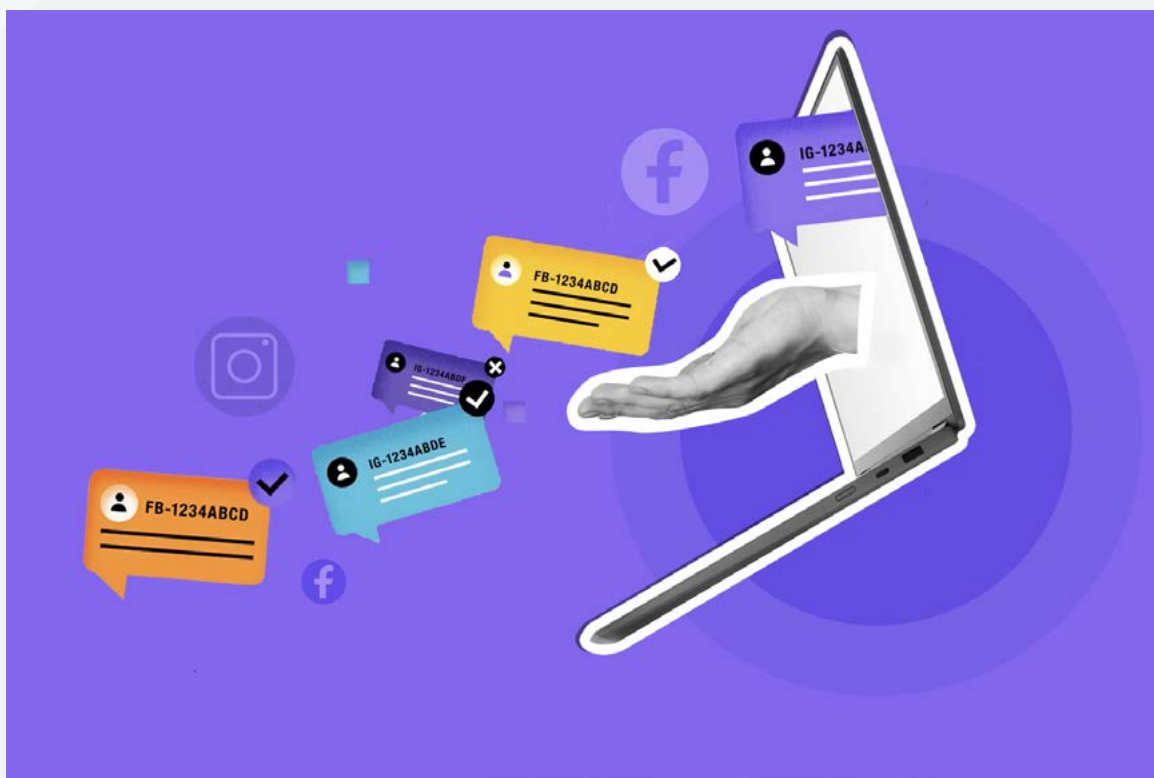


# REASONABLE, PROPORTIONATE AND EFFECTIVE?

Enhancing Respect for Freedom of Expression in Systemic Risk  
Assessments and Mitigation Measures under the European Union's  
Digital Services Act

THE OVERSIGHT BOARD, OCTOBER 2025





# 1. Executive Summary

In this industry-wide analysis, the Oversight Board offers insights to help very large online platforms and search engines respect freedom of expression, within the framework of the European Union's (EU) Digital Services Act (DSA) systemic risk assessments and related mitigation measures.

## 1.1 Context

The preamble to the DSA notes it seeks to, among other things, protect user rights by placing a regulatory requirement on very large online platforms and very large online search engines (hereafter, “designated providers”) to identify and mitigate systemic risks resulting from their online services.<sup>1</sup>

Specifically, DSA Article 34 requires that designated providers “identify, analyze and assess any systemic risks in the Union” arising from the design, functioning and use of their services, including “negative effects for the exercise of fundamental rights.” These have become known as systemic risk assessments.

DSA Article 35 requires that designated providers address the risks identified in systemic risk assessments with mitigation measures that are “reasonable, proportionate and effective,” with “particular consideration to the impacts of such measures on fundamental rights.”

Seeking to ensure that the DSA is interpreted in a way that respects global human rights standards, including those guaranteeing freedom of expression, the Board is concerned that the terms “reasonable, proportionate and effective” are not sufficiently defined. Left unchecked, this may incentivize platforms to pursue compliance through overbroad mitigation measures that adversely impact freedom of expression.

## 1.2. Purpose

This report builds upon [previous analysis](#) published by the Board promoting developments in systemic risk assessments and mitigation measures that enhance respect for human rights, with a particular focus on freedom of expression.

Specifically, this report proposes that determinations of whether mitigation measures are “reasonable, proportionate and effective” should be informed by Article 19 of the International Covenant on Civil and Political Rights (ICCPR), which requires that restrictions on freedom of expression be evaluated under the three-part test of legality, legitimate aim and necessity/proportionality. Based on the Board's experience to date, the three-part test provides a practical, proven and rights-based approach for designated providers, auditors and the European Commission to use. All EU member states are parties to the ICCPR.

<sup>1</sup> Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), preamble, at para. 3.



### 1.3 Current State

The lack of a definition for “reasonable, proportionate and effective” in the DSA creates uncertainty for designated providers and independent audit firms, who have some latitude to reach their own determination.

The Board’s review of systemic risk assessments and their accompanying audit reports reveals that designated providers and auditors (1) struggle to provide analysis justifying why the mitigation measures should be considered “reasonable, proportionate and effective”; (2) emphasize the process rather than the substance of determining whether mitigation measures are “reasonable, proportionate and effective”; and (3) pay more attention to “effective” than to “reasonable” and “proportionate” criteria.

### 1.4 Analysis

This report draws connections between the DSA’s requirement for systemic risk assessments and mitigation measures and the issues addressed by the Board in case decisions and policy advisory opinions to propose an approach to the design of mitigation measures. The Board’s work to date has been Meta-specific, but this analysis aims to provide industry-wide insights, especially for designated providers that moderate speech.

The report examines each principle in turn (i.e., reasonable, proportionate, effective) and uses the Board’s cases to demonstrate how the three-part test (i.e., legality, legitimate aim and necessity/proportionality) can provide a practical approach using rights-based methodologies to help ensure that mitigation measures respect freedom of expression and other human rights. This report builds on prior analysis of DSA systemic risk assessments and mitigation measures undertaken by various organizations.<sup>2</sup>

<sup>2</sup> For example, AccessNow, Centro de Estudios en Libertad de Expresión (CELE), DSA Civil Society Coordination Group, DSA Observatory, DTSP, European Center for Non-Profit Law, The Future of Free Speech, GNI, Integrity Institute, and the Knight-Georgetown Institute.



## 1.5 Conclusions

The Board reaches the following conclusions:

- **Reasonable:** The “legality” and “legitimacy” aspects of the three-part test required by Article 19 of the ICCPR can inform analysis of whether mitigation measures impacting freedom of expression are consistent with the principle of “reasonableness.” Further, an assessment of whether these mitigation measures are “reasonable” should be informed by the analysis of proportionality and effectiveness (below).
- **Proportionate:** Analysis of whether mitigation measures are “proportionate” should encompass the interlinked principles of both necessity (i.e., the least intrusive means) and proportionality (i.e., targeting a specific objective, without unduly intruding upon the rights of others). To ensure consistency with international human rights standards as defined in the ICCPR, designated providers and auditors should consider (1) whether the mitigation measures are “necessary and proportionate” to address the relevant systemic risk broadly and (2) whether the mitigation approach gives rise to “necessary and proportionate” measures on a case-by-case basis. The latter could be achieved by reviewing a sample of cases across different contexts.
- **Effective:** Analysis of whether mitigation measures are effective should encompass (1) relevant quantitative metrics; (2) feedback from affected stakeholders; and (3) evidence of whether mitigation measures are being implemented in an equitable and non-discriminatory manner, such as across languages and dialects. Additionally, human rights-based approaches imply that the “effectiveness” of mitigation measures should consider impacts on all global users, not only users in the EU. Finally, an analysis of effectiveness is relevant for reviewing whether a mitigation measure is the least intrusive means of achieving a legitimate aim.

The Board looks forward to further engagement with regulators, designated providers, auditors and other stakeholders on how best to ensure systemic risk assessments and mitigation measures enhance respect for freedom of expression and other human rights.





## 2. Introduction

### 2.1 Key Questions

This report builds upon the Board’s [previous analysis](#), promoting developments in systemic risk assessments and mitigation measures that enhance social media companies’ respect for human rights, particularly freedom of expression, both as an individual right and to enable other human rights.

Specifically, this report proposes practical human rights-based approaches for determining whether mitigation measures are “reasonable, proportionate and effective.” In doing so, it responds to a need arising from initial rounds of systemic risk assessment reports published by designated providers under the DSA. Two questions are central to the goal of ensuring respect for human rights, especially freedom of expression:

- How can DSA Article 35’s requirement that mitigation measures are “reasonable, proportionate and effective” be informed by the ICCPR Article 19 requirement that restrictions on freedom of expression be evaluated under the three-part test of legality, legitimate aim and necessity/proportionality?
- What tensions, trade-offs or tests should be considered when assessing whether different mixes of mitigation measures are “reasonable, proportionate and effective” and when evaluating the impact of mitigation measures on human rights, in particular freedom of expression?

The resolution of these questions is essential for placing freedom of expression at the core of systemic risk assessment and mitigation, thereby achieving the DSA’s goal of creating a trusted online environment where human rights are respected.

This report will draw connections between the issues addressed by the Board in its individual cases and the systemic risks addressed by the DSA to inform rights-respecting mitigation measures. By drawing on the Board’s practical experience, this report provides designated providers and auditors with workable methods and offers regulators insights to strengthen future guidance. The Board’s cases to date have been Meta-specific, but this analysis aims to provide industry-wide insights, especially for designated providers that host, moderate and disseminate user-generated speech.

This report addresses mitigation measures relating to the development and enforcement of content policies as part of the features, functionalities and systems of a designated provider’s service. The Board approaches the broad field of content moderation with a focus on what may impact freedom of expression and believes that the three-part test required by ICCPR Article 19 should be applied to all mitigation measures that may impact this right.

This report builds on prior analyses of DSA systemic risk assessments and mitigation measures, such as those undertaken by Access Now, Centro de Estudios en Libertad de Expresión (CELE), DSA Civil Society Coordination Group, DSA Observatory, Digital Trust and Safety Partnership (DTSP), European Center for Non-Profit Law, The Future of Free Speech, Global Network Initiative (GNI), Integrity Institute and the Knight-Georgetown Institute.



## 2.2 The DSA and International Human Rights Law

A comparison between DSA requirements and approaches to content moderation consistent with international human rights law (IHRL) principles, as defined in the ICCPR, provides the backdrop for the analysis provided in this report.

DSA Article 34 requires that designated providers “identify, analyze and assess any systemic risks in the Union” arising from the design, functioning and use of their services, taking into consideration their severity and probability, and including “any actual or foreseeable negative effects for the exercise of fundamental rights.” These have become known as systemic risk assessments.

DSA Article 35 requires that designated providers address the risks identified in systemic risk assessments with mitigation measures that are “reasonable, proportionate and effective,” with “particular consideration to the impacts of such measures on fundamental rights.”

The Board believes that these DSA provisions should be interpreted and applied in a way that is compatible with IHRL, including in particular the right to freedom of expression. Because the terms “reasonable, proportionate and effective” are not defined and the DSA relies on independent audit firms to reach a determination using information and benchmarks provided by designated providers, the Board is concerned that this ambiguity may incentivize overbroad mitigation measures with adverse impacts on freedom of expression.

DSA Article 35 lists 11 mitigation measures that “may” be used by designated providers “where applicable” to address identified risks, such as: adapting the “design, features or functioning of their services,” adapting “terms and conditions and their enforcement,” adapting “content moderation processes,” testing and adapting “algorithmic systems,” or “taking awareness raising measures.” However, the first systemic risk assessment reports indicate that many designated providers are using these suggestions as a checklist of pre-determined mitigation measures, or even requirements, against which to assess compliance.



A key theme of this report is evaluating how the definition of “reasonable, proportionate and effective” can be informed by the three-part test of legality, legitimate aim and necessity/proportionality of ICCPR Article 19.

In answering this question, the Board’s experience in interpreting and applying Article 19 to the evaluation of content restrictions can help inform how the DSA’s requirement for “reasonable, proportionate and effective” mitigation measures can be achieved in practice. This will build on pre-existing frameworks that enable a rights-respecting approach to assessing systemic risks and advance the DSA’s goal that designated providers “give particular consideration” to freedom of expression when deciding how to address systemic risks (DSA Recitals 86 and 90).

The Board acknowledges that IHRL applies primarily to states rather than companies. However, the [United Nations Guiding Principles for Business and Human Rights \(UNGPs\)](#) establish the responsibility of companies to respect internationally recognized human rights and not infringe upon the rights of others. Most designated providers have made a public commitment to the UNGPs (for example, [Meta](#), [Google](#), [Microsoft](#) and [TikTok](#)), and it is in this context that the Board draws inspiration from the UNGPs to place the three-part test at the center of its case decisions and recommendations.

The Board achieves this goal by using methods grounded in authoritative guidance on how to interpret ICCPR Article 19, most notably [United Nations \(UN\) General Comment No. 34](#) (CCPR/C/GC/34) and subsequent reports published by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. As the UN Special Rapporteur has stated, although “companies do not have the obligations of governments, their impact is of a sort that requires them to assess the same kind of questions about protecting their users’ right to freedom of expression” ([A/74/486, para 41](#)). Further, because the DSA represents government intervention into company content moderation, and because the DSA is subject to IHRL obligations, it is crucial for designated providers to ground their mitigation measures in IHRL.

The approach used by the Board is based on the following standards and principles:

- **ICCPR Article 19 (2)** protects everyone’s right to seek, receive and impart information and ideas of all kinds, regardless of frontiers and through any media. Article 19 (3) requires that any restrictions on freedom of expression be provided by law, have a legitimate aim and be necessary for the achievement of that aim.
- **Legality** means that any restrictions on freedom of expression must be provided by law, which means such limitations must also be formulated with sufficient precision to enable an individual to regulate conduct accordingly and provide appropriate guidance to those implementing the law (UN General Comment No. 34, paras. 24 – 25). In the context of content moderation by designated providers, the Board interprets this to mean two things. First, content policies and other communications should provide users with sufficient clarity, specificity and transparency to understand with reasonable certainty what expression is allowed on the service and/or may receive reduced or boosted visibility ([A/HRC/38/35, para. 46](#)). Second, content policies and internal enforcement guidance should also provide content reviewers with clear guidance to ensure their consistent and non-arbitrary enforcement.



- **Legitimate aim** means that any restrictions on freedom of expression must pursue a legitimate aim listed in the ICCPR, specifically the rights and reputation of others and the protection of national security, public health, public order and morals (ICCPR Article 19(3); UN General Comment No. 34, paras. 28 – 32). In the context of content moderation by designated providers, the Board interprets this by assessing whether the speech restriction and its stated aim are consistent with one or more of these public interest objectives. Section 4.1 of this report considers how mitigation measures should account for restrictions aimed at the specific purpose of the service, in addition to or instead of a public interest objective listed in ICCPR Article 19 (3).
- **Necessity and proportionality** mean that restrictions on expression “must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; they must be proportionate to the interest to be protected” (UN General Comment No. 34, para. 34). In the context of content moderation by designated providers, the Board interprets this to mean only restricting content when the same goal cannot be achieved by less intrusive means, such as limiting the visibility and reach of content rather than removing it altogether, and that the burden on freedom of expression is lower than the benefit achieved by restricting the right (A/HRC/38/35 para 47).

Furthermore, any restrictions on freedom of expression must not violate the principle of **non-discrimination**, for example, by seeking and taking into account the concerns of communities that have historically been at risk of censorship and discrimination (A/HRC/38/35, para. 48).

Necessity and proportionality intersect. Applying both the “least intrusive means” and “proportionality” tests provides a practical methodology for designated providers to meet their responsibility to respect freedom of expression and other human rights globally.





### 3. Current Practice for Complying with DSA Article 35

The absence of a clear definition for “reasonable, proportionate and effective” in the DSA creates uncertainty for designated providers and independent audit firms, which have some latitude to reach their own determination. A review of all the designated provider systemic risk assessments and audit reports available at the time of publication reveals four main themes:

- Several designated providers seek to define “reasonable, proportionate and effective,” but these are outliers rather than the standard practice.
- Designated providers and auditors struggle to provide analysis justifying why the mitigation measures should be considered “reasonable, proportionate and effective.”
- Auditors focus on the process steps used by designated providers, rather than the substance of their analysis.
- More attention is paid to “effective” than to “reasonable” and “proportionate” criteria.

#### 3.1 Defining Reasonable, Proportionate and Effective

There are two main examples of designated providers explaining how they consider the principles of “reasonable, proportionate and effective” in the context of their services. Meta (i.e., Facebook and Instagram) introduces its own definitions of “reasonable, proportionate and effective” that are used during systemic risk assessments and mitigation measures (see the table below). These definitions provide an indication of how mitigation measures have been assessed and contain some overlap with the three-part test. The following table is taken directly from Meta’s public systemic risk assessment reports:

Criteria	Mitigation Measure	Further Details
Reasonable	<ul style="list-style-type: none"><li>- Within Meta’s control to deploy with limited dependencies on external parties or non-Meta entities</li><li>- Appropriate, fair, and designed to address integrity risks or issues</li></ul>	Due to the residual risk exposure and/or the extreme criticality of a control in managing a systemic risk, it is appropriate to make investments to adapt, test, reinforce, initiate, adjust, and/or make changes to our systems, processes, and/or activities.
Proportionate	<ul style="list-style-type: none"><li>- Adequate, relevant, suitable and necessary to address specified systemic risks</li><li>- Not excessive in relation to a declared and specified purpose and residual risk exposure</li></ul>	The investment needed from a financial, technical, and operational perspective is commensurate with the current risk exposure or the risk exposure that will be created if the investment is not made. Additionally, in instances where rights, including fundamental rights, are in tension with a potential mitigation measure, a decision about a mitigation measure is based on the correlative impact a risk could have on users within the EU and society.
Effective	<ul style="list-style-type: none"><li>- Able to prevent, mitigate, or control the residual risk exposure as designed and intended</li><li>- Able to be monitored in order to measure its effectiveness</li></ul>	The investment needed to adapt, test, reinforce, initiate, adjust, and/or make changes to our systems, processes, and/or activities will effectively reduce the residual risk exposure of a systemic risk.



Google (i.e., Google Search, Play, Shopping and Maps, and YouTube) provides several narrative explanations for how it considers the “proportionality” of its mitigation measures. While Google does not give specific definitions, it offers a discussion on how necessary and proportionate removals may vary according to content, such as whether content is indexed or hosted, public or private, direct or indirect, recommended, or monetized. The question of how the outcomes of the three-part test (i.e., legality, legitimate aim and necessity/ proportionality) may differ across services with different purposes and characteristics is considered in more detail in section 4.1.

Elsewhere, the Google systemic risk assessment report discusses use of the proportionality principle when rights are in tension (e.g., freedom of expression and the rights of children), how proportionate mitigation measures may vary according to the purpose of the service (e.g., YouTube versus Maps) and how some mitigation measures may be more proportional than others (e.g., restricting access to content via interstitial warnings that appear before a user can access content versus removing content from the service entirely).

Finally, two of Google’s 40 risk statements (i.e., descriptions of risk that form the basis of Google’s systemic risk assessments) incorporate two parts (legitimacy and necessity/ proportionality) of the three-part test when assessing risks to freedom of expression:

- “Risk that a service removes content that does not constitute a necessary or proportionate removal of content with a legitimate purpose.”
- “Risk that children’s access to and/or use of a service is limited more than is necessary or proportionate for a legitimate purpose.”

While the attention shown to unpacking “reasonable, proportionate and effective” criteria by Meta and Google is a helpful start, further progress can be made towards approaches more directly grounded in IHRL. This would include more narrative relating to the tensions, trade-offs or tests to consider when assessing whether different mixes of mitigation measures are “reasonable, proportionate and effective” in various contexts.

### **3.2 Explaining their Analysis**

Designated providers tend to conclude that their mitigation measures meet the “reasonable, proportionate and effective” standard by referencing their main mitigation efforts, but without explaining their evaluation using criteria or clearly defined benchmarks.

For example, one designated provider reviewed each risk in turn, listed its mitigation measures and concluded that its services “have reasonable, proportionate and effective mitigation measures” in place for every risk. However, little evidence was presented about what criteria, benchmarks or metrics were used to reach this conclusion. This designated provider also appeared to be evaluating “reasonable, proportionate and effective” as a single and binary criterion, rather than multiple and distinct but interrelated concepts.

Many designated providers conclude that their mitigation measures are “reasonable, proportionate and effective” by simply cross-referencing the 11 sample mitigation measures listed in Article 35 of the DSA as a checklist. They do this rather than explaining why their mitigation measures should be considered “reasonable, proportionate and effective” for their specific circumstances or discussing the potential impacts on human rights resulting from their implementation.



### 3.3 Auditors Assessing Process Steps

Most audit firms focus on the process used by the designated providers to determine whether they have “reasonable, proportionate and effective” mitigation measures in place, rather than assessing the substance or merit of the designated provider’s conclusion.

For example, one audit firm examined the designated provider’s communications, meeting notes and dashboards relating to the process of reviewing mitigation measures, including whether the 11 sample mitigation measures listed in DSA Article 35 had been considered. Another audit firm inspected a sample of meeting notes and documents and reviewed whether the designated provider had considered the impact of mitigation measures on fundamental rights. However, the audit firm provided no analysis as to whether an appropriate mix of “reasonable, proportionate and effective” mitigation measures had been determined for the specific circumstances of the designated provider. The auditors’ focus in both cases was on process rather than outcome.

This theme also arose during the July 2025 [GNI and DTSP European Rights and Risks Stakeholder Engagement Forum](#), which emphasized the need to clarify the role of DSA audits and auditors.

### 3.4 Focus on Effectiveness

Systemic risk assessment reports and accompanying audit reports provide more analysis on the term “effective” than the terms “reasonable” or “proportionate.” This assessment of effectiveness typically takes one or both of two forms:

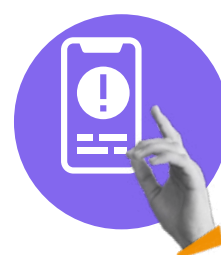
- **Quantitative metrics**, such as the prevalence of policy-violating content, appeals data, proactive detection rates and enforcement accuracy across different languages or turnaround times. These indicators are used by some designated providers as evidence of successful content policy enforcement and to substantiate a lower “residual risk” (i.e., risk after mitigation measures) rather than an “inherent risk” (i.e., risk before mitigation measures). For example, YouTube, Facebook and Instagram use “prevalence” and “violative view rate” metrics to evaluate effectiveness based on how widely policy-violating content is viewed. These metrics can indicate whether mitigation measures achieve some of their goals (e.g., timely and accurate removals) but do not indicate whether mitigation measures are reasonable or proportionate.
- **Controls or mitigations testing**, such as to determine whether mitigation measures are operating consistently, as intended and/or effectively. For example, Microsoft (i.e., LinkedIn and Bing) compares its mitigation measures to the [DTSP Safe Framework Maturity Rating](#), which uses a five-level scale to assess the maturity of a company’s trust and safety practices. Meta incorporates notions of reasonableness and proportionality into its definition of mitigation measure effectiveness and uses a mix of signals (e.g., control assurance/testing results, tracking remediation of known deficiencies) to review operating effectiveness.

The use of quantitative metrics to determine mitigation measure effectiveness is valuable and consistent with the UNGPs’ guidance that effectiveness should be tracked using quantitative key performance indicators (UNGPs Principle 20).



However, taken collectively, there is little evidence that quantitative methods are complemented by qualitative testing of mitigation measure effectiveness with affected stakeholders, and there appears to be little consideration of equity and non-discrimination in the implementation of mitigation measures. For example, designated providers and auditors could test the impact of mitigation measures designed to address certain policy violations (such as those related to hate speech, terrorism or harassment) on users who are most vulnerable to this content, as well as users most at risk of over-moderation. Illustrations of how this may be achieved in practice are provided in section 4.3 below.

Finally, conversations with experts to inform this report indicated that audit firms have greater familiarity with processes to test effectiveness (e.g., controls testing) than with the more nuanced and complex considerations surrounding “reasonable” and “proportionate” (e.g., identifying the least intrusive means). This appears to result in evaluations that scrutinize effectiveness more than reasonableness and proportionality, and/or assume that a process to consider all three principles equates to substantive compliance with them. There may be a case for the European Commission to require auditors to have expertise in IHRL and be able to evaluate the reasonableness and proportionality of mitigation measures, not just their effectiveness. The Delegated Regulation on Independent Audits under the DSA does allow auditors to contract with other organizations (or even a consortium of organizations) where there is a need for specific expertise, such as those relating to human rights (Delegated Regulation on Independent Audits, Recitals 3 and 9).





## 4. Human Rights-Based Approaches to Article 35

This section explores how a human rights-based approach can inform the principles of “reasonable, proportionate and effective” when designing mitigation measures that may impact freedom of expression. Although each principle is reviewed individually, the Board believes they are interconnected and integral to a designated provider’s cycle of ongoing human rights due diligence.

### 4.1 Reasonable

Both the International Court of Justice (ICJ) and the European Court of Human Rights (ECtHR) have affirmed that what is reasonable in any given case depends on its particular circumstances, though reasonableness typically implies notions such as legitimate expectations, certainty, due process, predictability, good administration, and non-discrimination.<sup>3</sup>

#### Reasonableness and the Three-Part Test

The adaptability inherent in the principle of “reasonableness” makes it well-suited to rapidly evolving digital environments and allows it to be used in the context of assessing designated provider mitigation measures under the DSA. At the same time, its open-ended, context-dependent and discretionary nature means that further specificity, based on IHRL, is needed to determine how to apply the criteria in the context of evaluating the impact of mitigation measures on freedom of expression and other human rights.

*Based on the Board’s experience, the principles of “legality” and “legitimate aim” from the three-part test required by Article 19 of the ICCPR can inform the principle of “reasonableness” for mitigation measures that impact freedom of expression.*

It is noticeable how these two parts of the three-part test overlap considerably with the characteristics of reasonableness and enable a direct connection to the context of content moderation:

- **Applying the “legality” requirement** helps ensure that mitigation measures, especially rules, guidance and systems that limit the availability and distribution of content, are formulated with sufficient clarity, specificity and transparency for users to regulate their conduct accordingly, and for content reviewers to understand with reasonable certainty what sorts of expression are properly restricted and what sorts are not (UN General Comment No. 34, para. 25).
- **Applying the “legitimate aim” requirement** helps ensure that mitigation measures affecting freedom of expression are undertaken for one or more of the legitimate aims listed in the ICCPR (i.e., rights and reputation of others and the protection of national security, public health, public order and morals) and that these aims are clearly stated in the designated provider’s content policies.

<sup>3</sup> For example, based on the jurisprudence of the Court of Justice of the European Union (CJEU) and relevant EU legal instruments



## Insights from Oversight Board Cases

A review of the Board's cases illustrates how the use of the "legality" and "legitimate aim" requirements of the three-part test can support the principle of "reasonableness" when considering mitigation measures that may impact freedom of expression. Using the "legality" and "legitimacy" requirements will help designated providers, auditors and other stakeholders understand whether compliance with the requirements of DSA Article 35 has been achieved in a manner consistent with these IHRL principles.

There are several mitigation measures listed in Article 35(1) where prior Board cases illustrate how the "legality" and "legitimacy" requirements can inform an assessment of whether "reasonable" mitigation measures are in place and/or what enhancements may be needed.

For example, DSA Article 35(1)(b) lists "adapting ... terms and conditions and their enforcement" (i.e., content policy) as a mitigation measure. Applying the "legality" and "legitimate aim" requirements will enable designated providers and auditors to review compliance with Article 35(1)(b) in more practical detail.

- **Policy clarity:** In the [Sudan's Rapid Support Forces Video Captive](#) and [Nazi Quote](#) cases, the Board emphasized that rules restricting expression must be clear, precise and publicly accessible, allowing users to adjust their conduct accordingly. The Board made recommendations to improve the communication of Meta's Dangerous Individuals and Organizations policy, address concerns of arbitrary enforcement and inform users which Community Standard they violated when their content was removed.
- **Policy precision:** In the [Statements About the Japanese Prime Minister](#), and [Iran Protest Slogan](#) cases, the Board recommended revisions to Meta's Violence and Incitement policy, and accompanying non-public internal guidelines, to more clearly distinguish between literal and figurative threats and the treatment of "public figures" and "high-risk persons," such as by publishing a general definition of high-risk persons and illustrative examples. The Board also recommended greater alignment between Meta's stated policy rationale and its actual enforcement practices by providing nuanced guidance to moderators on how to consider context.

In [Iranian Woman Confronted on Street](#) the Board recommended that Meta add a policy lever to its Crisis Policy Protocol, providing that figurative (or not literal) statements not intended to, and not likely to, incite violence do not violate the Violence and Incitement policy prohibition on threats of violence in relevant contexts. The Board recommended that criteria be developed for content reviewers on how to identify such statements in the relevant context.

- **Policy transparency:** In [Referring to Designated Individuals as "Shaheed"](#) policy advisory opinion, the Board made recommendations to further improve the accessibility of Meta's rules by clarifying the prohibition on "unclear references" to designated organizations, providing users with examples of violating content and increasing transparency of Meta's designated entities and events list. In [Homophobic Violence in West Africa](#), the Board recommended that Meta update its prohibition on "outing" to include illustrative examples of "outing-risk groups," including LGBTQIA+ people in countries where same-sex relations are forbidden and/or such disclosures create significant safety risks.



- **Policy rationale:** It is important for the substantive objective of content policy to align with a legitimate aim. In [AI-Manipulated Video Promoting Gambling](#), the Board determined that Meta’s prohibition on posts that establish fake personas seeks a legitimate aim because it protects people from scams and fraud (Article 17, UDHR), and protects the privacy rights and reputation of the persons depicted (Article 17, ICPPR). However, in [Russian Poem](#), the Board recognized that while protecting those targeted by hate speech is typically a legitimate aim, protecting soldiers from claims of wrongdoing in the context of a war and their role as combatants (rather than their nationality or another protected characteristic) during times of war is not a legitimate aim.

In [Images of Partially Nude Indigenous Women](#), the Board concluded that privacy rather than “community sensitivity” was the more appropriate legitimate aim and highlighted that any appeal to protecting morals must be based on principles not deriving exclusively from a single tradition.

DSA Article 35(1)(a) lists “adapting the design, features or functioning of their services, including their online interfaces” as a mitigation measure. Article 35(1)(i) lists “adapting their online interface in order to give recipients of the service more information” as a mitigation measure. Here too the “legality” and “legitimate aim” requirements can help designated providers and auditors interpret Article 35(1)(a) and Article 35(1)(i) in more practical detail.

- **Providing users with more information:** In [Reporting on Pakistani Parliament Speech](#), the Board noted how Meta had subsequently improved compliance with the legality requirement by adding an “awareness-raising” exception to the public-facing language of its Violence and Incitement Community Standard that did not exist at the time of the case. This “awareness raising” exception had previously only been included in Meta’s internal guidance to reviewers and was added in line with a Board recommendation in the [Russian Poem](#) case. This change removed a barrier to public interest discussions by making it clearer to users what content is permitted.
- **Informing users about the role and use of warning screens and newsworthy labels:** In the [Colombia Protests](#) and [Sudan Graphic Video](#) decisions, the Board recommended that Meta notify users when content remains on the platform due to a newsworthiness allowance, with the Sudan case also addressing notifications for application of warning screens to content. These cases illustrate how the “legality” and “legitimate aim” requirements can be applied to the online interface and visibility and reach of content, not simply what content is or is not allowed.





### Legality, Legitimate Aim and the Purpose of a Service

The “context-dependent” nature of reasonableness becomes relevant when considering whether, why and how content restrictions may vary across designated providers depending on the purpose of the service.

This is well illustrated by Google, whose services provide strikingly different benefits to users, and where restrictions that are reasonable for one service may not be reasonable for another. Google makes its case as follows:

- **Google Search:** “Content policies for Search are designed to minimize restrictions on freedom of expression and promote access to information. This design means that risks associated with potentially illegal or ‘legal but harmful’ content will always be present with Search because content may still be discoverable if it is available on the internet.”
- **Google Maps:** “The service emphasizes being a source of reliable information and a reflection of genuine user experiences. For this reason, we lean towards user-generated content policies that are designed to maximize the quality, accuracy, and authenticity of information for consumer and merchant user contributions. We go to great lengths to make sure content published by our consumer and merchant users is helpful and reflects the real world, recognizing that this means accepting some attendant limitations to freedom of expression.”
- **YouTube:** “YouTube values freedom of expression and is built on the premise of openness. Its policies aim to support the interest of its creators and their incredible array of diverse voices and perspectives. YouTube is committed to protecting its community from harmful content, while giving creators the freedom to share a broad range of experiences and perspectives through video. Because YouTube hosts and serves user-generated content, it has unique content policies.”



Google's case is that restrictions on freedom of expression are more reasonable for Maps than YouTube and Search, given the importance of objective accuracy for users in Maps, the emphasis on user-generated content and freedom of expression on YouTube, and user expectations that all legal content should be available on Search.

The Board's experience over the past five years has shown that the "legitimate aim" principle is more frequently complied with, primarily because content moderation decisions to enforce policies are often made to respect the rights of others, protect public order or support public health goals.

However, the Board is currently exploring when, and to what extent, content restrictions can pursue an aim related to the purpose of the service rather than a legitimate aim listed in the ICCPR. This question is likely to be prominent during the implementation of the DSA and assessments of whether the requirement for reasonable mitigation measures has been met in a way that respects the right to freedom of expression and other human rights.

The Board's initial premise is that mitigation measures that restrict freedom of expression for aims not listed in the ICCPR should (1) be informed by reasonable user expectations for the service based on its purpose; (2) meet the "legality" requirement by clearly explaining this difference in the designated provider's content policies; and (3) comply with other IHRL principles, such as non-discrimination and the least intrusive means test. The Board highlights this premise as one that will benefit from further research and exploration.

The Board also finds that an assessment of whether mitigation measures are reasonable should be informed by the analysis of proportionality and effectiveness, which are explored in sections 4.2 and 4.3.

## 4.2 Proportionate

Proportionality is a fundamental principle in IHRL for ensuring that measures taken in pursuit of a legal and legitimate aim are necessary and do not impose an excessive burden on the individual whose rights are restricted.

Specifically, UN General Comment 34 emphasizes that restrictions on freedom of expression should conform to the principle of proportionality by being "appropriate to achieve their protective function...the least intrusive instrument amongst those which might achieve their protective function; [and] proportionate to the interest to be protected" (UN General Comment No. 34, para. 34).

Further, the UN Special Rapporteur on freedom of opinion and expression has defined proportionality as requiring restrictions that (1) target a specific objective; (2) do not unduly intrude upon the other rights of targeted persons; and (3) ensure that interference with third-party rights be limited and justified in light of the interest supported by the intrusion ([Special Rapporteur Communication USA 6/2017](#)).

In EU law, "proportionality" is a general principle that restricts authorities by requiring them to strike a balance between the means used and the intended aim. For example, proportionality is enshrined in Article 5(4) of the [Treaty on European Union](#) (TEU), stating: "Under the principle of proportionality, the content and form of Union action shall not exceed what is necessary to achieve the objectives of the Treaties."



However, as can be seen in Article 5(4) of the TEU, the principle of “necessity” is a core element of achieving the overarching principle of “proportionality.” For example, the Court of Justice of the European Union (CJEU) applies a three-part test for proportionality that consists of “suitability” (i.e., appropriate to achieve the legitimate aim pursued with a rational connection between the means and the objective), “necessity” (i.e., no less onerous or restrictive alternatives available that could achieve the same legitimate aim), and “proportionality” (i.e., not impose a burden on the individual or entity that is excessive in relation to the objective sought to be achieved).

Further, Article 52(1) of the [Charter of Fundamental Rights of the European Union](#) states that “Any limitation on the exercise of the rights and freedoms recognized by [the] Charter must be [...] Subject to the Principle of Proportionality [...] made only if they are necessary and genuinely meet objectives of general interest recognized by the Union or the need to protect the rights and freedoms of others.”

### The Three-Part Test and Proportionality in the DSA

The link between proportionality and necessity is referenced in the DSA, with Recital 86 of the DSA stating that mitigation measures “should be proportionate in light of the economic capacity of the provider ... and the need to avoid unnecessary restrictions on the use of their service, taking due account of potential negative effects on those fundamental rights ... [giving] particular consideration to the impact on freedom of expression.”

Further, Recital 153 of the DSA states that when implementing the DSA, public authorities “should achieve, in situations where the relevant fundamental rights conflict, a fair balance between the rights concerned, in accordance with the principle of proportionality.”

Finally, it should be noted that the primary point of reference for proportionality in the DSA appears to be the overall system-wide risk, not a specific case in question (Del Campo, Zara, and Ugarte, 2025). DSA Article 35 requires that designated providers put in place reasonable, proportionate and effective mitigation measures that are “tailored to the specific systemic risks identified pursuant to Article 34.” Article 14 of the Delegated Regulation on Independent Audits under the DSA requires that auditors consider “whether [the mitigation measures] respond collectively to all the risks.”

### Insights from Oversight Board Cases: Relationship between Systems and Specific Cases

The Board’s experience with case decisions illustrates that protecting users’ freedom of expression requires mitigation measures that **are necessary and proportionate for system-wide risks and result in actions that are necessary and proportionate when applied to specific cases.**

To inform this report, we assessed whether mitigation measures that result in necessary and proportionate action in particular cases are also generally necessary and proportionate for the associated systemic risk, or whether there are inherent tensions between necessity and proportionality for the case as well as for the systemic risk.

In [Call for Women’s Protest in Cuba](#), the Board weighed the difficulties of moderating hate speech that includes comparisons to animals with the need to protect speech in contexts where there are strong restrictions on people’s rights to freedom of expression and peaceful assembly, especially in times of political protest. Using the [Rabat Plan of Action](#) as a guide, the Board focused on the social and political context, the speaker, the intent of the speech, the content



itself, the form of the speech and the likelihood and imminence of harm. The Board concluded that removing the content was neither necessary nor proportionate to achieve the legitimate aim of the Hate Speech policy and that content removal would have a disproportionate impact on the woman in the video, who overcame many difficulties that exist in Cuba.

This case illustrates the importance of ensuring mitigation measures that may be considered necessary and proportionate at the system level also result in outcomes that are necessary and proportionate in specific cases and contexts, especially since actions may be necessary and proportionate in one context but not in another. The Board came to a similar conclusion in [Iran Protest Slogan](#) and [Pro-Navalny Protests in Russia](#). When assessing compliance with the DSA, designated providers and auditors should test mitigations to determine whether the actions resulting from mitigation measures are also necessary and proportionate in a sample of specific cases; if they are not, this would call into question the necessity and proportionality of the collective mitigation measures.

The Call for Women’s Protest in Cuba case also illustrates the importance of considering the necessity and proportionality principle for both the specific case and the overall system. While the Board overturned Meta’s decision to remove the post, the Board also reiterated prior recommendations to improve how context and language expertise is incorporated into content moderation workflows.

This relationship between necessity and proportionality for system-wide risks and specific cases can be challenging when addressing the risk of “cumulative harm,” where one piece of content is unlikely to be directly connected to harm, but an accumulation of similar content (e.g., thousands of posts) may be linked to harm or even constitute a systemic risk. For example, this might occur if the accumulation of posts results in offline harm or undermines the freedom of expression for some users by causing them to leave the platform. While the notion of “cumulative harm” is contested, with Board decisions relying on this concept often including a minority and majority split, it is important for designated providers to consider the connection between individual cases and systemic risk.



In [Depiction of Zwarte Piet](#), the majority of the Board concluded that removing content with caricatures of Black people in the form of blackface was necessary to protect the rights of Black people to equality and non-discrimination. While it is challenging to establish a precise causal link between an individual post and the harms of discrimination, the majority argued that the accumulation of degrading caricatures on social media created an environment where acts of violence were more likely to be tolerated, reproducing discrimination in society and reinforcing ongoing structural racism. The majority concluded that less severe interventions, such as labels, warning screens or other measures to reduce dissemination, would not have provided adequate protection against the cumulative effects of leaving content of this nature on the platform. A minority of the Board, however, saw insufficient evidence to directly link this piece of content to the harm being reduced by removing it, and did not believe the requirements of necessity and proportionality had been met.

As can be seen from these cases, it is essential to ensure that mitigation measures considered necessary and proportionate for the magnitude of the risk result in actions that are also necessary and proportionate for specific cases and contexts. These cases also demonstrate that assessing the necessity and proportionality of the intervention must consider the perspectives of those most directly affected, especially those at the most significant risk of becoming vulnerable or marginalized (UNGPs General Principles, p2; DSA Recital 90).

It should not be enough for designated providers to simply demonstrate that mitigation measures meeting the descriptions in Article 35 are in place (e.g., adapting the design, features or functioning of their services; adapting their terms and conditions and their enforcement; adapting content moderation processes). Rather, designated providers should also demonstrate that the actions taken to implement these mitigation measures are necessary and proportionate for system-wide risks and result in content restrictions that are necessary and proportionate when applied to specific cases. The DSA requirement that mitigation measures are “proportionate” should, as interpreted through the lens of the ICCPR concepts of necessity and proportionality, serve as both a test for the overall mitigation strategy and a core principle for how designated providers define and enforce their content policies.

### Insights from Oversight Board Cases: Importance of Necessity

The Board’s experience illustrates that freedom of expression is best respected with approaches that encompass both necessity (i.e., least intrusive means) and proportionality (i.e., target a specific objective, without unduly intruding upon the rights of others), as espoused by the UN system (General Comment No. 34).

- **Visibility and reach of content:** In [Sudan Graphic Video](#), the Board upheld Meta’s decision to restore a post depicting violence against a civilian in Sudan because it raised awareness of human rights abuses and had significant public interest value. The Board concluded that placing a warning label on the content (rather than removal) was a necessary and proportionate restriction on freedom of expression. The Board also recommended that Meta add a specific exception to the Violent and Graphic Content Community Standard for raising awareness of or documenting human rights abuses, provided that a warning screen is displayed to inform users that the content may be disturbing.
- **Least intrusive means:** In [Altered Video of President Biden](#), the Board noted that in most cases, Meta can prevent harm to users caused by being misled about the authenticity of audio or audiovisual content through less intrusive means than removal or demotion, such as content labels. Rather than promote trust, content removal and demotion can sow



distrust and fuel accusations of cover-up and bias.

In [Claimed COVID-19 Cure](#), the Board emphasized that Meta should explain the range of options it has at its disposal in achieving legitimate aims and articulate why the selected one is the least intrusive means. The Board noted that Meta should publicly demonstrate three things in determining its least intrusive means: (1) the public interest objective could not be addressed through measures that do not infringe on speech; (2) among the measures that infringe on speech, Meta has selected the least intrusive means and (3) the selected measure helps achieve the goal and is not ineffective or counterproductive (A/74/486, para. 51-52). Here, the Board concluded that Meta did not explain how the removal of content constituted the least intrusive means of protecting public health, and the removal of the post therefore failed the necessity test.

In [Referring to Designated Dangerous Individuals as “Shaheed”](#), the Board concluded that, even though one meaning of “shaheed” does correspond to the English word “martyr” and is used in that way, it is not necessary or proportionate for Meta to remove all content solely for use of the word “shaheed” when referring to designated individuals. While Meta must seek to prevent its platforms from being used to incite acts of terrorist violence, a legitimate aim of its content moderation policies and a severe harm to address, removing all content solely for use of the word “shaheed” when referring to designated individuals was not necessary or proportionate for the pursuit of this policy goal.

However, determining what content policies and enforcement actions are necessary and proportionate is not always straightforward, and there are times when the Board has not reached consensus. This is to be expected and underscores the importance of designated providers describing their approach publicly for evaluation by stakeholders, as well as for auditors and regulators to deploy nuanced approaches that allow different designated providers to reach different conclusions.

- **Severity of risk:** In [Haitian Police Station Video](#), the Board looked to the [Rabat Plan of Action](#) to evaluate the necessity and proportionality of removing the content under review. A majority of the Board found that removing the content, nearly three weeks after it was posted, was no longer necessary, given the diminished likelihood of harm so long after the content was posted. However, a minority of the Board considered that while the risk of harm to the individuals depicted in the video was most acute in the days following the posting of the content, the risk that the video could lead to additional and retaliatory violence had not passed, given the overall context of ongoing violence and insecurity in Haiti.
- **Privacy considerations:** In [India Sexual Harassment Video](#), a majority of the Board found that removal of the content was not necessary and proportionate, but applying a warning screen and age restriction satisfied this test by reducing the probability of the victim being identified, thereby lowering the risk of re-victimization, social stigmatization and doxing. However, a minority of the Board disagreed, emphasizing that the remaining risk implied that the complete removal of the video would be necessary and proportionate.

By anchoring a methodology in IHRL that encompasses both necessity and proportionality, designated providers can take an approach consistent with the spirit of EU law, informed by a replicable model for rights-based analysis based on the application of the three-part test (such as in the work of the Board), deployable in non-EU contexts, and consistent with the IHRL obligations of EU states.



### 4.3 Effective

Unlike “reasonable” and “proportionate”, the principle of “effectiveness” does not explicitly appear in the three-part test required by Article 19 of the ICCPR.

However, the notion of effectiveness – i.e., whether a measure actually achieves its goals – has been proposed by the UN Special Rapporteur on freedom of opinion and expression as part of the test for assessing whether an intervention is the least intrusive means as a restriction on speech. (A/74/486, para. 51-52).

In addition, the principle of effectiveness is directly referenced in the UNGPs, which informs several relevant EU laws<sup>4</sup> and provides helpful direction for how the principle of “effective” can be interpreted by designated providers in practice. Specifically, the principle of effectiveness appears in two places:

- **Principle 20 Tracking Effectiveness:** To verify whether adverse human rights impacts are being addressed, companies should “track the effectiveness of their response.” This tracking should “be based on appropriate qualitative and quantitative indicators” and “draw on feedback from both internal and external sources, including affected stakeholders.”
- **Principle 31 Effectiveness Criteria for Non-Judicial Grievance Mechanisms:** To “ensure their effectiveness,” non-judicial grievance mechanisms should be legitimate, accessible, predictable, equitable, transparent, rights-compatible and a source of continuous learning. They should also be based on engagement and dialogue, consulting the stakeholder groups for whom they are intended to ensure that the design and performance of grievance mechanisms meet their needs.

Effectiveness is not merely aspirational in a human rights context, but a concrete requirement for actions that produce meaningful outcomes, help ensure that rights are protected in practice and for people affected by violations to have access to effective remedies. As directed by the UNGPs, evaluations of effectiveness should include both meaningful engagement with affected stakeholders, and the use of qualitative and quantitative data obtained via methods such as stakeholder engagement and reporting channels.

### Insights from Oversight Board Cases

The Board has made recommendations to Meta since January 2021 that have focused on service design and the effective enforcement of content policies, not just the substance of content policies. These recommendations have enhanced respect for freedom of expression and other human rights of users, such as by increasing alignment between what Meta’s policies aim to achieve and how they are implemented in practice.

The Board uses both publicly available and internal Meta data made available to it to understand the impact of the Board’s recommendations. A [public recommendation tracker](#) records both Meta’s response to the Board’s recommendations and the company’s implementation progress and is complemented by quarterly transparency reports. The staff of the Board has also published lessons learned from implementation tracking for regulators.<sup>5</sup>

<sup>4</sup> For example, the EU Corporate Sustainability Due Diligence Directive ([Directive 2024/1760](#))

<sup>5</sup> Naomi Shiffman, Carly Miller, Manuel Parra Yagnam, and Claudia Flores-Saviaga: [Burden of Proof: Lessons Learned for Regulators from the Oversight Board’s Implementation Work](#), Journal of Online Trust and Safety, (February 2024)



Based on experience, the Board believes that the determination of “effectiveness” in the context of the DSA should (1) include relevant quantitative metrics; (2) be informed by feedback from affected stakeholders; (3) consider whether the mitigation measure is being implemented consistent with the principles of equality and non-discrimination; and (4) consider impacts on all users globally, not only users in the EU.

This approach to “effectiveness” would be consistent with the spirit of Recital 90 of the DSA, which states that mitigation measures should be tested and designed “with the involvement of representatives of the recipients of the service, representatives of groups potentially impacted by their services, independent experts and civil society organizations.” This approach to effectiveness can also inform a designated provider’s cycle of ongoing human rights due diligence.

These factors (i.e., quantitative metrics, stakeholder feedback, non-discrimination and global relevance) will help ensure that mitigation measures account for the impact on human rights. They also provide additional meaning to the expectation of Article 35 of the DSA that designated providers consider adapting content moderation processes (1c), test and adjust their algorithmic systems (1d), and reinforce the internal processes, resources, testing, documentation or supervision of their activities (1f).

- **Metrics:** The Board reviewed Meta’s [cross-check program](#) (which aims to address mistaken removals by providing an additional layer of human review for certain posts) and made several recommendations for how the program could be improved for all users. The Board was concerned that cross-check granted some high-profile users greater protection than others and emphasized that any mistake-prevention system should not prioritize business concerns over speech that is in the public interest. In this context, the Board expressed concern that the metrics used to measure cross-check’s effectiveness did not capture all key concerns, such as whether decisions made through cross-check were more or less accurate than those made through its standard quality-control mechanisms.

In [Cartoon Showing Taliban Oppression Against Women](#), the Board highlighted shortcomings in Meta’s enforcement procedures, particularly in detecting and interpreting



images associated with dangerous organizations and individuals, and expressed concern that overenforcement of this policy could lead to the removal of artistic expression linked to legitimate political discourse. The Board re-emphasized its prior recommendations that Meta assess the accuracy of human reviewers enforcing the reporting allowance under its Dangerous Organizations and Individuals policy to identify systemic issues that may be causing enforcement errors.

In [United States Posts Discussing Abortion](#), in cases featuring posts arguing for and against abortion rights, the Board explored the potential for improvements in Meta’s machine learning and automated tools to reduce the number of false positives (when content is erroneously removed) without increasing the number of false negatives (when content is erroneously kept online). This is especially relevant in the context of political speech, where posts are more likely to use words and phrases that present an increased risk of being mistaken for violent threats (e.g., when the threat is not meant literally). The Board recommended enhanced use of enforcement accuracy data to inform necessity and proportionality analysis of the trade-offs in policy development and enforcement at scale. Similarly, in [Breast Cancer Symptoms and Nudity](#), the Board recommended that Meta implement an internal audit procedure to analyze a representative sample of automated content removal decisions to identify and learn from enforcement mistakes.

- **Stakeholder feedback:** A structured public comment period plays a key role in the Board’s process for standard case decisions and policy advisory opinions, providing an opportunity for organizations and individuals to help shape outcomes and recommendations by providing insights and expertise, such as those relating to language, culture and human rights, among others. The Board values these inputs for highlighting the various issues different cases and issues raise.

For example, in [Content Targeting Human Rights Defender in Peru](#), the Board received 65 public comments and consulted with advocacy organizations, academics, inter-governmental organizations and other experts on protecting human rights defenders online. Themes raised included the social and political context in Peru; the situation of human rights defenders; gendered dimensions of threats against defenders; recent legislative initiatives that impact the activities of NGOs in Peru; and social media narratives accusing NGOs, human rights defenders and civil society groups of “terrorism.”



The Board also emphasizes the importance of Meta undertaking its own stakeholder engagement. For example, in [Öcalan's Isolation](#), the Board recommended that Meta ensure meaningful stakeholder engagement on policy changes, emphasizing effective participation of individuals most impacted by the harms the policy seeks to prevent, as well as those with insights into the harms that may result from overenforcement. In [Criticism of EU Migration Policies and Immigrants](#), the Board recommended that Meta should undertake broad stakeholder engagement when auditing its slur lists, including consultation with impacted groups and civil society.

- **Non-Discrimination and Global Effectiveness:** The Board's cases illustrate that effective mitigation measures should adhere to the principle of non-discrimination, paying particular attention to the rights, needs and challenges of people who may be most vulnerable to adverse impacts (UNGPs General Principle, p2). The Board also finds that a single global approach based in IHRL is more likely to be consistent and replicable across different jurisdictions.

In [Mention of the Taliban in News Reporting](#), the Board overturned Meta's original decision to remove a post from a news outlet page reporting a positive announcement from the Taliban regime in Afghanistan on women and girls' education because there was no underlying violation of Meta's content policies. The Board expressed concern that Meta's systems for preventing enforcement errors of this kind were ineffective, particularly given the severity of the sanctions imposed. The Board recommended that Meta assess the accuracy of reviewers enforcing the reporting allowance under Meta's Dangerous Individuals and Organizations policy to identify systemic issues causing enforcement errors, particularly in languages other than English.

In [Sudan's Rapid Support Forces Video Captive](#) the Board recommended that Meta audit the training data used in its video content classifier to evaluate whether it has sufficiently diverse examples of content in the context of armed conflicts, including different languages, dialects, regions and conflicts. This would help ensure that content added to review queues is more equitably prioritized according to the probability and severity of potential violations, thereby enhancing the effectiveness of enforcement mechanisms.

The Board also raised concerns relating to disparate impact in [Homophobic Violence in West Africa](#). The Board examined Meta's enforcement practices in multilingual regions and expressed concern about the lack of human reviewers and market experts who speak Igbo, a language spoken by tens of millions of people in Nigeria and globally.

The Board acknowledges the practical reality that errors will be made when enforcing content policies at scale. For this reason, it is important for designated providers to pay special attention to the users most vulnerable to errors and to provide effective, accessible and transparent channels for mistakes to be reported and reversed. For example, in [Wampus Belt](#), the Board emphasized the importance of understanding which people and communities bear the greatest burden of mistakes and investigating the root causes.

Finally, given the Board's observation above (section 3.4) that designated providers emphasize "effective" more than "reasonable" or "proportionate," it is crucial to underline that effectiveness alone is not a sufficient reason for introducing mitigation measures. Instead, these measures must also pass the three-part test and comply with the principles of legality, legitimacy and necessity/proportionality. A mitigation measure that is effective but not necessary and proportionate should not be enforced.



## 5. Human Rights-Based Approaches to Article 34

While this report has focused on the mitigation measure requirements of Article 35 of the DSA, the Board’s analysis also has two important implications for the systemic risk assessments required by Article 34. First, systemic risk assessments should place IHRL at their center. Second, systemic risk assessments should consider the impact on users globally, not just users in the EU.

### 5.1 Placing International Human Rights Law at the Center

Article 34 of the DSA requires designated providers to assess risks relating to (1) illegal content; (2) “negative effects on fundamental rights;” (3) civic discourse and electoral processes, and public security; and (4) gender-based violence, the protection of public health and minors, and serious negative consequences to the person’s physical and mental wellbeing.

However, the goal of grounding analysis in the IHRL requirements of legality, legitimate aim and necessity/proportionality suggests that implementation of Article 34 of the DSA could usefully be restructured to place human rights at the center of evaluating mitigation measures, rather than as just one of four risk categories.

With human rights separated into a standalone category, it has become more challenging for designated providers to assess whether their mitigation measures achieve the legality, legitimate aim and necessity/proportionality tests because the impact on human rights has been framed as being separate from, rather than as inherent to, the other risks listed in Article 34.

For example, even where platforms are requested to remove unlawful content, questions of “negative effects on fundamental human rights” may still arise if national laws or their application are in tension with IHRL. Similarly, a human rights-based analysis is needed to understand what might constitute a negative effect on civic discourse, electoral processes or public security, especially because expression that could be reasonably classified under these categories is also lawful (Del Campo, Zara, and Ugarte, 2025).

Expressly treating negative impacts on human rights as a cross-cutting risk area can ensure closer adherence to IHRL in both identifying risks and ensuring reasonable, effective and proportionate mitigations.

### 5.2 Impact on Global Users

The UNGPs establish the responsibility to respect internationally recognized human rights as “a global standard of expected conduct for all business enterprises wherever they operate” (UNGPs 11), while the DSA’s scope is limited to “systemic risks in the Union”. This contrast is reflected in systemic risk assessment reports, which tend to emphasize the implementation of globally consistent content policies and human rights commitments, but also state the scope of their assessments is limited to the EU.



There is a risk that disproportionate focus on users in the EU contradicts one of the core concepts of the UNGPs, specifically that companies should prioritize adverse human rights impacts that are most severe or where a delayed response would make them irremediable. Despite the DSA's focus on users based in the EU, human rights-based approaches imply that designated providers should continue to prioritize efforts where impacts on people, society and the environment are most severe globally.

It is also essential to acknowledge that the interests of EU users are rarely self-contained but are often inextricably linked to the interests of global users. For example, many DSA systemic risk assessments consider enforcement accuracy across different languages by using 24 official EU member state languages as the reference point. This is an important step, but many users living in the EU speak other languages, and mitigation measures implemented on all languages can impact users both inside and outside the EU. For instance, implementation of the Board's recommendation in [Referring to Designated Dangerous Individuals as "Shaheed"](#) to address disproportionate restrictions on freedom of expression and civic discourse by ending Meta's blanket ban on the term "shaheed" will benefit Arabic speakers everywhere, including in the EU.

The Board believes that designated providers have a human rights responsibility to review the impact of mitigation measures established to fulfill Article 35 of the DSA on all global users (and non-users), particularly with regard to freedom of expression.





## 6. Conclusions

This report has evaluated how the three-part test required by ICCPR Article 19 can inform analysis of whether mitigation measures meet the requirement of being “reasonable, proportionate and effective” under DSA Article 35. The Board reaches the following conclusions:

- **Reasonable:** The “legality” and “legitimacy” aspects of the three-part test required by ICCPR Article 19 can inform analysis of whether mitigation measures that impact freedom of expression are consistent with the principle of “reasonableness”. Further, an assessment of whether these mitigation measures are reasonable should be informed by the analysis of proportionality and effectiveness.
- **Proportionate:** Analysis of whether mitigation measures are “proportionate” should encompass the interlinked principles of both necessity (i.e., least intrusive means) and proportionality (i.e., target a specific objective, without unduly intruding upon the rights of others).

To ensure consistency with IHRL standards, designated providers and auditors should consider (1) whether the mitigation measures are “necessary and proportionate” to address the relevant systemic risk broadly and (2) whether the mitigation approach gives rise to “necessary and proportionate” measures on a case-by-case basis. The latter could be achieved by reviewing a sample of cases across different contexts.

- **Effective:** Analysis of whether mitigation measures are “effective” should encompass (1) relevant quantitative metrics; (2) feedback from affected stakeholders; and (3) evidence of whether mitigation measures are being implemented in an equitable and non-discriminatory manner, such as across language and dialect. Finally, an analysis of effectiveness is relevant for reviewing whether a mitigation measure is the least intrusive means of achieving a legitimate aim.

Although each principle has been reviewed individually, the Board also finds that they are interconnected and integral to a designated provider’s cycle of ongoing human rights due diligence. Additionally, human rights-based approaches imply that the “reasonableness, proportionality and effectiveness” of mitigation measures should consider impacts on all global users, not only users in the EU.

The Board acknowledges that determining which mitigation measures to implement can involve tensions, trade-offs and various options. In this context, we believe designated providers should assess their mitigation measures against the global freedom of expression standard’s three-part test of legality, legitimate aim and necessity/proportionality, using the list of 11 mitigation measures found in Article 35 as an input rather than a definitive checklist of requirements.

The Board also notes that the outcomes of the three-part test may differ appropriately across designated providers, based on the different risk profiles and reasonable expectations that users have for different services. That said, the Board believes that all restrictions to freedom of expression should meet the “legality” requirement by clearly explaining restrictions in the designated provider’s content policies, and comply with other IHRL principles, such as



non-discrimination and the least intrusive means test. The Board highlights the question of different content restrictions across different service types as a topic that will benefit from further research and exploration.

While this report has focused on how the terms reasonable, proportionate and effective should be interpreted in the context of mitigation measures required by Article 35 of the DSA, it has also shed light on a crucial element of Article 34. Specifically, the Board believes that global human rights standards should be placed at the center of evaluating systemic risks and mitigation measures, rather than as just one of four risk categories. This would enable designated providers, auditors and regulators to more easily draw upon jurisprudence, precedent and case law relevant to DSA implementation, thereby enhancing the robustness of systemic risk assessments.

The Board looks forward to further engagement with designated providers and auditors on how best to ensure systemic risk assessments and mitigation measures enhance respect for freedom of expression and other human rights. The Board welcomes engagement with designated providers and auditors seeking to apply the insights shared in this report and use the three-part test to inform the design of mitigation measures that are “reasonable, proportionate and effective.”

The Board will continue to review systemic risk assessment and audit reports, providing analysis of the implications for human rights. The Board will also maintain engagement with a wide range of stakeholders – including the European Commission, Digital Services Coordinators, civil society organizations, academics and technology/policy experts – to advance human rights-based approaches to systemic risk assessments and mitigation measures.





## Annex: Methodology

This report combines (1) a desk-based review of all recently published designated provider systemic risk assessment and audit reports, (2) insights from the Board's case decisions and policy advisory opinions, and (3) perspectives of experts and other stakeholders.

### Assumptions

The Board's work begins with the following assumptions, which underpin the analysis presented in this report.

- **Companies should consider the global human rights impacts of their actions.** The Board believes that companies' responsibility to respect human rights extends globally, and that any effort to address human rights impacts in the EU must fully consider the consequences for human rights worldwide and respect the rights of the Global Majority.
- **IHRL standards should be a core expectation and reference point for content moderation and systemic risk.** Following the call of the UNGPs, the Board draws on the International Bill of Human Rights (consisting of the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights) for its analysis of content moderation and systemic risk. The Board also utilizes additional UN human rights instruments when relevant, such as the Convention on the Rights of the Child (CRC), the Convention on the Elimination of All Forms of Racial Discrimination (CERD), and the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW). (UNGP's Principle 12; A/HRC/38/35, paras 41 – 43).
- **The UNGPs provide a practical basis for policy, due diligence and access to remedy that can be applied globally.** These principles provide a well-established framework for considering the responsibilities of social media companies worldwide, including in areas such as content policy, due diligence and access to remedy. These responsibilities exist for companies independently of state obligations (A/HRC/32/38, paras 9 – 14) and are referenced in Recital 47 of the DSA.
- **Meaningful stakeholder engagement should inform decision-making.** The Board emphasizes the insights, perspectives and interests of people directly affected by company decision-making, particularly in relation to their freedom of expression and other human rights. Meaningful stakeholder engagement should be proactive, responsive and conducted prior to decisions being made (UNGP's Principle 18).

### Sources

The core of this report is a review of all recently published designated provider systemic risk assessment and audit reports in light of the accumulated insights the Board has gained from its prior case decisions (230 at the time of writing) and policy advisory opinions (four at the time of writing).



This report aims to promote developments in systemic risk assessment and mitigation measures that enhance respect for human rights by drawing connections between the issues addressed by the Board in cases and the systemic risks addressed by the DSA.

These case decisions and policy advisory opinions address some of the most complex content moderation issues affecting users, including crises and conflict situations, elections and civic space, gender, government interactions with platforms, hate speech, terrorism and violent extremism, and child safety. Collectively, the Board's cases have encompassed issues of content policies and their enforcement (including both human and automated review), algorithmic systems and platform design choices.

The Board's case decisions and policy advisory opinions overlap significantly with the topics outlined in Article 34 of the DSA, while [recommendations made by the Board](#) overlap significantly with the sample mitigation measures listed in Article 35 of the DSA. Both benefit from significant [public comments](#) and [stakeholder engagement](#).

The analysis provided by this report also benefits from participation in stakeholder engagements, such as those run by [the GNI, DTSP](#) and the European Commission, as well as focused discussions with experts and other stakeholders.

## Literature

This report also utilized literature evaluating the impact of the DSA.

- AccessNow and European Center for Non-Profit Law, [Towards Meaningful Fundamental Rights Impact Assessments Under The DSA](#)
- Centro de Estudios en Libertad de Expresión (CELE), [Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation](#)
- Centro de Estudios en Libertad de Expresión (CELE), [Reclaiming Human Rights for Platform Governance: Proposals for Restoring Their Centrality in the Era of Risks](#)
- [DSA Civil Society Coordination Group](#)
- [DSA Observatory](#) (various)
- European Contact Group, [ECG FAQs on the Delegated Regulation on Audits related to the Digital Services Act](#)
- The Future of Free Speech, [Thoughts on the DSA: challenges, ideas and the way forward through international human rights law](#)
- GNI, [Ensuring Digital Services Act Audits Deliver on Their Promise](#)
- GNI and DTSP, [European Rights & Risks: Stakeholder Engagement Forum 2025](#)
- Integrity Institute, [Global Transparency Audit](#)
- Integrity Institute, [Risk Assessment Guidance and Initial Analysis](#)
- Integrity Institute, [Risk Dimensions and Mitigation Effectiveness](#)
- Knight-Georgetown Institute, [Advancing Platform Accountability: The Promise and Perils of DSA Risk Assessments](#)
- Pielemeier, Jason S. and Sullivan, David, [Online Safety Regulation: Righting Risks or Risking Rights?](#)

