority fundamental rights

- The Office of the High Commission for Human Rights and the B-Tech initiative have published guidance on identifying human rights risks (available here).
- The Centre on Regulation in Europe has published a comparative analysis of recent initiatives targeting Digital Services in Europe comparing four pieces of adopted and draft legislation that deal with illegal and harmful content on digital services: the rules on video-sharing platforms (VSPs) contained in the Audiovisual Media Services Directive (AVMSD), the Terrorist Content Regulation (TERREG), the Digital Services Act (DSA) and the UK's proposed Online Safety Bill (OSB) (link here).