

RISK ASSESSMENT REPORT

on the results of the Systemic Risk Assessment
Under the Regulation (EU) 2022/2065 Digital Services Act (DSA)
conducted for [XNXX.COM](#)

NOVEMBER 2025

Content

1. Executive Summary	3
1.1 Update for Year 2 Risk Assessment Methodology	3
1.2 Risk Assessment Outcomes	4
2. Introduction	6
2.1 Purpose and Scope	6
2.2 Risk Management Framework	7
3. Platform Overview	9
3.1 Platform Features and User Interaction	9
3.2 Type of Content Hosted	9
4. Risk Identification	10
4.1 Risk Identification Methodology	10
4.2 Systemic Risks Identified	11
4.3 Risk Drivers Identified	21
5. Risk Assessment	29
5.1 Inherent Risk Assessment	29
5.2 Risk Drivers Assessment	36
5.3 Risk Mitigation	39
5.4 Residual Risk Assessment	52
6. Risk Response Strategy	59
6.1 Risk Response Strategy	59
6.2 Action Plan 2024-25 Implementation Status	59
6.3 Action Plan 2025-26	62
Action Plan	64

1. Executive Summary

The 2025 Systemic Risk Assessment for XNXX.com platform (also referred to as “Platform” or “XNXX”), conducted by NKL Associates s.r.o., reg. ID 02330482, with registered seat at Krakovská 1366/25, Praha 1, 110 00, Czech Republic (also referred to as “NKL”), the provider of the Platform, provides a comprehensive evaluation of systemic risks associated with the Platform’s operation as a Very Large Online Platform (“VLOP”) under the 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) (“Digital Services Act” or “DSA”).

1.1 Update for Year 2 Risk Assessment Methodology

During this assessment cycle, NKL has refined its systemic risk assessment methodology to strengthen compliance and transparency of methodological steps. The following methodological updates were carried out:

- 1) **Two-layer DSA analytical lens:** The methodology now explicitly distinguishes between
 - Systemic Risks – adverse societal or individual outcomes described under Article 34(1) DSA, such as the dissemination of illegal content, negative effects on fundamental rights, or impacts on the civil discourse, public health, users’ well-being, and
 - Risk Drivers – design or operational factors described under Article 34(2) DSA, that may contribute to those outcomes (e.g., recommender system design, content moderation activities, advertising systems, and data practices).
- 2) **Quantified residual-risk model:** To move beyond descriptive or qualitative assessments, NKL introduced a structured, two-step quantitative model for residual risk estimation. The detailed assessment procedure is presented in Section 5.4 This model does not replace professional judgment but provides a consistent assessment framework:
 - The first stage focuses on determining how far existing mitigation measures reduce the inherent level of each systemic risk. This expressed through the Risk Reduction Factor, which represents the average operational effectiveness of all controls that are formally linked to a given systemic risk and are recorded in the Mitigation Measures Register.
 - The second stage addresses residual exposure that may remain even after control preforms as designed. This residual influence is captured by the Driver Impact Score, that represents the continuing effect of the design and operational factors described in Article 34(2) DSA – the “risk drivers”.

This final score reflects the quantifiable effect of existing controls and the enduring contribution of operation features that continue to shape risk outcomes.

- 3) **Mitigation Measures Register:** Under Article 35 DSA, NKL introduced a structured Mitigation Measures Register to evidence all operational and design-based risk mitigations (see Section 5.3). Each control is assessed individually against three criteria – reasonableness, proportionality and effectiveness. Each control entry records the responsible control owner and applicable risk linkage.
- 4) **Cross-functional governance:** To ensure accountability, NKL formalised a cross-functional risk governance process. Workshops and validation sessions included participants from core operational functions. All workshops were documented (using a standard workshop-record template) to provide traceability of judgments.

- 5) **Protection of minors:** The methodology now incorporates a strengthened child-safety lens. Risk identification explicitly references the 5C framework and reflects limitations and proportionality concerns of available age-assurance mechanisms. In line with privacy-by-design principles, mitigations emphasise data minimalization and non-intrusive verification.
- 6) **Expanded traceability:** NKL is expanded its traceability model to enhances the transparency and continuity of its systemic risk management process. Every systemic risk entry now includes a full traceability chain – linking risk description, drivers, associated mitigation measures, responsible functions, quantified values (inherent and residual), and relevant risk strategies (where applicable).

1.2 Risk Assessment Outcomes

Inherent Risk Assessment

The assessment identified and evaluated 81 systemic potential risk scenarios across all DSA-defined systemic risks, with dissemination of illegal content, privacy and family life, data privacy and protection, and protection of minors emerging as the most significant areas of exposure. These categories showed the highest inherent probability and severity, primarily linked to non-consensual intimate imagery (“NCII”), child sexual abuse material (“CSAM”), doxing, data misuse, or inadequate age-assurance mechanisms. Additional potentials material risks were observed in areas concerning for instance gender-based violence, human dignity, and consumer protection, while lower inherent exposure was noted for freedom of expression, and civic and electoral impact. Overall, the platform’s inherent risk profile was assessed as medium, with nearly half of all the theoretical risk scenarios rated high before accounting for the mitigation measures.

Risk Drivers Assessment

76 operational and governance risk drivers were analysed on a residual basis to evaluate the effectiveness of existing mitigation measures. The majority were rated low to medium residual risk, confirming strong overall risk-management performance. Remaining high-influence drivers were concentrated in areas related to content-moderation transparency and user-rights communication, where incomplete feedback or appeal mechanisms can amplify the impact of incidents.

Mitigation Measure Assessment

The assessment confirmed that XNXX operates within a mature and well-structured risk-management framework. The Mitigation Measures Register consolidates 51 controls applied across the platform and evaluates them in terms of reasonableness, proportionality, and effectiveness. The results indicate that preventive and corrective mechanisms are generally robust, systematically implemented, and aligned with the Platform’s operational scale and regulatory obligations. Key strengths include for instance the integration of automated detection technologies with human moderation, clear advertising review protocols, and strong data-protection controls. Areas identified for improvement are limited and mainly procedural, such as strengthening user-verification standards and unifying feedback processes for user reports.

Residual Risk Assessment

Following implementation of multi-layered mitigation measures, the average residual risk declined from 14 to 6 (on a scale of 25), eliminating all high-residual risks. Remaining exposures are concentrated in medium-level categories. The most notable improvements occurred in privacy and family life, data protection, and illegal-content dissemination, driven by the combined effect of automated tools, structured moderation, and clearer user-reporting processes. Residual medium risks persist mainly in areas influenced by user behaviour or external dependencies.

Risk Response Strategy

The report concludes that NKL has established an effective, proportionate, and reasonable risk-management framework consistent with DSA obligations. The accompanying Action Plan 2025-26 focuses on further enhancing notice-mechanism transparency, strengthening account verification, refining age-assurance solutions, defining KPIs for measuring control effectiveness, and continuing structured engagement with regulators and NGOs on online safety. Continuous monitoring and annual reassessment ensure that systemic risks remain under control and aligned with evolving regulatory expectations.

2. Introduction

2.1 Purpose and Scope

This report presents a risk assessment conducted by NKL, the operator of the XNXX platform, in accordance with Articles 34 and 35 of the DSA.

It builds on last year's assessment as presented in the first iteration of this report and benefits from (i) our improved understanding of potential risks to users, society, and the company's compliance obligations; (ii) technical improvements that have been effected since; and (iii) insights obtained through the requests of information sent by the European Commission.

NKL operates XNXX, which is an online platform hosting content created by third parties, XNXX was designated by the European Commission as a VLOP in accordance with Article 33(4) of the DSA. This report aims at identifying, analysing and assessing, as comprehensively as possible, the risks associated with third parties' non-compliance with XNXX Terms of Services ("ToS") and NKL's compliance with the requirements set forth by the DSA. The primary objective is to assess risks that may affect users, the community or the Platform's compliance with the DSA including issues pertaining to the misuse by third parties of the Platform for distribution of illegal content, the Platform's compliance with transparency requirements, content moderation, data privacy, risks to fundamental rights, civic discourse and electoral processes, public health and persons' physical and mental well-being through the dissemination of content that promotes unhealthy or risky behaviours, misinformation about health, drug use, or encouraging self-harm, vulnerable groups such as minors, and victims of gender-based violence.

This report takes into consideration all elements related to our Platform's operational, technical and procedural aspects. It includes content moderation processes, data management, user security measures, advertising integrity and compliance reporting mechanisms. The assessment evaluates these risks in terms of their likelihood and potential impact on the platform's operations, users and the online community.

The development of this risk assessment considers the platform's diverse user base, reflecting the complexities of different languages and regional contexts. The risk assessment approach is based on the principles of proportionality and the risk-based approach outlined in the DSA, ensuring that risk management strategies are appropriately tailored to address both global and cultural/region-specific challenges.

The different languages and regional contexts, including when specific to a Member State, are reflected particularly in content moderation, where the Platform's content moderators are fluent in a broad range of languages, including German, French, Italian, English, Spanish, Polish, Russian, Slovak, and Czech, among others, enabling them to accurately assess and manage content across various regions. NKL employs a two-pronged approach to content moderation that combines the expertise of the platform's multilingual moderation team with reliable translation software, including Google Translate where necessary. While moderators handle content in their native languages, translation tools serve as supplementary resources for initial language support and to broaden coverage across languages not directly represented by the team.

The report comprises the following components to be read in conjunction with each other:

- The present *Risk Assessment Report*, which is the framework document explaining the concept of risk assessment, its methodology, and the strategy for risk mitigation;
- The consolidated *Risk Register*, which encompasses the following sub-components: *Systemic Risk Register*, *Risk Drivers Register*, *Systemic Risks Identification Catalogue*, *Risk Register Dashboard (overview of the assessment results)*;

- The *Mitigation Measures Register*, which is the catalogue of current mitigation measures, including owners and effectiveness assessment;
- The *Action Plan* document, which is capturing past and future action plans derived from the risk assessment strategy and ongoing monitoring.

All the above-mentioned documents are maintained in a manner that ensures easy accessibility for relevant internal stakeholders, as well as external auditors and regulators. Risk assessment is annually reviewed and updated to reflect new risks and measures, changes in operational practices, and shifts in regulatory requirements.

2.2 Risk Management Framework

The risk assessment process has been developed based on the principles outlined in **ISO 31000:2018 - Risk Management – Guidelines** (“ISO 31000”), which provides an internationally recognised framework for establishing a comprehensive and effective risk management system. This standard provides a clear framework for identifying, assessing, handling and monitoring risks, irrespective of their nature, emphasising the need for stakeholder involvement and senior management participation.

One of the core principles of ISO 31000 is continuous improvement, which is essential to maintaining a flexible and adaptable risk management culture. Leadership plays a crucial role in this framework, ensuring that risk management is integrated into NKL's strategic decision-making process. Management takes an active role in NKL, committing time, resources and oversight to ensure that risk management measures are effectively implemented and aligned with business objectives. This active involvement demonstrates a solid commitment to creating a sustainable, risk-informed environment.

In addition to these principles, the ISO 31000 framework is closely aligned with the requirements set out in the DSA for VLOPs. Although the DSA does not prescribe a specific risk management standard, the objectives of ISO 31000 - active risk identification, risk mitigation and ongoing monitoring – are particularly relevant and closely align with the expectations of the regulator and of the enforcer. For example, the systematic approach encouraged by ISO 31000 promotes compliance with the DSA's requirements for transparency in content moderation, data reporting obligations, and addressing the risks associated with the distribution of illegal content.

ISO 31000 has been chosen as the basis for a risk assessment and management framework for several key reasons:

- ISO 31000 is an internationally recognised standard that provides a comprehensive and structured approach to risk management. It covers all aspects of the risk management process, from risk identification and assessment to risk handling, monitoring and communication, ensuring consistency and thoroughness across the organisation.
- Although the DSA does not explicitly require ISO 31000, the principles underpinning the standard are closely aligned with the DSA's regulatory objectives. ISO 31000 promotes proactive risk identification, the development of mitigation strategies, and a focus on continuous improvement, which are key elements of any effective risk management system that addresses Platform management, transparency, and content regulation under the DSA.
- One of the most valuable aspects of ISO 31000 is its flexibility. This framework allows us to tailor our risk management approach to the unique requirements of the Platform while remaining flexible in response to changing regulatory requirements. This adaptability is necessary to integrate DSA's specific obligations into a broader risk management strategy, ensuring that NKL remains compliant as the regulatory environment evolves.

Role of the Compliance Function and Senior Management

The Compliance Function is NKL's second-line assurance unit, which operates independently from Senior Management. This function combines two mandates:

- the statutory Compliance Officer under Article 41 of the DSA, and
- the internal Risk Assessor under the ISO 31000 framework.

The purpose of this function is to provide continuous assurance that systemic risks are identified, assessed, and proportionately mitigated, and that the Platform fulfils its obligations under the DSA.

Key responsibilities related to risk management include the following:

- designing and maintaining the ISO 31000 risk assessment methodology and scoring criteria;
- chairing risk workshops and updating the Risk Register;
- drafting the annual systemic risk assessment and any ad-hoc updates required under Article 34 of the DSA;
- assigning, tracking, and verifying mitigation measures recorded in the Mitigation Measures Register; and
- reporting directly to Senior Management on systemic risks and potential non-compliance issues, with the authority to raise concerns and issue warnings when immediate action is required.

The Senior Management plays a crucial role in oversight and decision-making as it provides strategic direction, sets risk appetite, and is ultimately accountable for systemic-risk governance. Key responsibilities related to risk management include the following:

- setting the risk appetite and approving all risk-management documentation;
- confirming the adopted risk-management strategy and ensuring the implementation of required measures;
- reviewing the annual systemic risk assessment and relevant documentation;
- ensuring that risk-management efforts align with the organisation's strategic objectives and regulatory obligations; and
- allocating the necessary resources to support risk-mitigation and monitoring efforts.

3. Platform Overview

3.1 Platform Features and User Interaction

XNXX is a free, pornographic content-sharing platform that functions similarly to mainstream video-hosting websites but specializes exclusively in explicit adult material. Its interface and design are deliberately simple and optimized for accessibility, speed, and content discoverability.

Video Library and Search Functionality

XNXX hosts a catalogue of pornographic videos uploaded by its users. These are available for instant streaming without the need for registration or payment. The content is arranged into a vast taxonomy of categories (e.g., amateur, MILF, teen, fetish, interracial, lesbian, gay), allowing users to navigate intuitively through themes, and acts. A wide variety of content categories supports diversity and promotes freedom of expression within the sexual sphere.

Users can search the videos by keyword, video title, performer, or upload date. This algorithmically organized structure mirrors mainstream video-streaming services, offering personalized discovery through trending, recommended, or “related” videos. The homepage features a grid of video thumbnails, many of which play looping previews when hovered over.

User Accounts

XNXX can be used without registration; however, the Platform also offers optional user accounts for those seeking interactive uploading privileges. Where permitted by uploader settings and rights, users may also download videos to personal devices.

The Platform’s community and social features foster habitual engagement and create the perception of a participatory environment. These mechanisms (comments, “likes,” and replies) serve as low-barrier social signals rather than deep community exchanges.

3.2 Type of Content Hosted

The Platform’s content library consists particularly of user-uploaded **videos**, contributed by wide range of users from individuals and small groups to professional studios. These uploads span a vast spectrum of genres and explicitness and are often organized by user-applied tags (e.g., orientation, activity, body type, fetish, or region). This taxonomy is informal and community-driven rather than curated or standardized by the provider of XNXX.

Beyond video streaming, XNXX hosts a variety of static and text-based pornographic content, functioning as a broader adult community hub:

- **image galleries:** users can upload or browse curated photo sets, often featuring stills extracted from videos or professional photo sessions;
- **animated media (GIFs):** short, looping clips derived from videos are shared as animated GIFs, optimized for rapid viewing;
- **community forums:** the platform provides spaces for user interaction through comments.

4. Risk Identification

4.1 Risk Identification Methodology

Since the initial systemic risk assessment, XNXX has further enhanced its risk management framework to align with regulatory expectations and practical experience gained from ongoing operations. Several key developments have taken place across methodology and analytical depth.

Two-Layer Analytical Framework

In this assessment cycle, the Platform adopted a two-layer analytical framework that provides a structured and transparent way to analyse and manage risks under the DSA. The framework distinguishes between *Systemic Risks* and *Risk Drivers*, ensuring that the assessment captures both the potential outcomes (what could go wrong) and the underlying causes or mechanisms that make those outcomes more likely or severe.

- **Systemic Risks (SR):** Represent the adverse outcomes identified under Article 34(1) of the DSA. These include broad categories of harm such as the dissemination of illegal content, infringements of fundamental rights, and risks to public security, public health, or the protection of minors. Identified Systemic Risks are included in the *Systematic Risk Register* and are marked with a specific ID code (e.g. SR-IC-01).
- **Risk Drivers (DR):** Refer to the operational, technical, and governance factors that influence how these systemic risks may materialise or escalate. As outlined in Article 34(2) of the DSA, they include the design and functioning of recommender systems and other algorithmic processes, the effectiveness of content moderation systems, the enforcement of terms of service, the integrity of advertising systems, and data-related practices. Identified Risk Drivers are included in the *Risk Driver Register* and are marked with a specific ID code (e.g. DR-CM-01).

Each risk driver category contains several specific risk scenarios (e.g., delays in moderation, recommender bias, or insufficient age verification) that can increase the likelihood or impact of one or more systemic risks. For example, the systemic risk of “*dissemination of illegal content*” may be influenced by multiple drivers, such as delayed moderation response times, weak detection in non-English languages, or bot-driven amplification.

By combining both layers, the Platform moves beyond a static, one-dimensional risk register. This approach enables a causal understanding of risk formation: systemic risks define what harm may occur, while risk drivers explain why and how it might occur.

Overview of the Process

Risk identification followed an evidence-led and cross-functional approach, implemented through four phases described below:

Phase 1 – External knowledge base and desk research: To anchor the identification process in recognised standards and current evidence, the compliance team conducted a structured review of:

- The EU preliminary findings of the study on systemic risks and their mitigation¹;

¹ European Commission. (2025, March 18). Digital Services Act – Study on systemic risks and their mitigation: First workshop presentation. Directorate-General for Communications Networks, Content and Technology

- Regulatory guidance such as Ofcom’s risk assessment guidance on protecting people from illegal harms², or the children’s risk assessment guidance³;
- Academic and NGO reports on image-based abuse, deepfakes, discrimination, public-health impacts, and algorithmic or recommender effects;
- Case law and enforcement actions relevant to adult platforms (e.g., non-consensual content, trafficking, deceptive practices).

This corpus served as an external benchmark to verify completeness against Article 34(1) DSA scope and to identify emerging risks (e.g., AI-generated sexual imagery, grooming vectors via user-to-user features, harmful advertising verticals).

Phase 2 – Compliance team pre-work (structured brainstorming): Building on the external baseline, the compliance function prepared Platform-contextualised draft lists of systemic risks and risk drivers, including initial rationales and evidence tags (sources, operational observations, and prior incident learnings). This pre-work ensured that the proposed scenarios were not merely theoretical but grounded in XNXX’s operational realities and knowledge from previous risk-assessment cycles.

Phase 3 – Cross-functional internal workshops: The draft lists were reviewed through interactive workshops with cross-functional teams, including Content Moderation, Notice and Complaint, IT, and Advertising, with additional input from the Regulation Director and Legal Team. During the workshops, teams validated whether the proposed scenarios reflected actual failure modes and identified any missing vulnerabilities. Where gaps were identified, additional risk scenarios were added (for example, the Advertising Team expanded gender-based scenarios, while the Content Moderation Team added new illegal-content scenarios). The workshops also highlighted contextual differentiators to ensure the methodology captured disparate impact and non-discrimination considerations.

Phase 4 – Integration and documentation: Outputs from the previous phases were consolidated into two living artefacts:

- The *Risk Driver Register*, which includes each driver’s category (aligned with Article 34(2) DSA), its linkage to one or more systemic risks, and the relevant responsible teams;
- The *Systemic Risk Identification Catalogue*, which records each systemic risk’s category, scenario narrative and justification (regulatory basis and/or operational context), traceability to evidence (source references and internal observations), and team ownership.

These final lists formed the foundation for the subsequent risk assessment, as described in Section 5 of this report.

4.2 Systemic Risks Identified

This section provides a structured overview of the systemic risks identified through the methodology described in Section 4.1. Each risk category corresponds to Article 34(1) DSA and is informed by internal workshops and external evidence. The descriptions below outline the nature of the inherent risk, its relevance to adult-content platforms, and the justification for its classification as a systemic risk.

² Ofcom. (2024, December 16). *Risk Assessment Guidance and Risk Profiles: Protecting people from illegal harms online* (Guidance). <https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/risk-assessment-guidance-and-risk-profiles.pdf?v=390984>

³ Ofcom. (2025, April 24). *Children’s Risk Assessment Guidance and Children’s Risk Profiles*. <https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/statement-protecting-children-from-harms-online/main-document/childrens-risk-assessment-guidance-and-childrens-risk-profiles.pdf>

Dissemination of Illegal Content

The dissemination of illegal content remains one of the most material systemic risks for an adult content platform operating under the DSA. It encompasses the uploading, sharing, or promotion of material that violates applicable law.

Identified scenarios include:

- Upload or sharing of non-consensual intimate imagery, deepfakes, revenge porn, or doxxing (SR-IC-01);
- Dissemination of sexually exploitative or coerced material, including trafficking or extreme pornography (SR-IC-02);
- Use of comments to advertise sexual services or facilitate exploitation (SR-IC-03-04);
- Circulation of child sexual abuse material (CSAM), including via uploads or external links (SR-IC-05, SR-IC-06);
- Grooming or coercive contact with minors (SR-IC-07);
- Upload of animal cruelty or bestiality content (SR-IC-08);
- Distribution of malware or phishing links disguised as adult material (SR-IC-09);
- Uploading of copyrighted or counterfeit material (SR-IC-10);
- Terrorist or extremist propaganda, harassment, hate speech, and incitement to violence (SR-IC-11, SR-IC-12, SR-IC-18);
- Dissemination of illegal goods or instructions, including drugs, weapons, or unsafe sexual products (SR-IC-13, SR-IC-14, SR-IC-19);
- Encouragement of self-harm or suicide (SR-IC-15, SR-IC-16);
- Acts of foreign interference or disinformation using adult-content channels (SR-IC-17).

Evidence from the European Commission’s study on systemic risks and their mitigation (March 2025) (“the Commission’s workshop”)⁴ and Ofcom’s Illegal harms risk profiles (December 2024)⁵ indicates that adult platforms enabling upload of user-generated content are exposed to the dissemination of illegal content. Both sources identify adult-content services as high-risk environments due to their volume of user uploads, monetisation of explicit material, and open interaction features such as comments, live streams, and link sharing.

The Commission’s workshop specifically highlights categories including child sexual abuse material (CSAM), sexual exploitation, non-consensual intimate imagery (NCII), trafficking, and illegal or harmful goods and services as recurrent manifestations of systemic risk in the adult-platform context. Ofcom’s framework reinforces this by mapping relevant risk types, such as extreme pornography, sexual exploitation of adults, grooming, and illegal advertisement of sexual services, directly to the adult sector.

The combination of explicit sexual content, user-to-user functionalities, and financial incentives amplifies both likelihood and impact: illegal material can be uploaded, circulated, or monetised before detection. Moreover, technical and operational gaps, including moderation delays, weak detection in non-English languages, and misuse of hyperlinks, potentially enable the rapid re-emergence and persistence of unlawful content even after removal.

⁴ European Commission. (2025, March 18). Digital Services Act – Study on systemic risks and their mitigation: First workshop presentation. Directorate-General for Communications Networks, Content and Technology

⁵ Ofcom. (2024, December 16). *Risk Assessment Guidance and Risk Profiles: Protecting people from illegal harms online* (Guidance). <https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/risk-assessment-guidance-and-risk-profiles.pdf?v=390984>

Given this evidence, the category demonstrates high relevance (potential criminal and fundamental-rights violations), high prevalence (frequent appearance across upload and comment channels), and broad diversity of risk expressions, which makes the dissemination of illegal content the most significant systemic risk for the Platform under Article 34(1) DSA.

Human Dignity

According to findings from the compliance and moderation teams' workshops, adult platforms regularly faced user-generated content that may be considered degrading or dehumanizing (SR-HD-01). This might include, for example, scenes depicting gender-based humiliation, coercive role-play scenarios presented as authentic, verbal abuse, domination. Thus, this systemic risk is closely connected to the "Dissemination of Illegal Content" risk category. The moderation team has consistently identified these patterns through operational experience.

In parallel, the systemic commodification and exploitation of performers (SR-HD-02) potentially constitutes a structural risk to human dignity. Taking into account historic scandals involving the monetisation of non-consensual content⁶ it is evident that such harms may risk being embedded in governance and business practices, not merely in individual uploads. Empirical research cited in the catalogue (Donevan et al., 2025)⁷ documents continuous polyvictimisation and persistent mental-health challenges among individuals filmed for pornography, demonstrating that exploitation can be cumulative and enduring rather than episodic.

From a fundamental-rights perspective, this risk directly engages Article 1 (Human Dignity) and Article 3 (Integrity of the Person) of the Charter of Fundamental Rights of the European Union ("EU Charter"), as well as principles of equality and respect for a human person being embedded in the DSA's Article 34(1) framework.

Privacy and Family Life

The systemic nature of this category lies in its persistence and recurrence across multiple platform functions (e.g., uploads, comments, sharing tools) and in the intersection between privacy, dignity, and safety harms. Given that eventual breaches in this domain simultaneously infringe Articles 7 and 8 of the EU Charter (respect for private and family life and protection of personal data), the risk is both legally and socially severe.

Scenarios include:

- Non-consensual intimate imagery and leaks of private sexual material (SR-PF-01);
- Doxxing and exposure of personal data linking real identities to adult content (SR-PF-02);
- Blackmail or extortion through threatened publication of intimate material (SR-PF-03);
- Deepfake and manipulated imagery weaponised against private individuals (SR-PF-04);
- Exploitation of sensitive data such as metadata, location, or contact details (SR-PF-05);

The Commission's workshop highlights that adult platforms represent potential high-risk environments for privacy violations, where intimate data and imagery may be exposed, misused, or weaponised. These risks stem from the inherently sensitive nature of sexual content, which often embeds personal identifiers, sexual orientation, relationship details, and geolocation metadata.

Incidents such as non-consensual image sharing, doxxing, extortion, and deepfake abuse (SR-PF-01 – SR-PF-04) demonstrate how intimate material can be exploited to humiliate, threaten, or blackmail victims, resulting in reputational damage, emotional trauma, and secondary harm to family members. The catalogue further notes that

⁶ Uhl, C. A., Rhyner, K. J., Terrance, C. A., & Lugo, N. R. (2018). *An examination of nonconsensual pornography websites* [PDF]. *Feminism & Psychology*, 28(1), 50–68. <https://safeescape.org/wp-content/uploads/2022/01/examinationofsites2018.pdf>

⁷ Donevan, M., Jonsson, L. S., & Svedin, C. G. (2025). The experience of individuals filmed for pornography production: A history of continuous polyvictimization and ongoing mental health challenges. *Nordic Journal of Psychiatry*, 79(2), 156–165. <https://doi.org/10.1080/08039488.2025.2464634>

metadata and contact information linked to explicit content (SR-PF-05) can facilitate stalking and online harassment, extending harm beyond the original victims.

Data Privacy and Protection

From a fundamental-rights perspective, this risk category directly engages Article 8 (Protection of Personal Data) of the EU Charter.

Scenarios include:

- Data breaches or leaks exposing intimate content or identifiers (SR-DP-01);
- Profiling based on sexual identity or preferences that leads to discrimination (SR-DP-02);
- Surveillance or pervasive tracking discouraging lawful sexual expression (SR-DP-03);
- Invalid or opaque consent mechanisms and exploitation of personal data (SR-DP-04);
- Secondary exposure of victims' information through data linkage (SR-DP-05).

Adult platforms may process and store large volumes of highly sensitive personal data, including sexual imagery, user identifiers, behavioural data, and information revealing sexual orientation or preferences. Some material, by its very nature, falls within the special categories of personal data under the GDPR, demanding the highest level of protection. The systemic risks in this domain arise when intimate data are collected, shared, or exposed without sufficient safeguards or user control.

The Commission's workshop highlights recurring threats such as data breaches, invalid or opaque consent practices, and profiling based on sexual identity. These risks are amplified in adult-content ecosystems where personal and sexual-preference data could be exploited for advertising or recommender-system optimisation. Scenarios identified in the catalogue demonstrate how large-scale exposure of intimate content can lead to blackmail and harassment, how tracking and profiling undermine user autonomy, and how secondary data use without consent erodes trust and personal safety. Weak transparency and limited user control contribute to systemic violations of privacy and autonomy, rather than isolated compliance failures.

Freedom of Expression and Information

From a fundamental-rights perspective, this risk directly concerns Article 11 of the EU Charter, which protects freedom of expression and information, as well as media pluralism and diversity of opinion. Adult platforms operate at a unique intersection of sexual expression and content moderation. This category encompasses risks arising from excessive, arbitrary, or discriminatory moderation of lawful sexual content, leading to the suppression of legitimate expression and erosion of pluralism (SR-FE-01 – SR-FE-03).

The Commission's workshop indicates that disproportionate takedowns, opaque rules, and inconsistent enforcement are recurring risks on platforms hosting sexual or identity-based content. For adult services, these practices can result in over-removal of lawful material, marginalisation of minority voices, and unequal treatment of LGBTQ+, fetish, or non-mainstream creators. Examples such as the 2020 Pornhub⁸ content purge and the 2021 OnlyFans policy⁹ reversal illustrate how sudden enforcement shifts can restrict lawful expression on a massive scale.

The systemic nature of this risk lies in structural bias within moderation, algorithmic ranking, and advertising systems, where commercial and reputational incentives favour certain sexual expressions while suppressing others. This could create a chilling effect, discouraging creators and users from engaging in lawful discourse about sexuality.

⁸ Axios. (2020, December 17). *Pornhub's video purge poses a legal riddle*. <https://www.axios.com/2020/12/17/pornhubs-video-purge-legal-riddle>

⁹ Brus, D. (2021, August 19). *OnlyFans to prohibit "sexually explicit" content on platform*. Axios. <https://www.axios.com/2021/08/19/onlyfans-prohibit-sexually-explicit-content>

Non-Discrimination

From a fundamental-rights perspective, this risk directly engages Article 21 (Non-discrimination) of the EU Charter. The identified scenarios include:

- Exclusion of minority or LGBTQ+ creators from monetisation or visibility (SR-ND-01);
- Discriminatory advertisements that perpetuate stereotypes or hate (SR-ND-02);
- Unequal protection across languages or regions, especially non-English users (SR-ND-03);
- Racialised or fetishising taxonomies reinforcing harmful stereotypes (SR-ND-04).

Adult platforms can create or reinforce systemic inequalities through their design and moderation practices, resulting in unequal treatment of users or creators based on identity, or geographic location. The Commission’s workshop confirms that bias in recommendation algorithms, moderation processes, and ad targeting leads to structural exclusion of minorities and unequal access to visibility or revenue. On adult platforms, these effects manifest as reduced monetisation opportunities for LGBTQ+ or minority creators, amplification of racialised or fetishised stereotypes, and weaker protection against harmful content in non-English regions.

Protection of Minors

From a fundamental-rights perspective, this risk directly engages Article 24 (Rights of the Child) of the EU Charter. Adult platforms face one of the most critical systemic risks under the DSA concerning protection of minors, as exposure to pornographic or sexually explicit material is classified by EU and UK regulators as “primary priority content harmful to children”.

For the initial risk identification, the 5C typology of online risks to children were used (Content, Conduct, Contact, Consumer, Cross-cutting), as reflected in the European Commission’s Guidelines on measures to ensure a safety for minors (“DSA Article 28 Guidelines”)¹⁰. This category covers situations where minors are exposed to, exploited by, or otherwise harmed through access to or interaction with adult-content services.

Using the DSA Article 28 Guidelines (the 5C lens) and Ofcom’s Children’s Risk Assessment Guidance¹¹ at the identification stage, the following scenarios were considered:

Table 1: Mapping of systemic risk scenarios concerning the protection of minors to the 5C typology

ID	Systemic Risk Scenario	Primary 5C category	Possible secondary	Justification ¹²
SR-PM-01	Minors access pornographic content on the service, resulting in exposure to harmful or illegal sexual material	Content	Cross-cutting (safety/governance)	Ineffective or easily circumvented age assurance measures allow minors to gain access to harmful content. Both the Ofcom’s Children’s Risk Assessment Guidance and DSA Article 28 Guidelines stress that pornography is classified as primary priority content harmful to children and requires robust protection. Poorly designed or unreliable age checks increase the likelihood of minors being exposed to explicit sexual material.
SR-PM-02	Users upload, or share material depicting sexualized minors (CSAM) or images of	Content (illegal)	Conduct (illegal user behavior)	The platforms that allow user uploads face heightened risks of CSAM or sexualized images of minors being shared or redistributed. DSA Article 28 Guidelines makes clear that poor moderation practices enable this material to circulate, fuelling demand for further abuse and potentially encouraging the production of new material. Ofcom’s Children’s Risk

¹⁰ European Commission. (2025, October 10). *Guidelines on measures to ensure a high level of privacy, safety and security for minors online, pursuant to Article 28(4) of Regulation (EU) 2022/2065* (OJ C/2025/5519). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:C_202505519

¹¹ Ofcom. (2025, April 24). *Children’s Risk Assessment Guidance and Children’s Risk Profiles*. <https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/statement-protecting-children-from-harms-online/main-document/childrens-risk-assessment-guidance-and-childrens-risk-profiles.pdf?v=396653>

¹² The justification column is based on NKL’s internal document – *Systemic Risks Identification Catalogue*

	minors in sexualized contexts			Assessment Guidance also highlight that posting, sharing, and group messaging functionalities significantly increase the likelihood of minors being exposed to pornographic or abusive content. CSAM is not only illegal but also represents one of the most severe harms identified by regulators.
SR-PM-03	Minors are systematically exposed to explicit pornographic content via the platform's content delivery	Content	Cross-cutting (advanced tech/algorithms)	The recommender and personalization systems can amplify harm by repeatedly exposing minors to pornographic or otherwise harmful material once they have accessed the platform. The guidance stresses that recommender systems can quickly personalize content for minors while lacking equivalent safeguards, creating cumulative harm through repeated exposure. Ofcom's Children's Risk Assessment Guidance also identifies algorithmic systems as critical risk factors, given their role in disseminating content rapidly and widely to children.
SR-PM-04	Minors encounter harmful sexual content through discovery features (e.g., search suggestions, autocomplete, trending tags)	Content	Cross-cutting (design/UX)	The search functionalities like autocomplete or trending tags can directly expose minors to harmful sexual content, even if they did not initially intend to seek it out. It specifically required consideration of search content harmful to children, as minors can come across harmful material through system-driven outputs rather than deliberate searches. Regulators also emphasize that service design and interface choices can heighten risk by normalizing or pushing harmful material toward minors.
SR-PM-05	Minors who use comments on the platform are exposed to unwanted contact from unknown users, which may escalate into grooming	Contact	Conduct (predatory behavior), Cross-cutting (safety)	The DSA Article 28 Guidelines and the Ofcom's Children's Risk Assessment Guidance highlight that features such as direct messaging, friend requests, or live chat substantially raise the risk of unwanted contact, which can lead to grooming or coercion. Contact risks are a central category of harm identified in EU and UK regulatory frameworks, as they compromise children's privacy and safety.
SR-PM-06	Minors are extorted with threats to publish intimate images or coerced into producing sexual material	Contact	Conduct (illegal), Content (CSAM), Cross-cutting (privacy/safety)	The guidance identifies that functionalities enabling image sharing, private chat, or anonymous accounts increase the likelihood of minors being pressured into creating or sharing intimate material. Once produced, this material often becomes CSAM, which is both illegal and highly harmful. Extortion amplifies psychological distress and may result in repeated abuse. Regulators emphasize that cumulative harm occurs where minors are trapped in cycles of coercion, with the threat of exposure used to maintain control.
SR-PM-07	Minors are exposed to sexually explicit or pornographic advertisements, including ads for live sex cams, adult dating app, sex toys, and sexual enhancement products	Consumer	Content	The pornographic services are a key space where children can and do attempt to access adult content, often despite restrictions. Ineffective age assurance and weaknesses in ad targeting systems increase the likelihood that minors are inadvertently exposed to sexual material. Advertising models and recommender systems designed to maximize revenue can unintentionally push harmful or inappropriate ads.
SR-PM-08	Minors' well-being is impaired by exposure to pornography, leading to	Cross-cutting (health and well-being)	Content	Different studies (Andrie et al., 2021 ¹³ ; Common Sense Media, 2022 ¹⁴ ; Maheux et al., 2021 ¹⁵ ; Smahel et al., 2020 ¹⁶ ; Owens et al., 2012 ¹⁷) highlights that children at different developmental stages may encounter

¹³ Andrie, E. K., Sakou, I. I., Tzavela, E. C., Richardson, C., & Tsitsika, A. K. (2021). Adolescents' Online Pornography Exposure and Its Relationship to Sociodemographic and Psychopathological Correlates: A Cross-Sectional Study in Six European Countries. *Children*, 8(10), 925. <https://doi.org/10.3390/children8100925>

¹⁴ Common Sense Media. (2022). Teens and pornography: A survey of U.S. teens ages 13-17 [PDF]. <https://www.commonsensemedia.org/sites/default/files/research/report/2022-teens-and-pornography-final-web.pdf>

¹⁵ Maheux, A. J., Roberts, S. R., Evans, R., Widman, L., & Choukas-Bradley, S. (2021). Associations between adolescents' pornography consumption and self-objectification, body comparison, and body shame. *Body Image*, 37, 89-93. <https://doi.org/10.1016/j.bodyim.2021.02.004>

¹⁶ Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). EU Kids Online 2020: Survey results from 19 countries. *EU Kids Online*. <https://doi.org/10.21953/lse.47fdeqj01of0>

¹⁷ Owens, Eric W., et al. "The impact of internet pornography on adolescents: A review of the research." *Sexual Addiction & Compulsivity*, vol. 19, no. 1-2, Jan. 2012, pp. 99-122

	anxiety, confusion, and distorted views of relationships and sexuality			pornography online, often without seeking it. These include anxiety, confusion, unrealistic or distorted expectations of relationships, and reinforcement of harmful stereotypes. Vulnerable groups (such as those with pre-existing mental health challenges) are at even greater risk of harm. Ofcom’s Children’s Risk Assessment Guidance recognizes pornography as “primary priority content harmful to children,” meaning its potential impact is severe and requires specific mitigation measures.
SR-PM-09	Minors may encounter false or misleading information about sexual practices, contraception, or consent framed as “educational” porn	Content (disinformation)	Cross-cutting (health/well-being)	This systemic risk was identified following the compliance-team review of external research on protection of minors. Directly connected with SR-PM-08.
SR-PM-10	Weak or bypassable age verification measures allow minors to access pornographic content, exposing them to harmful or illegal sexual material	Content	Cross-cutting (safety/governance)	This systemic risk was identified following the compliance-team review of external research on protection of minors. Directly connected with SR-PM-01.
SR-PM-11	Intrusive or unsafe age verification systems (e.g., biometric scans, ID uploads) retain or process sensitive identity data without sufficient safeguards, exposing minors and adults to privacy violations	Cross-cutting (privacy and data protection)	Consumer (contractual/commercial data use)	The DSA Article 28 Guidelines require data-minimising, double-blind, privacy-preserving approaches to age assurance; providers should only receive a binary age result, and verification providers should not learn which services are accessed. They set non-intrusiveness as a criterion: process only strictly necessary attributes; do not store or repurpose age-assurance data. They also describe privacy-preserving architectures (e.g., anonymised age tokens, zero-knowledge proofs, EU Digital Identity Wallet / EU Age Verification Solution) as compliant exemplars.

The systemic dimension of this risk stems from the combination of service design choices or insufficient safeguards, which allow minors to be repeatedly exposed to explicit or harmful material even without intentional searching. Weak or privacy-intrusive age-verification processes compound the problem: while insufficient measures fail to prevent access, overly invasive systems may process sensitive identity data, creating secondary privacy risks for both minors and adults.

Consumer Protection

From a fundamental-rights perspective, these risks affect the right to fair and transparent treatment, closely linked to Article 38 (Consumer Protection) of the EU Charter. This category covers structural failures in transparency, fairness, and accessibility, which together affect users’ confidence and economic welfare.

Scenarios include:

- Opaque terms of service and unclear user rights (SR-CP-01);
- Barriers to reporting or appealing moderation decisions (SR-CP-02, SR-CP-05);
- Manipulative interface designs and dark patterns (SR-CP-03);
- Misleading or deceptive advertising and hidden charges (SR-CP-04);
- Unequal enforcement of rights across jurisdictions (SR-CP-06).

The Commission’s workshop identifies “lack of transparency, deceptive design, and barriers to complaint and appeal” as recurrent systemic failures across online services. The character of these risk lies in the business and design models that embed manipulative or non-transparent practices at scale, for instance, upselling subscriptions

through misleading cues, or concealing data-sharing policies that affect user autonomy. These practices, even when individually minor, collectively result in systemic deprivation of procedural fairness and effective redress, undermining user confidence and accountability mechanisms.

Civil and Electoral Impact

The combination of user-generated explicit content, anonymous participation, advertising systems accepting external creatives creates an environment susceptible to covert political messaging or manipulation.

The identified scenarios in this category include:

- Embedding political disinformation in pornographic material or advertising (SR-CE-01, SR-CE-02, SR-CE-03);
- Targeted manipulation of marginalised communities (SR-CE-04);
- Sexualised disinformation or deepfakes aimed at public figures, journalists, or women leaders (SR-CE-05).

The Commission's workshop identifies false or manipulated information, as emerging systemic risks relevant across online platforms, including those hosting adult or user-generated sexual content. Although not a primary platform function, the open and viral nature of adult-content environments allows political disinformation to be embedded within pornographic material, titles, comments, or advertisements.

The study further highlights that coordinated manipulation of public discourse and targeted micro-advertising can occur in non-traditional digital ecosystems where moderation or ad-screening is limited. Such vectors may be exploited to amplify political propaganda, distort electoral perceptions, or target marginalised audiences with tailored disinformation. A particularly harmful manifestation concerns sexualised or deepfake disinformation targeting political figures.

Public Security Concerns

The Commission's workshop identifies that online platforms may be misused to facilitate or conceal criminal activity, thereby posing risks to public safety and social order.

Scenarios include:

- Use by extremist groups for propaganda or recruitment (SR-PS-01);
- Trafficking or sexual exploitation disguised as consensual adult content (SR-PS-02);
- Upload or promotion of hate-driven or violent sexualised material (SR-PS-03);
- Harassment of journalists or activists using explicit material (SR-PS-04).

Extremist or radical groups may exploit adult platforms to distribute propaganda, recruit members, or normalise violence under the guise of explicit content. Criminal networks and traffickers can misuse creator accounts, advertising channels, or comment sections to facilitate sexual exploitation and human trafficking. Users may also upload or share hate-driven or violent sexualised content that glorifies aggression or fuels extremist narratives. Furthermore, coordinated actors have been observed targeting journalists, activists, and civil society figures with harassment or sexualised defamation campaigns, using explicit material to silence or intimidate.

The Commission's workshop notes that these behaviours exploit weak moderation and algorithmic amplification, allowing criminal or extremist content to persist and circulate rapidly.

Gender-Based Violence

The Commission's workshop together with supporting academic and enforcement sources, strongly indicates that gender-based violence manifests systemically within adult-content environments through a combination of non-consensual imagery, coercive sexual exploitation, algorithmic amplification of misogyny, and sexualised disinformation. Empirical evidence, such as research by Wright et al. (2016)¹⁸ and Lim et al. (2015)¹⁹, shows correlations between exposure to violent pornography and attitudinal support for sexual aggression, reinforcing the societal risk dimension of these harms.

The Commission's workshop findings further highlight algorithmic bias and recommender amplification as key enablers of systemic misogyny while documented enforcement actions (e.g., FTC 2025 settlement against Pornhub²⁰) illustrate industry-wide governance failures in preventing non-consensual and coercive content. Additional evidence from Bellingcat (2024)²¹ and Ajder et al. (2019)²² confirms that deepfake pornography overwhelmingly targets women, supporting the conclusion that it is a gender-specific digital violence vector.

The identified scenarios include:

- Violent or coercive sexual depictions normalising abuse (SR-GB-01, SR-GB-02);
- Harassment, stalking, or sextortion targeting women creators (SR-GB-03, SR-GB-04);
- Non-consensual or deepfake pornography disproportionately targeting women (SR-GB-05);
- Gendered disinformation campaigns silencing women in public life (SR-GB-06);
- Advertisements that monetise or normalise gender-based abuse (SR-GB-07, SR-GB-08).

The identified scenarios collectively demonstrate that adult platforms can both host and normalise violent or degrading portrayals of women, as well as facilitate direct harm through extortion, harassment, and deepfake abuse.

Finally, cross-linkages with civic and advertising risks reveal how sexualised disinformation and gender-harmful advertising extend gender-based violence beyond content into broader social and economic structures. Taken together, these patterns confirm gender-based violence as pervasive, cross-cutting systemic risk and among the most severe categories identified under Article 34(1) DSA.

Public Health

The Commission's workshop and public-health literature demonstrate that online platforms can materially influence sexual health behaviours and body image, thereby constituting a systemic public-health risk domain.

¹⁸ Paul J. Wright, Robert S. Tokunaga, Ashley Kraus, A Meta-Analysis of Pornography Consumption and Actual Acts of Sexual Aggression in General Population Studies, *Journal of Communication*, Volume 66, Issue 1, February 2016, Pages 183–205, <https://doi.org/10.1111/jcom.12201>

¹⁹ Lim, M. S. C., Carrotte, E. R., & Hellard, M. E. (2015). The impact of pornography on gender-based violence, sexual health and well-being: What do we know? *Journal of Epidemiology & Community Health*. Advance online publication. <https://doi.org/10.1136/jech-2015-205453>

²⁰ Federal Trade Commission. (2025, September 3). *FTC takes action against operators of Pornhub and other pornographic sites for deceiving users about efforts to crack down on child sexual abuse material and nonconsensual sexual content* [Press release]. <https://www.ftc.gov/news-events/news/press-releases/2025/09/ftc-takes-action-against-operators-pornhub-other-pornographic-sites-deceiving-users-about-efforts>

²¹ Koltai, K. (2024, February 23). *Behind a secretive global network of non-consensual deepfake pornography*. Bellingcat. <https://www.bellingcat.com/news/2024/02/23/behind-a-secretive-global-network-of-non-consensual-deepfake-pornography/>

²² Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019, September). *The state of deepfakes: Landscape, threats, and impact*. Deeptrace. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Academic and regulatory evidence cited in the catalogue (APHA 2010²³; Grudzen 2009²⁴; Paslakis 2022²⁵; Mundy 2025²⁶) substantiates that pornography and related advertising can normalise unsafe sexual practices, distort perceptions of sexual health, and encourage substance use in sexual contexts. These harms are intensified by platforms' monetisation structures, algorithmic amplification, and limited moderation coverage in health-related domains, which together facilitate wide and repeated exposure to harmful or misleading messages.

The identified scenarios cover multiple harm vectors:

- Promotion of unprotected or risky sexual acts (SR-PH-01);
- Dissemination of false health information (SR-PH-02);
- Unrealistic body standards and related mental-health harms (SR-PH-03);
- Depictions of drug use or “chemsex” normalising unsafe behaviours (SR-PH-04);
- Promotion of unsafe sexual-enhancement products (SR-PH-05).

Given the direct connection to users' physical and psychological health, and the systemic nature of the underlying mechanisms (user-generated uploads, ad placements) the public health risk domain constitutes a material systemic risk category.

Physic and Mental Well-Being

The Commission's workshop, supported by academic research (Wright et al., 2016²⁷; Hald et al., 2015²⁸; Bóthe et al., 2020²⁹; Hellevik et al., 2025³⁰), indicates that exposure to and engagement with adult-platform content can generate psychological and behavioural impacts affecting both users and victims.

Scenarios include:

- Exposure to violent or extreme sexual material influencing attitudes and relationship norms (SR-PW-01);
- Compulsive consumption patterns and associated health effects (SR-PW-02);
- Psychological trauma among victims of non-consensual content (SR-PW-03);
- Advertising stimuli that reinforce compulsive behaviour (SR-PW-04).

²³ American Public Health Association. (2010, November 9). *Prevention and control of sexually transmitted infections and HIV in the adult film industry* [Policy brief]. <https://www.apha.org/policy-and-advocacy/public-health-policy-briefs/policy-database/2014/07/28/15/23/prevention-and-control-of-sexually-transmitted-infections-and-hiv-in-the-adult-film-industry>

²⁴ Grudzen, C. R., Elliott, M. N., Kerndt, P. R., Schuster, M. A., Brook, R. H., & Gelberg, L. (2009). Condom use and high-risk sexual acts in adult films: A comparison of heterosexual and homosexual films. *American Journal of Public Health*, 99(Suppl 1), S152–S156. <https://doi.org/10.2105/AJPH.2007.127035>

²⁵ Paslakis, G., Chiclana Actis, C., & Mestre-Bach, G. (2022). Associations between pornography exposure, body image and sexual body image: A systematic review. *Journal of health psychology*, 27(3), 743–760. <https://doi.org/10.1177/1359105320967085>

²⁶ Mundy, E., Carter, A., Nadarzynski, T., Whiteley, C., de Visser, R. O., & Llewellyn, C. D. (2025, February 6). The complex social, cultural and psychological drivers of the ‘chemsex’ experiences of men who have sex with men: A systematic review and conceptual thematic synthesis of qualitative studies. *Frontiers in Public Health*, 13. <https://doi.org/10.3389/fpubh.2025.1422775>

²⁷ Wright, P. J., Tokunaga, R. S., & Kraus, A. (2016). Meta-analysis of pornography consumption and actual acts of sexual aggression in general population studies. *Journal of Communication*, 66(1), 183–205. <https://doi.org/10.1111/joc.12201>

²⁸ Hald, G. M., & Malamuth, N. M. (2015). Experimental effects of exposure to pornography: The moderating effect of personality and mediating effect of sexual arousal. *Archives of Sexual Behavior*, 44(1), 99–109. <https://doi.org/10.1007/s10508-014-0291-5>

²⁹ Beáta Bóthe, István Tóth-Király, Marc N. Potenza, Gábor Orosz, Zsolt Demetrovics, High-Frequency Pornography Use May Not Always Be Problematic, *The Journal of Sexual Medicine*, Volume 17, Issue 4, April 2020, Pages 793–811, <https://doi.org/10.1016/j.jsxm.2020.01.007>

³⁰ Hellevik, P. M., Haugen, L.-E. A., & Övertien, C. (2025). Outcomes of image-based sexual abuse among young people: A systematic review. *Frontiers in Psychology*, 16. <https://doi.org/10.3389/fpsyg.2025.1599087>

Continuous exposure to violent or extreme sexual material may shape attitudes towards sexuality and relationships, fostering desensitisation or acceptance of coercive behaviour. Prolonged use and compulsive consumption patterns can lead to health impacts such as sleep disruption, emotional distress, or reduced self-regulation, particularly when reinforced by the platform’s design or advertising dynamics. Victims of non-consensual content dissemination experience severe psychological trauma, including anxiety, depression, and social withdrawal, as documented by Hellevik et al. (2025)³¹. Additionally, advertising prompts that repeatedly encourage viewing (“live now,” “exclusive access”) can exacerbate compulsive behaviour and mental-health strain.

Collectively, these findings demonstrate that the risk to physical and mental well-being arises not from isolated incidents but from systemic exposure patterns and reinforcing platform mechanisms.

4.3 Risk Drivers Identified

This section summarises the main categories of risk drivers identified in the current assessment cycle. Risk drivers refer to the operational, technical, and governance factors that can affect the likelihood or impact of systemic risks. Each driver category is documented in the Risk Driver Register and was analysed through cross-functional workshops and operational evidence.

Content Moderation Drivers

Content moderation processes form the front line of defence against the dissemination of illegal or otherwise non-compliant material on the platforms. They are central to the platform’s ability to meet obligations under Articles 16-17 DSA (notice-and-action mechanisms, statement of reasons) and Article 20 DSA (internal complaint-handling systems).

Effective moderation requires not only technical detection tools but also human oversight and transparent procedures. Weaknesses at any point in this chain can result in continued exposure of users to illegal content, erosion of fundamental rights, and regulatory non-compliance. Conversely, excessive or inaccurate removals can lead to over-enforcement, suppression of legitimate expression, and discriminatory impacts on minority groups.

The current assessment identifies several interlinked risk mechanisms reflecting these operational and governance challenges:

- Moderation systems may fail to detect and remove illegal or harmful content promptly, allowing its continued visibility, re-upload, and dissemination (DR-CM-01). Delayed removal of illegal content, such as CSAM, revenge porn, or trafficking-related content poses severe legal and reputational risks.
- Automated moderation tools may misclassify legal content, such as consensual sexual role-play, or LGBTQ+ expression, as harmful or illegal (DR-CM-02). Such false positives lead to unjustified restrictions on freedom of expression and may disproportionately affect marginalised creators or minority communities.
- Inadequate user-verification and account-linking mechanisms allow offenders to re-register after enforcement actions, facilitating the persistent re-upload of prohibited material (DR-CM-03). This undermines deterrence and consumes moderation resources.
- Moderation teams may adopt an overly cautious approach to avoid liability, erring on the side of excessive removals (DR-CM-04). While this may reduce regulatory exposure in the short term, it risks chilling lawful expression and reducing media pluralism, especially in sensitive domains like sexual identity or artistic performance.

³¹ Hellevik, P. M., Haugen, L.-E. A., & Övertien, C. (2025). Outcomes of image-based sexual abuse among young people: A systematic review. *Frontiers in Psychology*, 16. <https://doi.org/10.3389/fpsyg.2025.1599087>

- Differences in interpretation of policies across teams, shifts, or linguistic markets can lead to uneven application of rules and under-enforcement of abusive or discriminatory content (DR-CM-05 – DR-CM-06). Inconsistency weakens user trust and may constitute indirect discrimination under fundamental-rights standards.
- Absence of clear escalation and triage protocols delays urgent action on high-severity material such as CSAM, terrorist content, or imminent self-harm (DR-CM-07). Without structured routing to specialised reviewers or law-enforcement liaison, critical incidents may be mishandled or ignored.
- Reporting tools that are confusing, inaccessible, or hidden within interface layers discourage users from submitting notices (DR-CM-08, DR-CM-15). When notices are filed, poor handling, such as long response times, unclear communication, or inconsistent decisions, undermines confidence in the Platform’s redress process (DR-CM-09 – DR-CM-12; DR-CM-16 – DR-CM-19). Lack of clear statements of reasons violates Article 17 DSA transparency duties.
- Insufficient allocation of trained moderators, coupled with sustained exposure to distressing or violent content, leads to fatigue, burnout, and reduced accuracy (DR-CM-13 – DR-CM-14). High turnover and lack of psychological support exacerbate inconsistency and delay.
- Failure to prioritise or systematically process notices from trusted flaggers (qualified entities under Article 22 DSA) diminishes the effectiveness of external oversight and cooperation with law-enforcement and civil-society partners (DR-CM-20).

Collectively, these drivers highlight that content-moderation risk is multifactorial, encompassing design, process, and human-resource dimensions. While automation increases scale, it also introduces bias and opacity; human moderation mitigates nuance but is limited by capacity and well-being. Ensuring effective measures is therefore critical to maintaining lawful, proportionate, and non-discriminatory moderation practices.

Manipulation and Inauthentic Use Drivers

Manipulation and inauthentic behaviour represent a cross-cutting operational risk for large user-generated content platforms. These behaviours may exploit platforms’ openness and engagement-based design to influence user perception and subvert recommendation systems. Coordinated inauthentic activity undermines the integrity and fairness of platforms. In the context of an adult-content platform, this is particularly concerning, because the same engagement mechanisms used for legitimate expression can be hijacked to spread CSAM links, coordinate abuse against creators, or artificially boost the visibility of exploitative material. The identified risk drivers are as follows:

- Insufficient safeguards against automated or coordinated manipulation allow bots, fake accounts, and engagement farms to artificially inflate metrics (DR-MA-01). This distorts recommender-system inputs and can elevate illegal or harmful material to wider audiences. The resulting artificial virality undermines user trust and creates unfair exposure advantages for malicious actors.
- Weak detection of brigading and mobbing campaigns in comment sections enables sustained, targeted harassment (DR-MA-02). Such coordination often takes the form of repeated abusive comments, doxing attempts, or coercive behaviour aimed at specific creators (e.g., women and LGBTQ+ performers). This contributes to a hostile environment, amplifies psychological harm, and may dissuade affected individuals from participation, infringing on the right to freedom of expression.
- Failure to adequately detect and filter URLs, QR codes, or link shorteners in comments allows malicious actors to distribute phishing, malware, or illegal content redirects (DR-MA-03). This not only facilitates dissemination of illegal content off platform but also exposes users to fraud and device compromise. Given that such links can mimic legitimate creator promotions, their early identification and blocking are critical to maintaining trust and legal compliance.

- Without robust safeguards against manifestly unfounded or coordinated abusive notices, hostile actors can misuse reporting systems to suppress lawful content and overwhelm moderation capacity (DR-MA-04). This “report bombing” creates procedural noise, delays legitimate enforcement actions, and can result in unjustified takedowns. The effect is both operational (diversion of resources) and rights-based (arbitrary restriction of speech).

The manipulation and inauthentic-use drivers demonstrate how platforms’ governance vulnerabilities can be exploited to reverse the intended protective function of moderation and recommendation systems. Inadequate detection of coordinated behaviour allows bad-faith actors to weaponize engagement features against both the Platform and its users. Addressing these risks requires a combination of technical countermeasures and policy-level interventions.

Regional and Linguistic Drivers

Regional and linguistic disparities represent a structural driver of systemic risk for platforms operating across multiple EU jurisdictions. Because most detection systems are designed initially for high-volume markets (particularly English-speaking ones), smaller or less-resourced linguistic environments often receive weaker oversight.

This imbalance creates unequal levels of user protection, inconsistent enforcement of national laws, and potential breaches of the principle of non-discrimination. In practice, gaps in linguistic and regional coverage allow harmful or illegal material to persist longer in some areas, while also giving malicious actors opportunities to exploit blind spots.

The identified risk drivers are as follows:

- Automated moderation and machine-learning models trained predominantly on English-language data may fail to identify illegal or harmful content expressed in less widely spoken EU languages (DR-RL-01). This includes regional slang, and dialects used in sexual exploitation or extremist material.
- A lower number of human moderators assigned to less common language markets leads to slower response times and delayed content removal (DR-RL-02). In small linguistic markets, even short delays can result in disproportionate exposure of victims and higher rates of content re-upload.
- Gaps in detection for non-major languages can be exploited by foreign interference operations, extremist propaganda groups, or traffickers who intentionally target weaker regulatory environments (DR-RL-03). These campaigns may use region-specific narratives to avoid detection and moderation.
- Moderation processes that are not adapted to national legal frameworks create compliance inconsistencies and potential conflicts with local authorities (DR-RL-04).
- Reporting and appeal mechanisms that are not translated or localised hinder users from exercising their rights to notify, appeal, and seek remedies in their own languages (DR-RL-05). This directly affects the Platform’s transparency and fairness obligations under Articles 16-20 DSA.
- Certain forms of CSAM, grooming, or trafficking-related content rely on region-specific slang, euphemisms, or coded references that automated detection systems and non-native moderators may fail to understand (DR-RL-06). These hidden cues can allow illegal networks to operate with relative impunity.
- Variations in recommender algorithm configurations across EU countries may result in uneven exposure to harmful or extreme content (DR-RL-07). Local engagement patterns can lead to feedback loops in certain linguistic markets, reinforcing harmful stereotypes or increasing exposure to illegal material.

Regional and linguistic drivers highlight the unequal distribution of risk across linguistic and geographic segments of the EU user base. They illustrate how operational asymmetries can translate into systemic inequality of protection. Mitigation requires sustained investment in linguistic diversity within moderation and detection tools.

Terms of Service Drivers

ToS serve as the central legal and operational framework that governs user behaviour and transparency obligations under Articles 14 DSA. Weaknesses in ToS design, implementation, or enforcement can significantly increase the likelihood of systemic risks, including inconsistent rule application, lack of user understanding, and erosion of fundamental rights such as freedom of expression and effective redress.

In practice, poorly structured or ambiguously worded, ToS provisions can create confusion for both users and internal reviewers, resulting in uneven enforcement, adversarial effects on freedom of speech, and exposure to liability. The assessment identifies multiple risk drivers that illustrate how ToS deficiencies can cascade into systemic failures.

The analysis identifies multiple potential risk drivers:

- Vague or inconsistent definitions of prohibited illegal or incompatible content cause interpretation discrepancies across jurisdictions and moderation teams (DR-TS-01). Such ambiguity undermines predictability and fairness in enforcement.
- Rules that are applied differently between countries/languages, or user groups create discriminatory enforcement outcomes and may violate the principle of equality of treatment (DR-TS-02).
- Where the ToS fail to explain the enforcement process, users cannot understand the basis for decisions or effectively contest them (DR-TS-03). This limits compliance with Article 17 DSA on statements of reasons and Article 20 DSA on complaint handling.
- Fear of regulatory penalties or reputational damage may lead to overly cautious removals of borderline content (DR-TS-04). This results in systematic over-removal, reducing media pluralism and curtailing lawful sexual or minority expression.
- A lack of robust escalation measures against repeat violators, such as re-registering non-consensual content uploaders, allows persistent harm to continue (DR-TS-05).
- Insufficient explanation of notice-and-action, complaint, or statement-of-reasons workflows limits user engagement with procedural rights (DR-TS-07, DR-TS-08).
- ToS and related rights information not localised to all EU languages hinder users in smaller or non-dominant linguistic markets from understanding their rights (DR-TS-09).
- Absence of clear and accessible contact points limits users' ability to seek assistance or resolve issues effectively (DR-TS-10).
- Missing or insecure contact-point procedures for official notices risk non-receipt or delayed enforcement (DR-TS-11 – DR-TS-13). These failures overlap with the Contact-Point driver category but originate from policy-design weaknesses embedded in ToS governance.

The ToS driver category highlights how governance and documentation quality directly influence compliance risk. Poorly structured or inconsistently applied rules compromise both regulatory obligations and user rights.

Contact Point Drivers

Maintaining a reliable and secure contact point for authorities is a core compliance obligation under Article 11 DSA. This function ensures that competent authorities across EU Member States can submit lawful orders (e.g., removal requests or data-disclosure demands) and receive timely, authenticated responses. Deficiencies in this process can cause regulatory non-compliance, delayed response to urgent law-enforcement requests, or even accidental disclosure of user data to fraudulent entities. As a result, this driver category directly influences platforms' ability to

cooperate effectively with national authorities while safeguarding users' rights to due process, privacy, and data protection. The identified potential risk drivers are as follows:

- Where no official and continuously monitored communication channel is available, legitimate removal orders, preservation requests, or information demands may not be received or processed at all (DR-CA-01). This creates a risk of non-compliance with national authorities and undermines trust in the platform's responsiveness.
- Weak or non-standard verification of incoming official requests can lead to erroneous compliance with fraudulent or spoofed notices (DR-CA-02). Such incidents can result in unlawful disclosure of user data or unjustified content removals.
- Official notices or judicial orders issued only in national languages that internal staff cannot interpret may cause delays in response or misinterpretation of legal obligations (DR-CA-03). This is particularly relevant for a cross-border platform serving all EU Member States, where procedural formats and terminologies differ significantly.

The contact-point function, though administrative in nature, is a critical control in platforms' regulatory interface. Weaknesses in this area can lead to serious procedural failures. The robustness of this process directly underpins the platforms' credibility with regulators and law-enforcement agencies, ensuring that lawful orders are executed diligently while protecting against unlawful or abusive requests.

Advertising System Drivers

The Platform relies on advertising to generate income. Advertising systems represent a critical intersection of commercial operations and systemic-risk governance. Under Articles 26 and 39 DSA platform providers must ensure that advertising on their services is transparent and does not contribute to illegal or harmful content dissemination. In adult-content environments, advertising carries heightened sensitivity due to its proximity to sexual material, the potential targeting of minors, and the risk of monetising exploitative or degrading content.

Weaknesses in ad-system governance can therefore magnify several systemic risks, including discrimination, public-health harms, consumer deception, and exposure of minors to explicit or unsafe material. Algorithmic optimisation for revenue and engagement can also incentivise harmful amplification, undermining platforms' duty of care and compliance obligations. The identified risk drivers are as follows:

- When the Platform fails to clearly disclose advertiser identity, sponsorship, or targeting criteria, hidden actors can manipulate users or launder influence through pornographic contexts (DR-AS-01). This impairs transparency and violates Article 26 DSA.
- Ad-ranking systems optimised for engagement may disproportionately promote degrading or fetishised porn categories, reinforcing harmful gender or racial stereotypes (DR-AS-02). Such bias contributes to long-term discrimination and reputational risk.
- Algorithms that prioritise click-through rates over content integrity can systematically surface risky, extreme, or unlawful ads because they attract higher engagement (DR-AS-03).
- Weak verification of advertisers enables traffickers or fraudulent entities to purchase ad inventory anonymously (DR-AS-04). Such actors can weaponize the Platform's monetisation ecosystem to profit from illegal activity.
- Lack of effective ad-category screening allows unsafe sexual-enhancement products, unverified treatments, or illicit substances to be promoted to users (DR-AS-05). This raises both public-health and consumer-protection concerns.

- Absence of safeguards against covert political or state-linked advertising allows disinformation and electoral manipulation within pornographic spaces, exploiting anonymity and reduced scrutiny (DR-AS-06).
- When ad-systems fail to exclude or detect non-consensual, coercive, or abusive sexual content, they inadvertently incentivise its continued circulation for profit (DR-AS-07). This driver represents one of the most severe forms of systemic risk, connecting financial incentives directly to fundamental-rights violations.

Advertising-system drivers highlight the inherent tension between commercial optimisation and fundamental -rights compliance. Ad revenue models built on engagement or click-through metrics can unintentionally reward illegal or incompatible materials. Strong advertising governance not only mitigates legal and reputational risk but also supports user trust, public-health protection, and fair treatment of creators and consumers.

Recommender and Algorithmic Systems Drivers

Recommender and algorithmic systems are among the most influential components of platforms' risk architecture. By determining what content users see, how it is prioritised, and how engagement is rewarded, these systems act as key amplifiers of both beneficial and harmful dynamics. Under Articles 27 and 38 of the DSA, platforms must ensure that recommender logic is transparent, offers user choice, and does not systematically amplify illegal or harmful content.

In the adult-content context, recommender algorithms shape user exposure to sexual material, affecting not only freedom of expression and pluralism but also public health, gender equality, protection of minors, and civic integrity. Poorly designed or unmonitored algorithms can inadvertently prioritise illegal or extreme material, reinforce discriminatory stereotypes, or promote coercive or exploitative content that breaches the platforms' duty of care.

The identified risk drivers are as follows:

- Algorithms designed to maximise engagement may unintentionally prioritise illegal or harmful material, including non-consensual intimate imagery, revenge porn, and deepfakes (DR-RS-01), or automatically promote new uploads and trending categories without prior moderation review (DR-RS-02).
- Engagement-based sorting of comments, chats, and interactions (DR-RS-03) can elevate grooming attempts, phishing links, or sexually explicit spam, while the same algorithms may favour pirated or stolen paywalled material because it attracts higher click-through rates (DR-RS-04). This dynamic undermines both user safety and intellectual-property rights.
- Machine-learning models trained on unbalanced data can reinforce degrading or discriminatory categories—such as racial fetishisation, misogyny, or body-shaming tropes, while suppressing minority or LGBTQ+ creators (DR-RS-05). Furthermore, similarity-based clustering and “related content” features can form echo chambers (DR-RS-06), gradually normalising violent or extremist content. Over time, these processes foster cultural and psychological harms, including body-image distortion and desensitisation (DR-RS-07).
- Users frequently lack meaningful insight into how recommendation parameters' function or how to adjust them (DR-RS-09). The absence of a non-profiling or chronological feed option (DR-RS-10) restricts user autonomy and contravenes DSA user-choice duties. In parallel, the use of sensitive personal data or inferred characteristics (e.g., sexual orientation, religion, or political views) in profiling logic (DR-RS-11) creates discrimination and data-protection risks under Article 9 GDPR.
- The recommender system may inadvertently surface disinformation or polarising political narratives embedded within adult-content contexts (DR-RS-12). Such covert dissemination risks influencing public opinion and undermining democratic discourse, particularly during election periods.

The analysis shows that algorithmic systems are central amplifiers of systemic risk. When governed solely by engagement or monetisation metrics, they can transform from neutral discovery tools into risk-propagation

mechanisms. By embedding transparency, fairness, and accountability into recommendation design, platforms can significantly reduce its exposure to DSA systemic-risk categories and strengthen user trust.

Data-Related Practice Drivers

Data-related practices underpin nearly every aspect of platforms' operation, from content personalisation and age assurance to advertising and analytics. Consequently, weaknesses in data protection or access control can directly generate systemic risks under both the DSA and the GDPR, particularly in areas involving sensitive or intimate data.

For an adult-content platform, data risk is heightened: user interactions, uploads, and metadata inherently carry special-category information that can expose individuals to serious privacy harm if mismanaged. The identified drivers reflect governance, security, and transparency challenges across the full data lifecycle:

- Inadequate technical and organisational safeguards can lead to unauthorised access or breaches of highly sensitive materials, including intimate videos and personal identifiers (DR-DP-01). Weak encryption, poor key management, or unmonitored external storage significantly increase exposure. Compromised accounts or leaks can have long-term consequences for affected users, including harassment and extortion.
- Continuous tracking through cookies, device fingerprints, or ad-tech identifiers (DR-DP-02) allows the platform — and potentially third parties — to infer intimate behavioural patterns. Combined with insufficient consent mechanisms (DR-DP-03), this undermines user autonomy and may violate principles of purpose limitation and data minimisation under the GDPR.
- Users often face complex or opaque privacy interfaces that lack clear opt-in/opt-out options, granular controls, and understandable language (DR-DP-03). As a result, they are unaware of what data are collected, how profiling operates, or with whom information is shared, eroding informed consent and user trust.
- Even when explicit content is stored securely, associated metadata (e.g., geolocation, IP address, contact details, timestamps) may be insufficiently protected (DR-DP-04). When combined with external datasets, such metadata can deanonymize users, revealing private behaviours or identities.
- The indefinite retention of sensitive material and metadata without effective deletion or erasure rights (DR-DP-06) breaches both GDPR principles and users' rights to privacy and control over personal information. Such practices also enlarge the impact of any subsequent data breach.
- Sharing of user or content-derived data with advertisers, affiliates, or analytics vendors without clear safeguards or audit trails (DR-DP-07) creates secondary exposure channels. Weak vendor-due-diligence or reliance on opaque ad-tech intermediaries may enable unauthorised secondary processing.
- Weak internal permissions or audit logging allow staff or contractors to view, copy, or leak intimate material (DR-DP-08). Such incidents pose severe reputational and legal risks, especially if no effective whistle-blower or accountability mechanism is in place.

Data-related risks are both foundational and systemic: they do not only concern cybersecurity, but also transparency and governance. Mitigating these drivers requires a comprehensive data-protection and governance framework.

Age-Assurance Drivers

Age assurance is a critical protective and compliance function for platforms hosting adult or explicit content. Under Articles 28 and 35 of the DSA, platforms must take appropriate and proportionate measures to prevent minors from accessing pornographic material and to ensure that data processing linked to age assurance complies with privacy and data-protection law.

Inadequate or poorly designed age-assurance systems expose both minors and adults to harm. Weak age-assurance mechanisms may allow underage users to access or be exposed to explicit content and advertising, while excessively intrusive methods may compromise data protection and privacy. Striking the right balance between protection of minors and respect for users' fundamental rights (privacy and informational self-determination) is therefore central to the Platform's compliance strategy.

The analysis identifies multiple risk drivers:

- Insufficient or absent age assurance in interactive features such as comments or private messaging enables malicious actors to potentially contact and groom minors (DR-AV-01). This risk is heightened in environments where pseudonymity is allowed.
- Inadequate age-assurance controls may allow minors to view or interact with adult advertisements (DR-AV-02). These exposures can cause psychological harm and contravene legal restrictions on advertising to minors under EU consumer-protection frameworks.
- Certain age-assurance solutions require users to upload biometric data or government-issued identification without adequate technical safeguards (DR-AV-03). This introduces significant data-protection and cybersecurity risks, as intimate personal data could be misused or unlawfully shared.
- Age-assurance systems that lack clear explanations of how data are processed, retained, or anonymised erode user trust (DR-AV-04). Without transparency, users may feel coerced into invasive methods, undermining informed consent and DSA transparency requirements.
- Where age-assurance controls can be easily circumvented, minors can still access the Platform and encounter explicit content (DR-AV-05). This directly conflicts with the duty of care to protect minors and exposes the Platform to regulatory sanctions.
- Weak encryption, excessive retention, or unclear data-deletion protocols for age-verification materials (IDs, biometric scans, or tokens) create risks of data breaches, and internal misuse (DR-AV-06). The sensitivity of such data magnifies the impact of any incident.

Age-assurance drivers demonstrate the tension between protection of minors, privacy protection and security of users. Systems that are too lenient risk exposing minors to potential harm, while systems that are too intrusive may infringe privacy, cause security risks and deter legitimate adult use. Effective mitigation therefore requires a risk-based and privacy-preserving architecture, guided by principles of data minimisation, proportionality, and transparency.

5. Risk Assessment

5.1 Inherent Risk Assessment

Inherent Risk Assessment Methodology

An inherent risk assessment evaluates the potential risks to NKL, Platform's users or society without regard to any existing controls or risk mitigation strategies. This assessment focuses on two main criteria: the likelihood of a risk event occurring and the potential impact of that event. Each criterion is scored on a scale of 1 to 5, with higher scores indicating greater likelihood or more severe impact.

The **likelihood** of a risk event is assessed using the following scale:

- **Rare (Score 1):** The risk event is highly improbable to take place within the next year considering no historical occurrences and robust preventative measures elsewhere.
- **Unlikely (Score 2):** The risk event has a low chance to take place within the next year, possibly due to rare instances or minimal vulnerabilities.
- **Possible (Score 3):** The risk event has a moderate chance of occurring within the next year, supported by occasional past occurrences and some identified vulnerabilities.
- **Likely (Score 4):** The risk event is expected to occur within the next year, indicated by frequent past instances and significant vulnerabilities.
- **Highly Likely (Score 5):** The risk event is almost certain to occur within the next year, with ongoing incidents or critical vulnerabilities present.

The potential **impact** of a risk event is evaluated on the following scale:

- **Insignificant (Score 1):** Minimal or no potential harm to users, society, or the organization. Regulatory action is unlikely, and any reputational impact would be immaterial.
- **Minor (Score 2):** Potential harm to users or society, such as exposure to inappropriate content or limited misinformation spread. The organization might face minor reputational damage or warnings from regulators.
- **Moderate (Score 3):** Notable potential for harm to users or society, including safety risks, privacy breaches, or erosion of trust. The organization could experience moderate reputational damage or regulatory investigations.
- **Major (Score 4):** Significant potential for widespread harm to users or society, such as extensive exposure to harmful content or manipulation of essential processes. The organization may face substantial reputational damage, operational suspensions, or significant fines.
- **Critical (Score 5):** Severe potential for extensive harm to users or society, including exploitation, psychological harm, or significant disruptions to societal functions. The organization could suffer severe reputational damage, loss of operational licenses, or complete shutdown.

Once the likelihood and impact are determined, the inherent risk value is calculated using the Risk Value Matrix (see Figure 1 below). This matrix cross-references the likelihood score with the impact score to assign an overall risk value:

$$\text{Inherent Risk Score} = \text{Likelihood Score} \times \text{Impact Score}$$

Figure 1 – The Risk Value Matrix determines the categorisation of risk level
(source: NKL’s internal document – Risk Register and Dashboard)

RISK VALUE MATRIX		(x-axis: impact; y-axis: likelihood)				
		1	2	3	4	5
		Insignificant	Minor	Moderate	Major	Critical
1	Rare	1	2	3	4	5
2	Unlikely	2	4	6	8	10
3	Possible	3	6	9	12	15
4	Likely	4	8	12	16	20
5	Highly Likely	5	10	15	20	25

The color-coded risk values provide a visual representation of the risk level:

- **Low Risk (Green):** Risk values between 1 and 5 indicate a low level of inherent risk, representing isolated or unlikely events with minimal Platform-wide consequences. These scenarios typically involve rare occurrences, limited exposure, or narrow systemic implications.
- **Medium Risk (Yellow):** Risk values between 6 and 15 represent a medium level of inherent risk, typically involving known vulnerabilities or systemic weaknesses that could become severe if not effectively managed.
- **High Risk (Red):** Risk values between 16 and 25 indicate a high level of inherent risk, denoting significant exposure where both probability and impact are elevated. These risks are often systemic, recurring, or structurally embedded, and could substantially affect user safety or Platform viability.

The inherent risk framework establishes a baseline risk landscape that guides prioritisation of mitigation and governance resources. It does not account for control strength or detection capacity but helps to identify areas where structural safeguards must exist by design (e.g., proactive moderation, privacy-by-default, and child protection measures).

Inherent Risk Assessment Approach

Cross-Functional Workshops

An inherent risk assessment approach was applied through a series of structured cross-functional workshops. The workshops were designed to foster collaborative discussion, ensure multi-functional input, and capture interdependencies among risk categories. Each team assessed risks considered directly and indirectly relevant to its domain. The following teams participated in the workshops:

- Content Moderation Team,
- Notice & Complaint Team,
- IT Team,
- Support Team,
- Advertising Team.

Additionally, the Legal Team provided cross-cutting legal analysis and supported impact evaluation across all risk categories, and the Regulatory Director focused primarily on assessing risks related to the protection of minors.

The workshops were facilitated and coordinated by the Compliance Team, which also ensured documentation quality and alignment of methodologies.

The assessment was conducted through a multi-functional approach to ensure comprehensive coverage of all relevant systemic risk domains. The focus areas varied according to the nature of each team's activities and their connection to specific systemic risk categories. Certain teams assessed risks directly linked to their core operations and subject-matter expertise. Because systemic risk categories are interdependent, the boundaries between assessment areas were intentionally flexible. Many scenarios were evaluated collaboratively to reflect overlapping domains. This integrative approach enabled a more accurate reflection of real-world interconnections, where a single risk event can influence multiple categories simultaneously (e.g., cross-connections between fundamental rights and the dissemination of illegal content).

Each workshop followed a structured agenda:

- Identification and discussion of key systemic risk scenarios relevant to the teams' area,
- Evaluation of likelihood and potential impact using the predefined scale,
- Analysis of risk drivers relevant to the teams' area,
- Review of existing mitigation measures (for context only - these were not factored into inherent scoring),
- Documentation of conclusions and validation of interdependencies with other teams' scenarios.

The overall assessment therefore combined distinct expert inputs into a unified view of inherent risk exposure. To ensure a balanced representation of viewpoints, the final inherent risk score for each scenario was calculated as the arithmetic mean of all team submissions. This averaging method provided a composite measure reflecting the collective judgment of all relevant functions. The finalized set of inherent risk scores represented the collective institutional view of systemic risk exposure prior to any mitigating measures.

Compliance Team Review

After the cross-team workshops, the Compliance Team conducted a review to perform an independent, evidence-based sense check of the workshop outputs and to ensure that the final inherent risk reflects internal expertise and the available external evidence. Where public sources provided clear, credible indications about prevalence or harm, the Compliance Team aligned ratings accordingly - raising or, in rare cases, lowering the inherent score to maintain methodological consistency with the external record.

The Compliance Team review covered all risk scenarios included in the inherent risk matrix. Particular attention was given to cases meeting, where external or regulatory evidence indicated a material difference in impact or harm severity compared to the workshop consensus. During this process, the Compliance Team reviewed both the likelihood and impact dimensions. The resulting final inherent risk score therefore represents the arithmetic mean of all team inputs, following any adjustments made during the Compliance Team validation step.

To ensure objectivity, the Compliance Team triangulated information across several classes of independent, reputable sources:

- Independent research and NGO studies: including academic papers, sectoral meta-analyses, and publications from recognized human rights and digital safety organizations;
- Industry and Platform reports: such as transparency data published by comparable services and advertising networks;
- Verified media reporting: considered only when supported by verifiable evidence or corroborated by official or academic sources.

Where internal data were robust and verifiable, they remained the primary basis for evaluation. External references were used primarily to contextualize or calibrate the likelihood or impact, particularly in areas with limited internal sample size or high uncertainty.

The following examples illustrate how the Compliance Team applied this validation approach:

- IR-IC-05 – CSAM Distribution: Operational teams reported recurring detection of attempted CSAM uploads, confirming the threat’s persistence. Compliance Team raised the likelihood from Likely to Highly Likely based on verified patterns across major industry actors. Mindgeek, the parent company of multiple adult content companies including Pornhub, YouPorn, RedTube, Brazzers, and more, reported 13,229 instances of child exploitative content that year; 4,171 of those were unique reports (Cole, 2021)³². Between 2021 and mid-2024, Pornhub’s transparency reports show a decline in removals related to child sexual abuse material (CSAM) from 11.7% to 0.9%, though total removals rose due to broader moderation efforts. XVideos and Stripchat reported smaller shares (0.96% and 0.6%, respectively) (Happ, Harpenau, & Wiewiorra, 2024)³³. Given the legal and moral gravity of CSAM, the impact remained Critical, as any single verified instance can trigger severe sanctions, reputational harm, and irreversible victim trauma.
- SR-GB-05 – Deepfake and NCII: According to Keepnet Labs (2025)³⁴, 96–98 % of all deepfake content online consists of non-consensual intimate imagery (NCII), and 99–100 % of victims are female. Researchers (Mosqueda 2020)³⁵ documented a Telegram bot that auto generated over 100 000 fake nude images of women by mid-2020. ESET UK (2023)³⁶ found that two in five (39 %) people view deepfake pornography as a significant personal risk of sharing intimate content. Among those whose images were misused, 46 % felt shame or embarrassment, and 19 % said they would not seek help, underscoring under-reporting and long-term psychological harm. This supports a “Likely” or even “Highly Likely” likelihood score. Impact was raised from Moderate to Major to reflect the substantial reputational and legal implications.
- SR-PH-02 – False or misleading health information: For this risk scenario the Compliance Team confirmed the other teams’ evaluation. It was found that this topic is primarily relevant for social media platforms - 82% of adult social media users report seeing some or a lot of false or misleading health information (46% “some,” 36% “a lot”), while 18% report seeing none or a little. Additionally, 67% of respondents say they find it difficult to determine whether health information on social media is true or false (Vivion et al., 2025)³⁷. However, no reliable evidence was found indicating that misleading health information is being disseminated on adult tube platforms. Therefore, the likelihood assessment was based mainly on the operational teams’ evaluation. The impact assessment is also confirmed, because the World Health Organisation (WHO)³⁸ has warned that widespread medical misinformation, such as false claims about treatments or vaccines, poses

³² Cole, S. (2021, April 2). Pornhub just released its first transparency report. VICE. <https://www.vice.com/en/article/pornhub-just-released-its-first-transparency-report/>

³³ Happ, M., Harpenau, F., & Wiewiorra, L. (2024). Economics and regulation of adult online content (WIK Working Paper No. 9). WIK Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste. <https://www.econstor.eu/bitstream/10419/308077/1/1913368831.pdf>

³⁴ Keepnet Labs. (2025, September 24). Deepfake statistics & trends 2025: Growth, risks, and future insights. <https://keepnetlabs.com/blog/deepfake-statistics-and-trends>

³⁵ Mosqueda, S. (2020, October 22). Deepfake bot creates pornographic images of thousands of women. ASIS International. <https://www.asisonline.org/security-management-magazine/latest-news/today-in-security/2020/october/thousands-of-women-abused-by-deepfake-porn-bot/>

³⁶ ESET UK. (2023, December 12). Nearly two-thirds of women worry about being a victim of deepfake pornography, ESET UK research reveals. ESET. <https://www.eset.com/uk/about/newsroom/press-releases/nearly-two-thirds-of-women-worry-about-being-a-victim-of-deepfake-pornography-eset-uk-research-reveals/>

³⁷ Vivion, M., Reid, V., Trottier, V., Bergeron, F., Savard, I., Dionne, E., & Tourigny, A. (2025). Interventions to counter health misinformation among older people: Protocol for a scoping review. JMIR Research Protocols, 14, e74138. <https://doi.org/10.2196/74138>

³⁸ World Health Organization. (2024, February 6). Disinformation and public health [Questions & answers]. <https://www.who.int/news-room/questions-and-answers/item/disinformation-and-public-health>

a serious public health threat, as it fosters distrust toward experts and discourages adherence to legitimate medical advice. Nevertheless, because adult content platforms are not generally perceived as credible sources of health information, the impact level has been assessed as “Moderate”, which is considered appropriate and proportionate.

- SR-PM-09 – Exposure of minors to misleading “educational” pornography: The Compliance Team reviewed the initial team assessment that categorized this risk as “Possible” and “Moderate”. While operational teams noted that explicit “educational” or pseudo-informational content forms only a small portion of total uploads, the Compliance Team review identified credible evidence that exposure of minors to misleading or distorted sexual information through pornography is frequent across digital ecosystems: 52% of teen viewers reported seeing pornography depicting rape, choking, or someone in pain; only 33% said they had ever seen porn where someone asked for consent before sex. Teens who viewed violent porn were more likely to think “most people like being hit during sex” (28%) or that putting hands around someone’s neck is safe (20%). The report also documents early and accidental exposure (avg first exposure ~12; 15% by age ≤10). These findings speak directly to “distorted views” and “confusion.” (Common Sense Media, 2022)³⁹. Exposure to sexualized media increased acceptance of token resistance (belief that “no” may mean “yes”), a specific consent-related distortion relevant to safety design and education (Van Oosten et al, 2015)⁴⁰. Other research summarizes longitudinal findings where adolescent pornography use predicted more permissive sexual attitudes over time (not vice-versa in several studies). Good umbrella reference (Peter, J., & Valkenburg, P. M., 2016)⁴¹. Based on the evidence the Compliance Team re-evaluated the likelihood of this risk scenario as “Likely”. The team concluded that misleading or pseudo-educational pornographic content represents a plausible and recurring exposure pathway for minors.

Downward revisions were exceptionally rare and only applied where the available data demonstrated a low level of exposure and minimal foreseeable harm, or where the risk description overlapped with another, higher-priority scenario.

The Platform recognizes that some systemic-risk domains currently lack robust or standardized public datasets. Where comprehensive or comparable statistical evidence was unavailable, the Compliance Team relied on the expert judgment and operational experience of internal functional teams as the primary validation source. In these instances, the team’s practical exposure to moderation data was treated as credible qualitative evidence.

Furthermore, the Platform does not yet operate a fully developed quantitative Key Performance Indicator (KPI) or metrics framework for systemic-risk tracking. As a result, it was not possible to corroborate every scenario with internal numerical indicators such as detection rates or incident volumes. In the absence of these quantitative datasets, the Compliance Team applied an approach that combined internal insights with peer-platform data and academic research, where available.

Inherent Risk Assessment Outcomes

Overall Posture

A total of 81 inherent risk scenarios were evaluated across all categories. The average inherent risk score of 14 places the Platform within the medium overall risk band. However, the distribution of scores is strongly skewed toward higher-severity scenarios, with 38 classified as high risk and 36 as medium risk (see Figure 2). This distribution

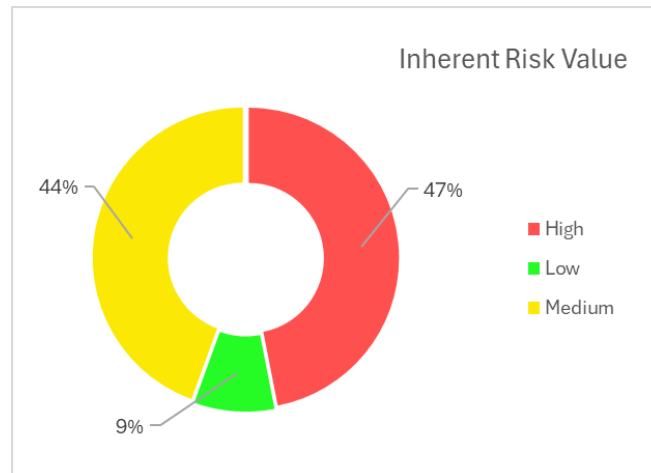
³⁹ Common Sense Media. (2022). Teens and pornography: A survey of U.S. teens ages 13–17 [PDF]. <https://www.commonsensemedia.org/sites/default/files/research/report/2022-teens-and-pornography-final-web.pdf>

⁴⁰ Van Oosten, J. M. F., Peter, J., & Valkenburg, P. M. (2015). The influence of sexual music videos on adolescents’ misogynistic beliefs: The role of video content, gender, and affective engagement. *Communication Research*, 42(7), 986-1008. <https://doi.org/10.1177/0093650214565893>

⁴¹ Peter, J., & Valkenburg, P. M. (2016). Adolescents and pornography: A review of 20 years of research. *The Journal of Sex Research*, 53(4-5), 509-531. <https://doi.org/10.1080/00224499.2016.1143441>

indicates that nearly half of all identified risks fall into the high-risk category, signalling a materially elevated baseline exposure characteristic of adult-content environments. High inherent risk levels reflect structural vulnerabilities that exist before the application of any mitigating controls.

Figure 2: Distribution of inherent risk values across all categories
(source: NKL's internal document – Risk Register and Dashboard)



Concentration of Risks

A total of 12 risk scenarios were assessed at the maximum inherent severity (25/25) representing the Platform's most critical exposure areas. These concentrate in:

- **Dissemination of Illegal Content:** Includes repeated high-severity cases of NCII, revenge pornography, and AI-generated deepfakes depicting real individuals without consent. This category also encompasses CSAM, both in direct uploads and through off Platform linking mechanisms (e.g., QR codes, shortened URLs, redirections).
- **Privacy and Family Life:** Encompasses severe invasions of privacy, including NCII and doxing, the leakage of paywalled or private intimate material, and blackmail-related dissemination,
- **Data Privacy and Protection:** Refers to large-scale exposure of sensitive identifiers (e.g., metadata, images, contact details) that can be used for blackmail, profiling, and unlawful data exploitation.
- **Protection of Minors:** The presence or accessibility of CSAM, grooming, and weak or bypassable age-assurance mechanisms make this category critical. Exposure is compounded by secondary pathways, such as discovery features, recommendation systems, or comments, that inadvertently expose minors to explicit or harmful material.

The analysis of the inherent risk distribution across categories (see Figure 3) reveals that the Platform's exposure is defined by the concentration and systemic clustering of severe scenarios.

Figure 3 Average inherent risk scores by risk category
 (source: NKL's internal document – Risk Register and Dashboard)

Risk Category	Number of Risks	Inherent Risk Score
Privacy & Family Life	5	22
Data Privacy and Protection Risks	5	20
Protection of Minors	11	18
Physical and Mental Well-being	4	15
Dissemination of Illegal Content	19	15
Consumer Protection	6	14
Gender-Based Violence	8	14
Human Dignity	2	14
Public Health	5	12
Public Security Concerns	4	10
Non-Discrimination	4	8
Freedom of Expression and Information	3	5
Civic and Electoral Impact	5	5
Total / Average Inherent Risk Score	81	14

Category-Level Analysis

The Privacy and Family Life category records the highest average inherent risk score (22) across five risk scenarios, making it the single most critical area of exposure. This domain is dominated by cases of NCII, doxing, and deepfake exploitation involving real individuals. These practices violate personal dignity and expose the Platform to significant cross-border legal liability under human rights and data protection frameworks.

Closely following is the Data Privacy and Protection category with an average inherent score of 20 across five risk scenarios. The nature of adult-content platform creates inherent systemic vulnerability, because highly sensitive and identifiable material is processed and stored. Risks in this category arise from large-scale exposure of intimate identifiers, metadata, or sexual-preference data, which can be exploited for blackmail, coercion, unlawful profiling, or secondary data use. This creates material legal risk under the GDPR and reputational exposure.

The Protection of Minors category ranks third in severity, with an average inherent risk score of 18 across 11 identified scenarios. This represents a cluster of persistent, high-critical risks linked to CSAM, grooming, minor exposure to pornography through weak or bypassable age-assurance, and harmful or sexualised advertising. Given the strict legal and moral frameworks governing protection of minors, these scenarios are universally considered “critical.”

When considering exposure through volume, the Dissemination of Illegal Content category stands out as the largest and most complex risk surface, comprising 19 distinct risk scenarios with an average inherent score of 15. Although its average severity sits at the upper end of the medium range, the breadth of this category significantly amplifies systemic exposure. It includes diverse and high-impact threats such as NCII, CSAM uploads and off-Platform link-sharing, malware and phishing links in comments, hate speech, copyright and IP violations, prohibited or violent sexual content, and even extremist propaganda.

The Gender-Based Violence category, while containing fewer scenarios (8 risks; average score 14), remains a persistent exposure area. It captures the normalisation and monetisation of NCII or violent sexual material, the creation and distribution of deepfake pornography targeting women, and advertisements that promote or profit from abusive content. These issues are particularly sensitive in terms of ethical responsibility and compliance with gender equality and anti-violence regulations.

Beyond content harms, advertising and consumer-protection risks (e.g., deceptive designs, dark patterns, inadequate user remedies) extend exposure into the Platform’s commercial practices, directly affecting trust and regulatory compliance. Additionally, comment-based threats (phishing, malware, or illicit solicitation) reveal secondary but significant risk vectors that might bypass formal moderation.

At the lower end of inherent severity, categories like Freedom of Expression, Civic and Electoral Impact, and Public Security Concerns carry modest scores but remain policy sensitive. Missteps in these areas could rapidly escalate into reputational or legal challenges.

These results provided the analytical foundation for the subsequent Residual Risk Assessment and for the development of mitigation and monitoring strategies.

5.2 Risk Drivers Assessment

Risk Drivers Assessment Methodology

The assessment methodology for risk drivers was conducted on a residual-risk basis. This approach was adopted because the risk drivers represent the influencing factors described under Article 34(2) of the DSA, which concern the design and functioning of systems and processes already established within the Platform. The assessment therefore does not evaluate hypothetical or uncontrolled situations but rather considers the potential risks of failure or limitation within existing operational mechanisms. Since these drivers correspond to systems that are already functioning, the appropriate analytical lens is residual risk, which reflects the level of exposure that remains after current controls have been applied and are operating as designed.

Residual risk scores were derived using the formula ***Likelihood × Impact = Score (1–25)***, where both parameters were estimated qualitatively, considering the presence and maturity of existing safeguards. Control measures were recognized as implemented and operating as designed, but their precise efficiency was not subjected to numerical modelling or audit-level validation.

An influence type was assigned to each driver to indicate its dominant mode of effect, whether it primarily increases likelihood (e.g., weak URL filtering, inconsistent moderation enforcement) or amplifies impact (e.g., opaque user communications or limited redress mechanisms). This classification enables a clearer understanding of how operational and technical factors shape the overall residual exposure.

To reflect the relative influence strength of each driver on the underlying systemic risks, a driver-modifier scale was applied. This scale allows small, evidence-based adjustments to the final residual score of systemic risks in cases where the driver demonstrably shifts either the likelihood or impact dimension despite the presence of established controls. The following adjustment bands were used:

- **Low influence:** The driver exerts negligible residual effect after controls; no adjustment applied.
- **Medium influence:** The driver retains a limited but measurable influence on either likelihood or impact, as supported by workshop evidence or observed performance patterns.
- **High influence:** The driver exerts a strong and persistent effect, either by significantly raising the probability of occurrence or by materially worsening potential consequences, even after mitigations are in place.

These modifiers were applied qualitatively and judgement-based, reflecting expert assessment. The purpose was to acknowledge observable differentials across drivers without implying precision beyond the level of organisational risk tolerance or data availability. All adjustments were discussed during cross-functional workshops and recorded transparently in the Risk Drivers Register to ensure traceability and methodological consistency.

Risk Drivers Assessment Approach

A residual-risk assessment approach was used to evaluate the risk drivers, applying the same structure and process as the systemic inherent risk assessment (see Section 5.1, *Inherent Risk Assessment Approach*). In practice, both assessments were carried out together during the same cross-functional workshops. The workshops followed a two-part analytical sequence. In the first stage, teams reviewed systemic risk scenarios to align on the nature of potential

harms. In the second stage, the same participants proceeded to the risk driver assessment. Teams based their judgements on direct operational experience.

Each driver was reviewed in terms of:

- The existing controls and measures already in place,
- The residual likelihood and impact of failure or underperformance,
- The influence type (whether the driver primarily affects likelihood or impact).

Following the workshops, a structured validation process was undertaken. The Compliance Team verified that each driver’s rationale and influence type were logically consistent with the internal processes and related systemic risks.

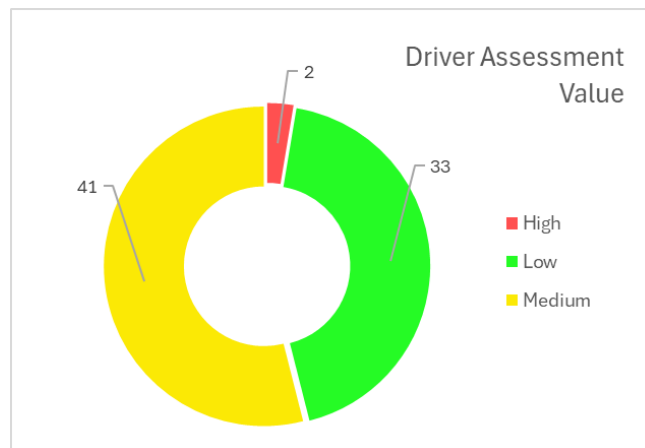
The Compliance Team also ensured that each risk driver was connected to the corresponding systemic risk entry in the *Systemic Risks Identification Catalogue*, which serves as a detailed analytical foundation for this assessment. This catalogue provides for every systemic risk linked risk drivers and related influence factors aligned with Art. 34(2) DSA.

Risk Drivers Assessment Outcomes

Overall Distribution

The evaluation of 76 Risk Drivers was completed across all operational and governance domains. Across all categories, the majority of Risk Drivers were rated as medium (41 drivers; 54%) or low (33 drivers; 43%), with only 2 drivers (3%) rated as high (see Figure 4). This distribution confirms that most operational processes are functioning effectively under current control conditions, with residual vulnerabilities concentrated in a small number of specific governance and user-rights areas.

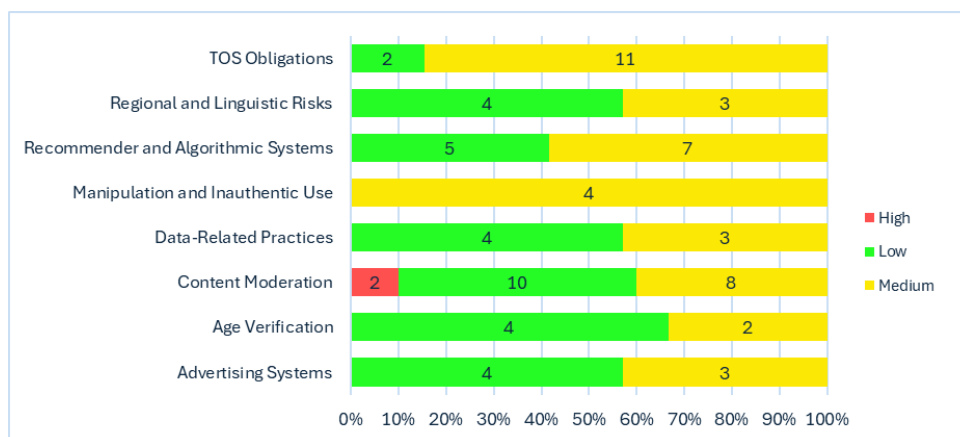
Figure 4: Distribution of risk driver values
(source: NKL’s internal document – Risk Register and Dashboard)



Concentration of Drivers

The distribution of residual risk by driver category is visualised in Figure 5. In the figure, each category’s count of drivers is broken down by risk band (low, medium, high), providing an at-a-glance view of where residual exposure is most concentrated.

Figure 5: Overview of risk drivers across categories and risk levels
 (source: NKL's internal document – Risk Register and Dashboard)



The two high-influence drivers, both within Content Moderation, arise from transparency and redress processes. Opportunities for enhancement were observed in the clarity of statements of reasons provided for enforcement actions involving textual and static image content. These factors were assessed as contributing to a higher relative impact severity within this content category.

The medium-influence drivers, which constitutes the operational core of exposure, captures drivers that either raise the likelihood of harm (e.g., through moderation coverage gaps, URL/link exposure, recommender effects, or parity issues) or amplify impact when incidents occur (e.g., through limited transparency, incomplete consent management, or minor-access vulnerabilities). These risks reflect areas where controls are present but not yet fully optimised or uniformly applied.

The low-influence drivers represent areas where layered controls are demonstrably effective, including automated detection systems, law-enforcement cooperation, internal access controls, and GDPR-aligned data management. While these drivers maintain strong performance, several relate to scenarios that would still produce critical impact if controls failed (e.g., data breaches, illegal content exposure, or moderation escalation lapses). As such, they are retained under continuous monitoring to ensure that control performance remains stable and verifiable.

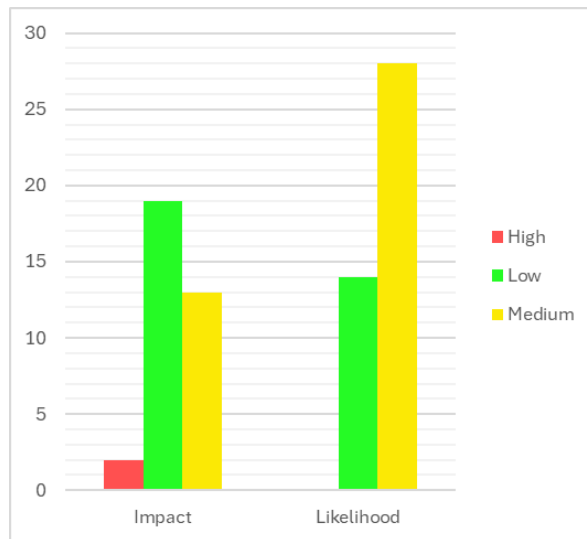
Influence Dynamics

The analysis of influence type (see Figure 6), whether a driver primarily affects the likelihood of an adverse event or the impact once an event occurs), reveals a clear structural pattern.

Out of 76 total risk drivers, 42 (55%) were classified as likelihood-oriented, while 34 (45%) were identified as impact-oriented. This near-balanced split indicates that systemic exposure is shaped both by operational conditions that make harmful events more probable and by governance factors that determine how severe their consequences become once they occur.

Impact-type drivers include 2 high, 13 medium, and 19 low items. These tend to correspond to transparency and rights-protection processes, such as clarity of statements of reasons and accessibility of appeals with respect to certain types of content, consent management, and user-communication flows. Impact drivers are consequential, because they define the severity and public visibility of harm. Failures here can transform otherwise contained incidents into rights-level or regulatory crises.

Figure 6: Influence of risk drivers on systemic risk
 (source: NKL's internal document – Risk Register and Dashboard)



Likelihood-type drivers with 28 medium, and 14 low items, dominate the operational risk space. These relate to process performance and technical effectiveness, including moderation coverage, language parity, comment and link filtration, recommender logic, and system scalability. Their residual ratings show that while strong control baselines are in place, a number of mechanisms continue to exert incremental probability pressure on systemic risks such as Dissemination of Illegal Content, Protection of Minors, and Consumer Protection. Addressing these areas yields immediate benefits in lowering overall risk frequency.

Taken together, the impact amplification is concentrated but sharp, while likelihood elevation is broad and operationally distributed. In effect, the Platform's residual-risk profile is defined less by missing safeguards than by the variable strength of existing ones.

5.3 Risk Mitigation

Mitigation Measures Register

In previous iteration, the risk assessment relied exclusively on the Risk Register, which mapped mitigation measures to individual risk scenarios. However, this structure did not enable effective tracking of each measure's effectiveness, proportionality, and reasonableness on a standalone basis. To address this limitation, the current iteration introduced a dedicated Mitigation Measures Register, which consolidates all measures and presents the results of a measure-centric assessment. The Mitigation Measures Register includes the following elements:

- Structured list of mitigation measures in place,
- General information about listed measures detailing Measure ID, Category, Name, Description, Type and Owner,
- Mitigation measure assessment outcomes, providing ratings of reasonableness, proportionality and Effectiveness.

The Mitigation Measures Register was developed using information from previous risk assessments and additional input provided by the respective Platform teams during the current assessment iteration. As a first step, a consolidated list of measures from the earlier assessment was compiled, ensuring consistent terminology and eliminating any duplications. The relevance of each measure was subsequently verified against the recommended

safeguards outlined in Article 35 of the DSA, information presented in the EU preliminary findings of the study on systemic risks and their mitigation⁴², and Guidelines on measures to ensure a high level of privacy, safety and security for minors online⁴³.

The collected information was then reviewed and discussed during cross-functional workshops (further described in the following section). These workshops served to validate the final list of measures, refine their specific descriptions and assess their adequacy and effectiveness, forming the foundation of the Mitigation Measures Register.

Mitigation Measure Assessment Methodology

Mitigation measure assessment focuses on evaluation of measures in place. Each measure was reviewed and rated across three assessment dimensions: reasonableness, proportionality and effectiveness. The criteria of reasonableness and proportionality served as an indicator of whether the measure is adequately implemented and the results do not enter the next phase of assessment – residual risk assessment. The criterion of effectiveness served for evaluation of the effectiveness of individual measures implemented. The results were then used as input for the residual risk calculation. Section 5.4 describes how the results of the effectiveness assessment were subsequently handled.

The assessment criteria were based on a combination of information presented in the EU preliminary findings of the study on systemic risks and their mitigation⁴⁴ and an established qualitative assessment approach. For each criterion, a qualitative rating scale (High, Moderate, Low) was applied, supported by predefined guidance criteria. The following section outlines the assessment criteria applied during the evaluation, defining the underlying principles used to determine each rating.

Reasonableness and Proportionality

The descriptions provided in the tables below represent general guidance for assessing the levels of Reasonableness (see Figure 7) and Proportionality (see Figure 8). They serve as a reference framework to ensure consistent interpretation of evaluation criteria across measures.

*Figure 7: Reasonableness assessment levels
(NKL's internal document – Mitigation Measures Register and Dashboard)*

Reasonableness Level	Description
High	The measure is applied promptly and systematically, without unnecessary delays. Measure application relies on established internal rules, expert knowledge, and/or predefined processes rather than ad hoc judgment/random execution.
Moderate	The measure is generally applied in a timely manner, but with occasional delays or inconsistencies. Some steps or decisions rely on informal practices or limited documentation. Existing knowledge and rules are used, though not always comprehensively.
Low	The measure is applied irregularly or with significant delays. There are no clear procedures or consistent use of existing knowledge. Decisions depend mainly on individual discretion, leading to unpredictable outcomes.

⁴² European Commission. (2025, April 29). Digital Services Act – Study on systemic risks and their mitigation: Second workshop presentation. Directorate-General for Communications Networks, Content and Technology

⁴³ Guidelines on measures to ensure a high level of privacy, safety and security for minors online, pursuant to Article 28(4) of Regulation (EU) 2022/2065.

⁴⁴ European Commission. (2025, April 29). Digital Services Act – Study on systemic risks and their mitigation: Second workshop presentation. Directorate-General for Communications Networks, Content and Technology

Figure 8: Proportionality assessment levels
(NKL's internal document – Mitigation Measures Register and Dashboard)

Proportionality Level	Description
High	The measure effectively protects users and platform integrity while fully respecting fundamental rights such as privacy and freedom of expression. Implementation is scaled appropriately to the platform's size and operational capacity.
Moderate	The measure contributes to user safety but may introduce minor unnecessary restrictions or gaps in rights protection. Resource allocation is adequate but not fully proportionate to the platform's scale.
Low	The measure may unduly restrict user rights or lack necessary resources for consistent application. Implementation is misaligned with the platform's size or operational needs.

For each individual measure, 2 specific assessment questions for reasonableness and 2 specific assessment questions for proportionality were developed based on above predefined levels and applied, tailored to the measure's purpose, scope, and operational context. Each question could be answered with Yes (2 points), Partly (1 point), or No (0 points). The total score was then used to determine the final rating according to the following scale:

- **4 points:** High Reasonableness (Proportionality)
- **2–3 points:** Moderate Reasonableness (Proportionality)
- **0–1 points:** Low Reasonableness (Proportionality)

This structured scoring ensured a consistent and transparent evaluation across all assessed measures.

Effectiveness

The effectiveness assessment was conducted qualitatively based on expert judgment provided by the respective measure owners. These individuals or teams are directly responsible for the practical implementation of the measures and have detailed knowledge of their operational performance and limitations. The evaluation followed the predefined effectiveness levels in the table below (Figure 9). Where necessary, external studies were also considered to support and validate the evaluation.

Figure 9: Effectiveness assessment levels
(NKL's internal document – Mitigation Measures Register and Dashboard)

Effectiveness Level	Description
High	With defined KPIs: ≥ 80% of targets achieved or strong positive trend. Without defined KPIs: qualitative assessment shows the measure is robust, well-integrated, and reliably operates.
Moderate	With defined KPIs: 50–79% of targets achieved or mixed results. Without defined KPIs: qualitative assessment shows the measure works in most cases but impact is limited (e.g. by scope, accuracy, user engagement, and/or incomplete process coverage).
Low	With defined KPIs: < 50% of targets achieved or consistently poor results. Without defined KPIs: qualitative assessment shows narrow or unreliable impact (e.g., easily bypassed, dependent on external factors, partial/dysfunctional implementation).

The best practices recommend quantitative assessment using predefined KPIs. However, at the time of the risk assessment, such assessment was possible only for the measures of automated tool Vercury, Advertisements review before publication, Advertisements review after publication, and Advertisers Certification Program, where KPIs were

established. Therefore, the predefined effectiveness levels (see Figure 9) provide criteria for the quantitative assessment using KPIs.

During the workshops, possible KPI options were proposed and discussed with the respective teams. Where available, indicative or estimated performance data were taken into account, though such estimates were incorporated only within the qualitative part of the assessment. NKL acknowledges this methodological limitation and addresses it in the Action Plan (see Section 6.3), which includes a specific action point aimed at defining measurable KPIs for each relevant measure.

Mitigation Measure Assessment Approach

The assessment of mitigation measures was conducted to evaluate the adequacy and effectiveness of the Platform's existing controls in addressing inherent systemic risks. The evaluation process combined two complementary approaches:

- Desk research and data review, and
- Cross-functional workshops.

The desk research and data review involved analysing internal documentation and materials, transparency reporting data, information gathered during the workshops, and complementary external studies.

The cross-functional workshops were designed to verify the relevance of identified measures, their operational implementation and the assessment of reasonableness, proportionality and effectiveness. Workshops were held with key teams, including:

- Content Moderation and Advanced Review Team,
- Notice and Complaint Team,
- Tech Team,
- Support Team,
- Advertising Team,
- Legal Team.

The teams discussed the measures within their respective operational areas. First, each team, guided by the Compliance Team as the workshop facilitator, reviewed the operational relevance of the identified measures. Second, the adequacy (reasonableness, proportionality) of each measure was assessed through targeted evaluation questions tailored to each measure, with responses provided based on the teams' practical experience and expert judgment. Finally, each team evaluated the effectiveness of the measures as well as discussed existing and proposed potential KPIs to support future quantitative assessment of effectiveness.

The Compliance Team then consolidated all findings and integrated the results into a comprehensive Mitigation Measures Register, which now serves as a central repository of all controls and their evaluations.

Mitigation Measure Assessment Outcomes

Reasonableness Assessment

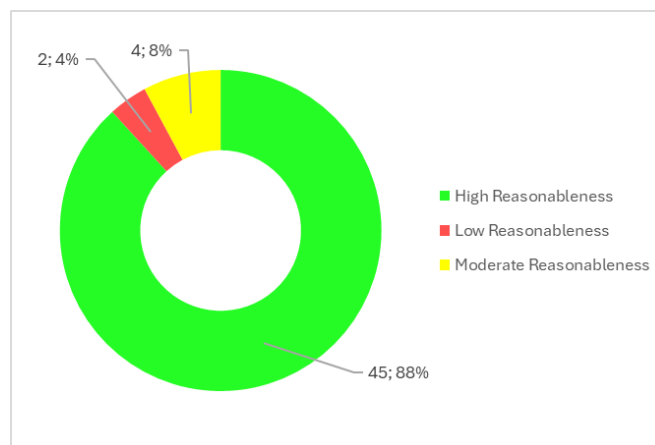
The results of the assessment are visualized in Figure 10 and indicate that the majority of mitigation measures demonstrate **high reasonableness** (45 out of 51). These measures are applied promptly and systematically, without unnecessary delays, their application relies on established internal rules, expert knowledge, or predefined processes rather than random execution.

Moderate reasonableness was identified for 4 measures. The procedures are applied systematically, however lacks uniformity across all processes. For instance, regular code reviews to avoid security issues only partially follow predefined security guidelines or checklists. However, they are conducted systematically before deployment – each update undergoes a review to ensure stability and security before it goes live.

Low reasonableness was identified for 2 measures. The procedures are only partially applied or voluntary for users, resulting in limited consistency.

Overall, the results show that the Platform’s risk-management framework is well-structured and that most measures are implemented in a manner consistent with regulatory expectations and operational feasibility. Relevant action points have been included in the Action Plan (see Section 6.3) for those measures where further adjustments would meaningfully strengthen the Platform’s overall risk management, taking into account the overall risk assessment results and the Platform’s operational capabilities.

Figure 10: Reasonableness Assessment Outcomes
(NKL’s internal document – Mitigation Measures Register and Dashboard)



Proportionality Assessment

Similarly, the outcomes of the assessment, visualized in Figure 11, show that the majority of mitigation measures are **highly proportionate** (43 out of 51). These measures effectively protect users and Platform integrity while fully respecting fundamental rights such as privacy and freedom of expression. Moreover, their implementation is scaled appropriately to the Platform’s size and operational capacity.

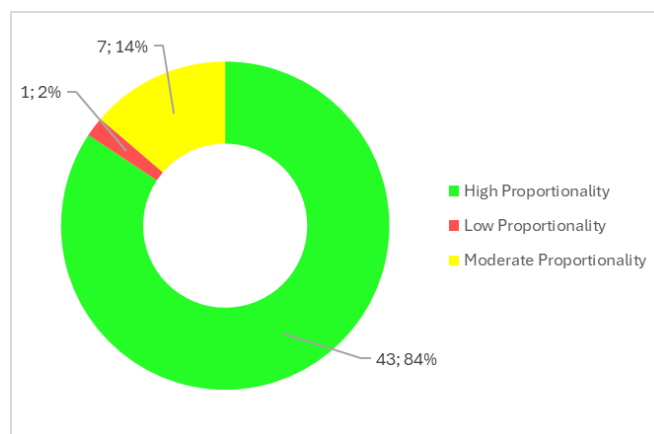
Moderate proportionality was recorded for 7 measures:

- MD5 hash allows to block uploads without manual review, which, however, applies only to content meeting strict evaluation criteria. On the other hand, the hashing infrastructure is integrated into the upload pipeline and scaled to handle the Platform’s volume of incoming files.
- Page blurring enables to remove the blur with a single confirmation click, allowing adults immediate access after acknowledging age and ToS requirements, and functions reliably across all modern devices and browsers.
- Same limitation applies to Cookie banner function.
- However, the cookie banner allows to users who reject non-essential cookies to access the site and view content, including videos. Thus, the Platform maintains its basic functionality for all users.
- Further, the system of firewalls and access restrictions is designed to target harmful connections, however, occasional false positives occur. On the other hand, the reliability is presumed high.

Low proportionality was recorded for 1 measure, specifically, for XNXX uploaders verification process.

Overall, the results indicate that the Platform’s mitigation framework is well balanced, ensuring that implemented measures manage identified risks while remaining proportionate to the Platform’s nature and operational capacity. Relevant action points have been included in the Action Plan (see Section 6.3) for those measures where further adjustments would meaningfully strengthen the Platform’s overall risk management, taking into account the overall risk assessment results and the Platform’s operational capabilities.

Figure 11: Proportionality Assessment Outcomes
(NKL’s internal document – Mitigation Measures Register and Dashboard)



Effectiveness Assessment

The outcomes of the assessment are visualized in Figure 12, showing that more than half of the mitigation measures (26 out of 51) are **highly effective**. These include, for example, automated tools (Vercury, Hive), data protection measures (e.g., Multiple access checks, Strong password internal policy, Firewalls and access restrictions), content moderation practices (Manual review of uploaded and reported videos, Cooperation with law enforcement agencies, Translator tools and Multilingual content moderation team, Content moderation procedures, Repeat infringers policy, Notice mechanism via Abuse Reporting Form and via Copyright Infringement Reporting Form, Complaint-handling system), or Advertising (e.g. Advertisements review before and after publication). Many of these measures are technically robust, consistently implemented, and supported by automated enforcement or structured human oversight, they reliably achieve their intended outcomes, demonstrate strong performance and/or coverage.

Moderate effectiveness was recorded for 17 measures. These include, for example, automated tools (Google SafetyNet API, Keyword search and match), protection of minors measures (Parental controls instructions available for visitors/users, Warnings about adult content, User confirmation of Terms of Service before accessing content, Page blurring), data protection measures (e.g., User in full control of tracking and data collection, Regular code reviews to avoid security issues, Cookie Banner), and policy-related instruments (ToS, Cookie and Privacy Policies). Many of these measures demonstrate consistent implementation and contribute meaningfully to risk reduction, however, their effectiveness is moderated by partial coverage, technical or contextual limitations (e.g. user engagement, circumvention potential, incomplete feedback loops, or restricted scope).

Low effectiveness was recorded for 8 measures. These measures include, for example, automated tools (MD5 hash, Safer), protection of minors’ measures (Self-validation, RTA label), or user verification procedures for XNXX uploaders and regular users. The measures provide only limited mitigation capability, as their functionality is either technically constrained, dependent on external factors or user behaviour, or not fully operational.

The effectiveness of individual measures in each category is shown in Figure 13. Overall, the results indicate that the Platform has established a solid and multi-layered system of controls capable of effectively addressing most identified risks.

The less effective measures remain functional but offer reduced assurance. Relevant action points have been included in the **Action Plan** (see Section 6.3) for those measures where further adjustments would meaningfully strengthen the Platform’s overall risk management, taking into account the overall risk assessment results and the Platform’s operational capabilities.

Figure 12: Effectiveness Assessment Outcomes
(NKL’s internal document – Mitigation Measures Register and Dashboard)

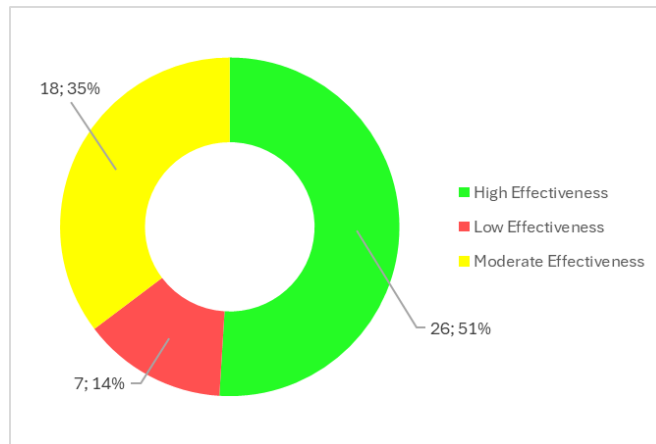
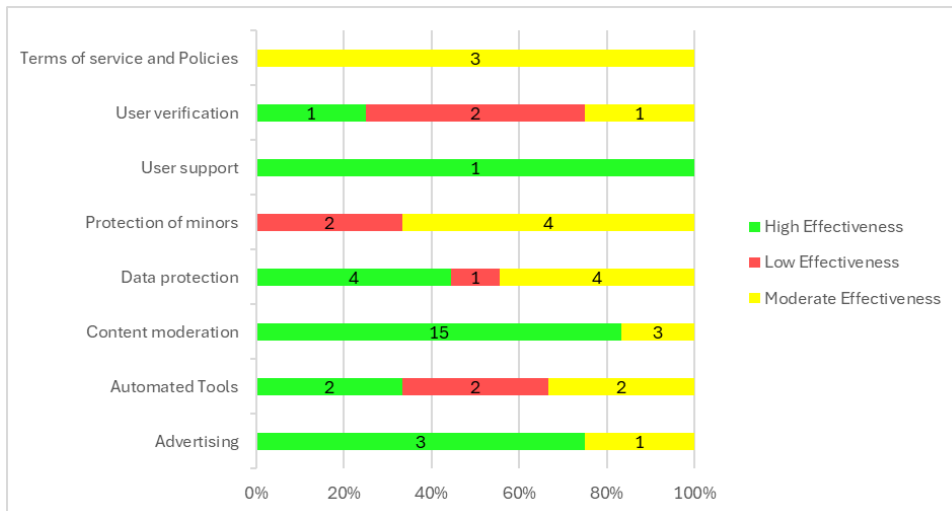


Figure 13: Effectiveness Assessment Outcomes per Measure Category
(NKL’s internal document – Mitigation Measures Register and Dashboard)



Mitigation Measures in Place

Automated Tools (M-AT)

The Automated tools category comprises a set of systems designed to detect, flag, or block illegal or policy-violating content through automated analysis and comparison techniques (MD5 hash, Vercury, Hive, Google SafetyNet API, Safer, Keyword search and match).

The MD5 hash tool generates a fixed digital fingerprint for each file, allowing precise detection of duplicate or known illegal materials, addressing risks related to dissemination of illegal content, gender-based violence, and physical and mental well-being.

Vercury applies internal signature matching to compare video or image content against a proprietary database, assigning a match score that helps identify potential violations for review. Hive leverages artificial intelligence to analyse visual content and detect behaviours or objects linked to prohibited categories such as violence or firearms. The Google SafetyNet API uses similar AI-based analysis to evaluate images and assign a risk score, particularly focused on detecting underage or abusive content. Safer complements these tools by maintaining a database of verified content fingerprints to efficiently identify known harmful materials. Additionally, the keyword search and match system monitors text-based inputs through dynamic blacklists and grey lists, automatically blocking or flagging specific words and phrases to ensure continuous control over potentially harmful or illegal textual content. These measures provide a combined mitigation for wide range of risks related to, for instance, dissemination of illegal content, human dignity, gender-based violence, privacy and family life, non-discrimination, protection of minors, public security concerns, public health and physical and mental well-being.

Content Moderation (M-CM)

The Content Moderation category encompasses measures ensuring that all user-generated materials are reviewed, assessed, and managed responsibly and consistently. Measures in this category address a wide range of risks related to Dissemination of Illegal Content, Human Dignity, Gender-Based Violence, Privacy & Family Life, Freedom of Expression and Information, Non-Discrimination, Protection of Minors, Consumer Protection, Civic and Electoral Impact, Public Security Concerns, Public Health, and Physical and Mental Well-being.

Manual review of uploaded videos

The manual review of uploaded videos is carried out by a dedicated moderation team that examines content flagged by automated systems. Moderators assess user age, participant consent, legality, and copyright compliance, taking appropriate actions such as validation, ghosting, takedown, or deletion. All moderation activities follow standardized content moderation procedures defining review and decision-making steps to maintain legal compliance.

Moderation team training

To ensure quality and consistency, moderation team training provides comprehensive onboarding and continuous training. Content moderators undergo trainings. Training is conducted by the team leader and experienced moderators, ensuring new recruits and existing team members are guided by seasoned professionals. Newly hired moderators undergo a mentorship-based onboarding process, observing experienced colleagues before gradually taking on tasks under supervision. Training emphasizes the responsible handling of sensitive content, the identification and addressing of illegal material, and a strong familiarity with moderation processes and classification methods.

Translator tools & Multilingual content moderation team

To address linguistic diversity, translator tools support moderators in understanding non-supported languages, while multilingual moderation teams ensure coverage of multiple EU languages. The Platform uses integrated translation tool for languages not directly supported within the group. For more complex or nuanced cases, moderators verify translations using contextual checks, cross-referencing with online resources, or consulting colleagues who are native speakers or proficient in English. At minimum, all moderators are required to demonstrate advanced proficiency in English.

Cooperation with law enforcement agencies

The Platform maintains cooperation with law enforcement agencies. The Platform applies an internal classification system to identify content that is undoubtedly illegal. When a reviewer identifies and flags such content, the system automatically saves the related data. This database is made available to law enforcement agencies. The Platform also cooperates with the official Czech reporting portal stoponline.cz operated by the CZ:NIC, through which illegal content outside of XNXX websites and where the review team has no authority or means to deal with is reported via an established reporting form.

Notice mechanisms, Repeat infringers policy & Manual review of reported videos

The Platform also maintains a comprehensive notice system that enables users, trusted flaggers, and authorities to report illegal or Terms of Service–violating material through multiple dedicated channels.

Users and third parties can report illegal or ToS violating content via the Report button next to each comment. Upon submission of a comment report, the reporter receives an automatic email acknowledgment confirming that the report has been successfully submitted. Similar system is deployed for reporting pictures as well. Users and third parties can report illegal or ToS violating pictures via the Takedown amateur form and the Privacy take down form in the section Content removal on the button of the webpage. Upon submission of a report, the reporter receives an automatic email acknowledgment confirming that the report has been successfully submitted.

Illegal or ToS violating videos can be reported via the Abuse Reporting Form in the Content Removal section. The following is submitted via the form: exact electronic location of the content, specific category of the reported content, explanation of why it is considered illegal or in violation of the ToS, contact details of the reporting party, and a statement confirming the reporting party's bona fide belief that the information and allegations provided are accurate and complete. Upon submission, reporters receive an on-screen and email confirmation including a link to ticketing system and later a reasoned decision including outcome, rationale, and redress options. Each video also includes a Report button enabling immediate flagging of suspected illegal or harmful material. After submission, the reporter receives an automatic email confirmation acknowledging receipt, with a link to the reported video. A subsequent email informs them of the decision and reasoning, with access to appeal mechanisms.

Users and third parties can also report copyright-related video issues via the Copyright Infringement Reporting Form. Within the form, the following data is requested: URL of the content and contextual information describing copyrighted content being infringed. Upon submission, reporters receive an on-screen and email confirmation and later a reasoned decision including outcome, rationale, and redress options. Users who repeatedly upload content violating copyright receive strikes after prior warnings. Accounts are terminated after three strikes, unless a strike is proven to be a mistake.

Reported videos and other content are reviewed by the Notice and Complaint Team and Advanced Review Team. The team assesses the notice, may request clarifications from reporter or uploader (including copyright documentation), and issues a content decision (validation, ghosting, takedown, deletion) or account terminations.

Trusted Flaggers

The Platform also allows registration of certified Trusted Flaggers. Trusted Flaggers can register through a dedicated "Trusted Flaggers – Registration" page. The registration system is available only for organizations that are established

in the EU, officially designated as “trusted flaggers” by the competent Digital Services Coordinator, and listed in the public database maintained by the European Commission. The following information is requested within the registration form: the name of the organization, address of the registered office, country of establishment, and the official email address listed in the European Commission’s Trusted Flaggers register. After completing the registration, the organization receives both an on-screen confirmation and an email confirmation acknowledging receipt of the application. Once the registration request of a Trusted Flagger has been verified and approved, the applicant receives a notification email confirming that their account has been activated. The email provides a link to create a password for the account. Trusted Flaggers’ notices are then prioritised and reviewed with urgency. They are flagged and automatically placed at the top of the moderation queue.

Points of contact for Member States’ authorities

The Platform has established a contact point for the receipt and processing of official removal orders concerning terrorist content issued by the competent state authorities. The contact point is accessible via a link in the ToS and dedicated Platform page (Contact Form - Authorities). The contact point may be addressed in English or Czech.

Complaint-handling system

Users can further challenge moderation outcomes via the complaint-handling system, ensuring transparent appeals and corrective actions. Complaint-handling system allows recipients to appeal moderation decisions. Complaints can be submitted either through a ticketing system accessible in the user dashboard or via email using a link provided in the moderation notice. The ticketing system ensures all communications are logged, tracked, and resolved in a transparent and organised manner. Each appeal case is reviewed by a dedicated team, which assesses the justification of the moderation action. The review follows internal guidelines. Where an error is identified, corrective measures are taken, such as restoring content or reversing an account suspension. Feedback gathered through appeals is also used to refine moderation systems.

User Verification (M-UV)

The User verification category defines the procedures applied to verify the users contributing content to the Platform as well as regular users.

Uploader verification

Users who upload videos via XNXX are not required to verify themselves. However, for uploading tagged content, uploader verification is required. This verification can be completed through alternative upload channels and may include several procedures depending on the chosen route. The verification process may require submission of an official ID document and content samples demonstrating that the uploader is the same person appearing in the video. Verification checks confirm the uploader’s name, date of birth (age verification), and consent and ownership of the content. In other cases, verification involves a profile picture and a video upload, confirming the visual content matching (authorship).

Regular user verification

For regular users, those who watch or comment on videos, verification is limited to basic email validation upon account registration – which is, however, not required -, confirming authenticity while maintaining accessibility for non-uploading users.

The process of account creation includes, inter alia, age declaration, agreement with ToS, which clearly demonstrates that users under 18 or the age of most of their jurisdiction are prohibited from accessing the website (Article 2). Furthermore, the process as well as ToS are available in all EU languages, ensuring easy navigation.

User Support (M-US)

The User support category ensures that users have accessible and reliable channels for communication with the Platform, mitigating risk related to consumer protection (SR-CP-05). The contact points for users are established and accessible through the “Contact Us” form available via a dedicated webpage, and directly from the user dashboard through Discussion. These channels allow users to submit general inquiries, raise concerns, or request assistance related to the use of the service.

Protection of Minors (M-PM)

The Protection of minors category includes measures designed to prevent underage users from accessing adult material and to support parents and guardians in applying appropriate safeguards. These measures specifically address risks from the category Protection of Minors.

Warnings about adult content & Page blurring

The Platform applies a default blurring effect to its front page, ensuring that explicit content is not immediately visible upon entry. Users must take an additional action (confirm age and agree to the terms and conditions, see described below) to access the site and view the content in full. Moreover, upon a user’s initial visit, the Platform displays a clear warning that its content is restricted to adults.

Self-validation & User confirmation of Terms of Service before accessing content

Upon a user’s initial visit, users are presented with a clear message stating that by entering the website, they confirm their age. The Platform displays a requirement to confirm that the user is above the legal age threshold. Similarly, the Platform requires user confirmation of the Terms of Service before entry, ensuring informed consent. Until the user explicitly confirms their consent, access to the website remains blocked and the underlying content is blurred.

Consistent with the DSA Article 28 Guidelines that self-validation does not meet the age assurance requirements, the effectiveness of this measure is rated as low. NKL is aware that for access to pornographic content is recommended the age verification method. However, NKL holds the view that available age-verification approaches present deficits in accuracy, privacy, proportionality, and security. Expert bodies note that no current website-level method reliably balances effectiveness with data-minimisation.⁴⁵ In practice, such measures tend to displace users to less regulated sources, undermining child-safety objectives and weakening overall trust-and-safety outcomes. Therefore, the Action Plan includes action points related to age verification tools and protection of minors, prescribing the continuous cooperation in this field and monitoring of latest age verification developments. See measures Collaboration with NGOs and experts on online protection of minors, and Collaboration with the regulators in Section 6.3. Similar initiatives are continuous for NKL as has been reflected in the previous iteration of the risk assessment. For more detail on current status of these activities, please see Section 6.2.

Parental controls instructions available for visitors/users & RTA label

Information about practical tools aimed at helping guardians prevent minors from accessing adult online material are available via dedicated page of the Platform. Users are also provided dedicated link when initially entering the page, before the content is displayed. Instructions inform about activating technical restrictions on both devices and networks used by children. On the browser level, users are provided links on filtering modes in common search engines, such as Google SafeSearch, Microsoft Bing search settings as well as Yahoo! SafeSearch, to prevent adult site such as XNXX to appear in the search results. Recommend is also the use of a safe visual search engine tailored

⁴⁵ EDRI. (2024, September). Joint Statement on the Dangers of Age Verification Proposals to Fundamental Rights Online. <https://edri.org/wp-content/uploads/2024/09/Joint-Statement-on-Dangers-of-Age-Verification.pdf>; Australian Government, (2023, August). Government response to the Roadmap for Age Verification. <https://www.infrastructure.gov.au/sites/default/files/documents/government-response-to-theroadmap-for-age-verification-august2023.pdf>; CNIL. (2022, September). Online age verification: balancing privacy and the protection of minors, <https://www.cnil.fr/en/online-age-verification-balancing-privacy-and-protection-minors>.

for kids (Kiddle). On the device label, the instructions also inform about content-control options available within standard operating systems (Windows 10, Android, Apple). On the network-level, information regarding the activation of network-level content filters is available. Such measures are not deployed on platform level, therefore the scope of options for adults is limited, given the requirements listed in the DSA Article 28 Guidelines. However, the measures are easy to use and access, applicable regardless of the operating systems as well as proportionately restrict minor's access platform, given the minimum age for access. Additionally, these measures are deployed in combination with other measures preventing minor's access to the Platform, as suggested in the guidelines.

Additionally, the RTA label ("Restricted to Adults") is embedded across all pages, signalling to browsers and filtering systems that the content is intended solely for adult audiences.

Data Protection (M-DP)

The Data protection category covers measures designed to ensure the security, integrity, and lawful handling of user information.

User in full control of tracking and data collection

Users retain full control over the collection and use of their data. Before accessing the homepage for the first time, users are prompted to set their cookie preferences. These choices can later be modified at any time through a dedicated link available at the bottom of the website. In addition, users are provided with the option to enable or disable their viewing history. By deactivating this feature, users can prevent their past activity from influencing future content recommendations.

Transparency is reinforced also by a cookie banner. Before interacting with any content on the Platform, users are presented with this cookie banner. The banner provides three options: accept all cookies, manage cookie preferences, or reject all cookies. It also contains a short explanation of the use and management of cookies, along with a direct link to the full Cookie Policy.

Thus, risks related to privacy and family life, data privacy and protection, consumer protection and civic and electoral impact are managed.

Regular code reviews to avoid security issues & HackerOne program

Technical safeguards include regular code reviews. Reviews of Platform's codebase are conducted to identify vulnerabilities or errors that may lead to security risks. Code reviews are performed internally, primarily before new features or updates are deployed. The process focuses on detecting bugs that may have security implications, even if the code appears to function normally.

Moreover, under the HackerOne program, independent ethical hackers are encouraged to identify vulnerabilities or weaknesses in the Platform's systems, such as unauthorized data access. The program operates entirely externally, without granting hackers access to internal systems.

Firewalls and access restrictions, Multiple access checks, Strong password internal policy & Alerts for low password complexity

To protect infrastructure, firewalls and access control mechanisms are deployed to protect Platform's network and systems against unauthorised connections, malicious traffic, and external threats. Also, a layered access control mechanisms is implemented to ensure that only authorised users can reach sensitive systems, data, or functionalities. Access protection includes multiple authentication layers: (i) VPN connection to prevent external unauthorized network access, (ii) Web server password securing entry to server-level environments, (ii) for some applications a Application-level password /two-factor authentication/ for an additional security layer. Access monitoring then logs both successful and failed login attempts. In the event of unusual or failed access attempts, notifications are sent automatically via email, enabling administrators to verify or report unauthorized activities.

A strong password internal policy mandates the use of complex credentials, while real-time alerts for low password complexity prompt users to strengthen weak passwords. Warnings are displayed in real time during the password creation process, encouraging users to select stronger credentials that meet defined security standards.

Data retention practices (IPs)

Additionally, data retention practices limit the storage of IP address data. Retention practices set that IP address data is stored only for a limited period. Automatic deletion processes run weekly to remove old records. Through these measures, risk related to privacy and family life and data privacy and protection are managed.

Terms of Service and Policies (M-LE)

The Terms of Service and Policies category establishes the legal and procedural framework governing the Platform's operations, user conduct, and data management.

Terms of Service

The ToS define the fundamental rules for accessing and using the Platform, outlining conditions for account creation, content submission, and user behaviour. They explicitly prohibit illegal, harmful, non-consensual, or terrorist material, set age restrictions, and include references to parental control tools. The ToS also address intellectual property rights, sponsorship transparency, and dispute resolution procedures. ToS are accessible via a dedicated link when accessing the platform as well as via a dedicated platform subdomain, translated into all EU languages, and accompanied by a user-friendly summary to support understanding.

Thus are the ToS addressing risks related to dissemination of illegal content, human dignity, gender-based violence, privacy and family life, freedom of expression and information, non-discrimination, protection of minors, consumer protection, civic and electoral impact, public security concerns, public health and physical and mental well-being.

Cookie Policy

Cookie Policy explains the use of cookies to ensure proper website functionality, enhance user experience, and support security and performance. The Policy distinguishes between essential cookies, which are necessary for site operation and security, and non-essential cookies, which require user consent and are used for personalization, advertising, and improving services. Users are provided with transparent information on the purposes, duration, and categories of cookies, and are informed about the management or withdrawal of their consent. The Policy also clarifies that data collected is not shared with unrelated third parties and is not transferred outside the EEA without adequate safeguards.

Privacy Policy

The Privacy Policy describes how the Platform collects, uses, processes, and discloses user information, including personal information, in conjunction with user's access to and use of the Website. The Platform processes personal data only in accordance with this Privacy Policy and the relevant legislation, in particular GDPR. By accessing the website, the user acknowledges that they have read the Privacy Policy.

These policies address risks related to data privacy and protection, consumer protection and civic and electoral impact.

Advertising (M-AD)

The Advertising category encompasses measures ensuring transparency, legality, and compliance in the display and management of advertisements across the Platform, directly addressing risks related to misleading promotions, discriminatory advertising, dissemination of illegal content, gender-based violence content as well as risk related to civic and electoral impact and public security concerns.

Clear Visual Marking of Advertisements

All ads are clearly identifiable through visual marking, where each advertisement features an “i” icon leading to an “About This Ad” section that discloses the advertiser’s identity and targeting parameters, promoting transparency for users.

Advertisements review before and after publication & Advertisers Certification Program

All advertisements are manually reviewed prior to approval and publication. Each element is checked against internal rules. Non-compliant ads are declined. A wide range of tools are used, including antivirus scanners, URL analysers, VPNs, and translation utilities. Internal rules define accepted and labelled (prescription drugs if legal) and not accepted ad formats and types (e.g. non-consensual content, violence, incest, scatophilia, malware, phishing, ransomware, prostitution, racism, political ads). For ads targeting the EU, UK or California (US), additional checks are performed to confirm that they do not drop cookies, ensuring compliance with data protection requirements. Submitted ads are placed in a pending status and enter a review queue processed in chronological order.

Upon advertisements' approval, they are subject to continuous scanning to detect policy violations. Detected violations trigger investigation. A predefined tiered violation process is in place to handle detected issues. Minor violations trigger labelling, major violations lead to campaign declines and repeat violators may be suspended, severe violations result in immediate bans and permanent account closure.

Moreover, the Advertisers Certification Program ensures that only qualified and compliant advertisers gain access to specific advertising zones. To become certified, advertisers must meet predefined eligibility criteria. Applicants are subject to a formal assessment including assessment of related risks and verification of adherence to advertising rules. Certified advertisers are reviewed periodically to confirm continued compliance. Any detected violations may result in removal of certification, suspension of campaigns, or termination of the advertising account.

5.4 Residual Risk Assessment

Residual Risk Assessment Methodology

The residual risk assessment evaluates the level of risk that remains after the implementation of mitigation measures and existing internal controls. The assessment methodology was updated from the prior assessment cycle to more transparently reflect Articles 34(2) and Article 35 DSA. In this iteration, residual risk is derived through an explicit two-step, quantitative process that makes clear how control effectiveness reduces risk and how drivers may still shape outcomes.

Whereas the previous iteration primarily relied on qualitative reasoning and control descriptions, the revised approach introduces a semi-quantitative reduction model that directly links mitigation performance to the numerical risk outcome. This adjustment ensures that the reduction of inherent risk is not only narratively explained but also quantitatively evidenced in the residual score.

Step 1: Adjustment for Control Effectiveness

Each inherent risk score was quantitatively adjusted to reflect the mitigating effect of existing controls. This adjustment was made through the application of a **Risk Reduction Factor (RRF)**, which expresses the average operational effectiveness of all safeguards linked to the given risk. Effectiveness ratings were assigned using a three-level scale, reflecting control maturity and operational reliability:

Effectiveness Rating	RRF
High Effectiveness	0,8
Moderate Effectiveness	0,5
Low Effectiveness	0,2

The reduction is calculated using the following formula:

$$\text{Residual Risk Score (pre-driver)} = \text{Inherent Risk Score} \times (1 - \text{RRF})$$

This step quantifies the extent to which risk is reduced through control performance rather than through qualitative description alone.

Step 2: Adjustment for Risk Driver Influence

Once the base residual value is established, the model incorporates the influence of risk drivers that can still elevate risk levels despite effective mitigations. These drivers correspond to the influence factors defined under Article 34(2) DSA. Each driver was analysed to determine whether it primarily affects the likelihood (increasing probability of occurrence) or the impact (amplifying severity of harm) of the related systemic risk. The strength of its residual influence was expressed through a **Driver Impact Score (DIS)**, which functions as an additive value to the pre-driver residual score.

Influence Level	DIS
Low Influence	0
Medium Influence	0,5
High Influence	1

Where multiple drivers affected the same systemic risk, their modifiers were aggregated to produce a single DIS, reflecting the total cumulative influence.

Step 3: Adjustment for Risk Driver Influence

The final residual risk value integrates both the mitigating effect of controls and the amplifying effect of risk drivers:

$$\text{Final Residual Risk Score} = (\text{Inherent Risk Score} \times (1 - \text{RRF})) + \text{DIS}$$

This combined formula ensures that reductions in exposure are quantified transparently through control effectiveness, while the persistent influence of systemic drivers is explicitly represented. The resulting residual scores are directly comparable to inherent risk scores.

Residual risks were classified using the same three-band scale as the inherent risk assessment, maintaining continuity and comparability.

Residual Risk Assessment Approach

The residual risk assessment was conducted by the Compliance Team using a structured and formula-based approach to ensure methodological transparency. Residual scores were not assigned manually or based on qualitative judgment; instead, they were calculated systematically in accordance with the defined quantitative model. The calculation was applied to each systemic risk entry to reflect the combined effect of implemented control effectiveness and residual driver influence. All resulting values were validated and formally recorded in the Risk Register. This approach guarantees that residual exposure is represented through evidence-based metrics rather than subjective estimation.

Notable methodological points:

- For certain scenarios, lower baseline RRF values (e.g., 0.5 or 0.65), were manually adjusted upward to 0.8. Such adjustments were made only when there was clear, documented justification showing that the measures were operationally effective. Each adjustment was accompanied by a rationale grounded in evidence of sustained control performance and alignment with regulatory or ethical standards.

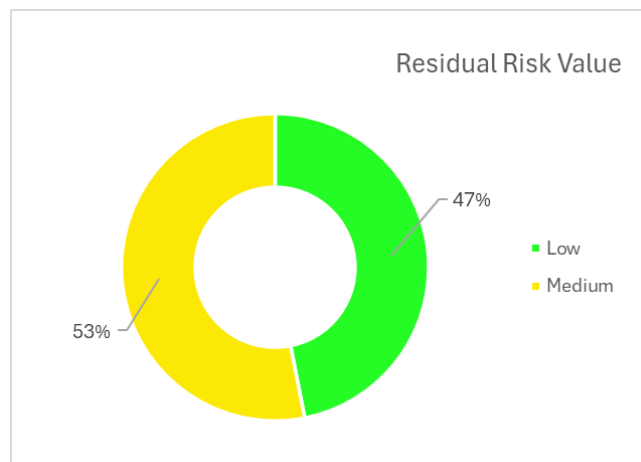
- Certain safeguards (parental controls, warnings, page blurring) were intentionally excluded from RRF calculations in scenarios assessing risks to minors who have already accessed the Platform. This methodological decision reflects a conservative assessment stance, acknowledging that such measures primarily function as preventive barriers rather than active mitigations once access has occurred.

Residual Risk Assessment Outcomes

Overall Posture

The residual risk landscape demonstrates a balanced and controlled post-mitigation profile. As illustrated in Figure 14 medium residual values account for 53% of all assessed risk scenarios, while low residual values account for the remaining 47%. The chart confirms that no high residual risks persisted after mitigation. This indicates that the implemented controls and preventive measures have been effective in reducing overall systemic exposure. However, the predominance of medium values highlights that some risk areas continue to be influenced by factors beyond full technical or procedural control.

Figure 14: Distribution of residual risk values across all categories
(source: NKL's internal document – Risk Register and Dashboard)



Change from Inherent to Residual Risk

The average residual risk score decreased from 14 (inherent) to 6 (residual) (see Figure 15). This reduction is driven primarily by a combination of automated detection tooling, structured human moderation, and clearer user-facing processes for reporting and redress.

Figure 15: Average residual risk scores by risk category
 (source: NKL's internal document – Risk Register and Dashboard)

Risk Category	Number of Risks	Residual Risk Score
Physical and Mental Well-being	4	10
Protection of Minors	11	9
Consumer Protection	6	8
Privacy & Family Life	5	7
Data Privacy and Protection Risks	5	7
Gender-Based Violence	8	7
Dissemination of Illegal Content	19	6
Human Dignity	2	6
Public Health	5	5
Public Security Concerns	4	5
Freedom of Expression and Information	3	4
Civic and Electoral Impact	5	4
Non-Discrimination	4	4
Total / Average Residual Risk Score	81	6

Privacy & Family Life experienced the most pronounced improvement, moving from an average inherent score of 22 to a residual score of 7. Scenarios such as non-consensual intimate imagery (including deepfakes) and doxing now sit at Medium residual levels. The shift is attributable to layered safeguards operating in concert: hashing and AI-assisted detection (to surface abusive material quickly), trained manual review (to confirm and contextualise), standardised takedown and notice mechanisms (to remove at scale and offer redress), cooperation with law-enforcement where appropriate, and user-verification plus Terms of Service enforcement (to deter repeat abuse and raise accountability).

In Data Privacy & Protection, the average risk declined from 20 to 7. The improvement reflects the maturation of privacy governance - GDPR-aligned policies and consent flows, access controls, security hardening, and disciplined retention practices. Where evidence indicated strong real-world effectiveness with limited impact on user engagement, the RRF was conservatively set or adjusted to 0.8 to reflect that controls meaningfully reduce likelihood and/or impact without introducing offsetting risks. Even after reductions, residual scores remain Medium in several scenarios because adversarial misuse of data and the scale of processing require continuous vigilance.

For Dissemination of Illegal Content, residual exposure decreased from 15 to 6 on average. This is the outcome of a defence-in-depth posture that blends specialized tools (e.g., Vercury, Hive, Google SafetyNet API, Safer), keyword matching and link-pattern detection, and a multilingual moderation capability able to evaluate context across languages. Trusted-flagger channels and law-enforcement cooperation further compress response times. While some vectors (e.g., CSAM attempts, extremist propaganda, or obfuscated phishing links) are persistent and adaptive, the layered approach consistently contains them to Low/Medium residual levels.

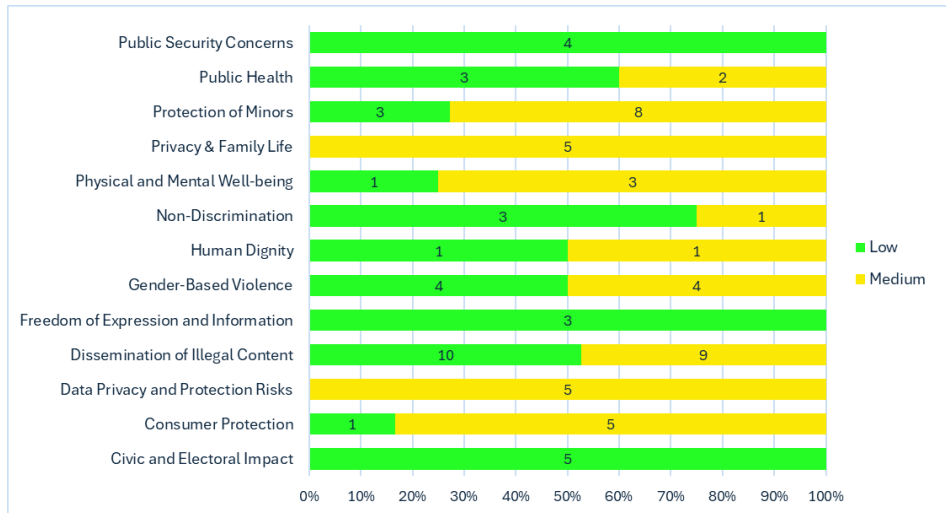
Finally, Human Dignity and Gender-Based Violence both show meaningful reductions (from 14 to 6, and 14 to 7, respectively). Here, clearer content standards, proactive detection, moderator training, and accessible reporting/complaint routes reduce the prevalence and persistence of degrading or violence-normalising material. That said, several scenarios remain at Medium because visibility dynamics and broader cultural normalisation can re-amplify harmful tropes even when individual items are removed. In practice, sustaining low residual risk in these areas depends on continuous tuning of detection criteria, consistent enforcement across languages and regions, and user-education that discourages harmful engagement patterns.

In sum, the controls have eliminated high residual ratings and concentrated remaining exposure in medium and low bands.

Residual Risk Assessment Outcomes by Category

The analysis of residual risks by category, as illustrated in Figure 5 provides a consolidated view of how the implemented control framework has reduced inherent risks to manageable levels across all systemic categories.

Figure 16: Residual risk distribution by category
(source: NKL's internal document – Risk Register and Dashboard)



The outcomes are as follows:

- Public Security Concerns (4 risks – 100% Low residual risk):** The use of automated detection tools, active cooperation with law enforcement agencies, and robust advertisement and account monitoring procedures ensure that risks relating to extremist or criminal exploitation are effectively contained. Residual exposure is negligible, and the category maintains a stable low risk level.
- Public Health (5 risks – 3 Low, 2 Medium residual risks):** Residual risk remains Medium where exposure to unrealistic body standards or risky sexual behaviours could affect users' perceptions and health attitudes. These risks are intrinsic to adult entertainment and user-generated media. Other vectors, such as the spread of misinformation or unregulated drug promotion, are largely mitigated by moderation procedures and content review protocols, resulting in mostly low residuals.
- Protection of Minors (11 risks – 3 Low, 8 Medium residual risks):** Despite multi-layered safeguards risks affecting minors persist at medium levels. The primary constraint is the limited effectiveness of age-assurance mechanisms and the reliance on indirect controls (e.g., parental controls, content warnings, and self-declaration). These measures help mitigate exposure but do not fully prevent minors from accessing or being exposed to adult material. Given the inherently high likelihood and impact, the residual exposure remains significant and requires ongoing monitoring of privacy-preserving age verification tools.
- Privacy and Family Life (5 risks – 100% Medium residual risk):** All risks remain at Medium, primarily because the harm vectors (NCII, doxxing, extortion) are driven by user behaviour and external dissemination. Although multi-layered controls (AI, manual review, takedown cooperation, user verification, LEA collaboration) prevent prolonged exposure, they cannot fully reverse harm. Residual exposure therefore represents the boundary between feasible mitigation and user-driven offences that extend beyond Platform control.

- **Physical and Mental Well-being (4 risks – 1 Low, 3 Medium residual risks):** Risks related to compulsive content consumption and exposure to extreme or violent sexual material remain at a medium level. These risks are inherent to open user-generated content ecosystems and engagement-optimised digital environments. Existing measures cannot fully regulate individual consumption patterns without compromising user autonomy or freedom of expression. As a result, further risk reduction in this category depends on broader product-level design considerations.
- **Non-Discrimination (4 risks – 3 Low, 1 Medium residual risk):** Residual exposure in this category is limited. The multilingual moderation system and translator tools ensure consistent enforcement, but slight residual disparities across languages and regional interpretations keep one scenario at medium.
- **Human Dignity (2 risks – 1 Low, 1 Medium residual risk):** While moderation and ToS enforcement minimize exposure to degrading or dehumanising content, contextual ambiguity (e.g., distinguishing satire from harm) prevents complete elimination. Residual medium risk persists where enforcement must carefully balance user rights and content interpretation.
- **Gender-Based Violence (8 risks – 4 Low, 4 Medium residual risks):** The Platform’s detection models and advertiser review processes have significantly reduced harmful or exploitative material. Still, medium residual risk remains due to the rapid evolution of synthetic media (e.g., deepfakes) and persistent normalization of harmful stereotypes through borderline user content. Complete elimination of these risks would require content restrictions that could infringe lawful expression.
- **Freedom of Expression and Information (3 risks – 100% Low residual risk):** Residual risk is low owing to clear moderation standards, complaint-handling procedures, and the balance between removal of illegal content and protection of lawful expression. Controls effectively prevent systemic suppression of legitimate content.
- **Dissemination of Illegal Content (19 risks – 10 Low, 9 Medium residual risks):** This category presents the largest volume of assessed risks but also demonstrates robust mitigation. Automated and manual moderation, AI detection tools (Vercury, Hive, SafetyNet API, Safer), and trusted flagger mechanisms have substantially lowered exposure. However, medium residuals persist due to adversarial adaptation (e.g., re-uploading altered files, using link shorteners, multilingual evasion). These limitations are intrinsic to user-generated content ecosystems but are continuously managed through iterative tool improvement and moderator training.
- **Data Privacy and Protection Risks (5 risks – 100% Medium residual risk):** All risks in this category remain at medium due to the structural sensitivity of large-scale data processing. Although GDPR-aligned controls, explicit consent mechanisms, and strong access security have reduced exposure significantly, the inherent vulnerability of personal and sensitive data, including risks of re-identification, third-party misuse, and data inference, cannot be eliminated. Residual exposure persists mainly because of external dependencies (e.g., service providers) and the impossibility of absolute prevention against advanced profiling or unauthorized linking of anonymised data.
- **Consumer Protection (6 risks – 1 Low, 5 Medium residual risks):** Risks within this category primarily stem from uneven user experience in redress and transparency processes. While key measures such as complaint-handling systems, transparent advertising, and clear ToS are in place, users may still face variability in reporting/appeals experiences and consistency across jurisdictions and languages. In several cases the RRF has been conservatively set or adjusted to 0.8 to reflect additional effectiveness from the free-content model and clearly signposted cancellation flows; even so, structural process complexity prevents a uniform shift to low.
- **Civic and Electoral Impact (5 risks – 100% Low residual risk):** Residual risk in this category is minimal, reflecting the effectiveness of keyword detection, advertisement review, and moderation controls in

identifying and removing politically sensitive or manipulative content. As a result, these risks are considered well-contained.

Risk levels declined where controls directly disrupt the harm pathway. Automated detection and hashing/AI tools (such as Vercury, Hive, SafetyNet, Safer, and MD5) help identify illegal or harmful content early. Layered human review improves accuracy. Structured notice-and-action systems (for example, report buttons, abuse and copyright forms, and complaint handling) shorten the time needed for takedown and redress.

Policy and governance measures, such as ToS confirmation, repeat-infringer enforcement, advertiser certification, and a GDPR-aligned privacy framework, set clear behavioural standards. Collaboration with law enforcement, trusted flaggers, and relevant government bodies further accelerates responses to serious harms.

In summary, the current risk-management framework has driven a broad reduction from inherent to residual risk, eliminating High residual ratings and concentrating exposure within Medium and Low bands. Accordingly, the Platform's Risk Response Strategy (outlined in the next Section 6) provides a structured mechanism for determining how these residual risks will be managed moving forward.

6. Risk Response Strategy

6.1 Risk Response Strategy

NKL has adopted specific response strategies tailored to the residual risk scores that align with the risk management strategy and are based on the residual risk scores as follows:

- Low (Score 1-5): **Accept**
- Medium (Score 6-15): **Accept / Disclose** or **Maintain / Monitor**
- High (Score 16-25): **Reduce**

Accept

For risk scenarios with residual risk “**Low**”, the risk is accepted. Under the Accept strategy, the Platform acknowledges the existence of a given risk but does not implement new controls, as its probability and severity are minimal or already fall within the Platform’s defined tolerance level. In such cases, no additional mitigation measures are considered necessary at this stage, although periodic reviews are conducted to ensure that the risk level does not increase over time.

Accept / Disclose

For risk scenarios with residual risk “**Medium**” the risk is either accepted (disclosed) or maintained and monitored. The Accept / Disclose strategy is applied where the Platform recognises that a particular risk cannot be realistically reduced, as it stems primarily from human or societal behaviour rather than from the Platform’s own actions. In these instances, the Platform either discloses the nature of the residual risk and continues to monitor relevant trends, user feedback, and complaints to detect any material changes.

Maintain / Monitor

For risk scenarios with residual risk “**Medium**” the risk is either accepted (disclosed) or maintained and monitored. The Maintain / Monitor strategy is used where all practical and proportionate risk-reduction measures have already been implemented. In such cases, the Platform focuses on maintaining existing controls, actively monitoring early warning indicators, and ensuring continued compliance with regulatory developments. Escalation occurs only if there is a significant change in the probability or severity of the risk.

Reduce

Finally, the Reduce strategy applies where the residual risk “**High**” is calculated, thus the Platform identifies the need to lower the probability or severity of a given risk to an acceptable level. This involves introducing or strengthening mitigation measures, such as policies, automated tools, training, or technical solutions, and assigning clear ownership, responsibilities, and implementation timelines to ensure the effective application of the chosen controls.

6.2 Action Plan 2024-25 Implementation Status

The last iteration of risk assessment conducted in 2024 has defined additional action points planned for implementation. Below is described the state of progress of these mitigation strategies as of November 2025.

Networking with experts and NGOs focusing on the health and social impacts of adult content / Collaboration with NGOs and experts on online protection of minors

NKL considers protection of minors as high priority field of interest and has assigned a dedicated person for maintaining the cooperation with NGOs regarding this issue (see also measure “Strengthening age verification and protection mechanisms” below). Such cooperation represents an ongoing and multi-dimensional engagement process as part of NKL’s strategy to prevent and mitigate online harms, particularly in relation to illegal content and child protection. The cooperation is not a one-off or isolated activity but an ongoing process. For example, StopNCII’s recent conference provided opportunity to network with a wide range of stakeholders from survivors to government, regulation to large online Platforms and increases understanding of the harms, mitigation measures and grows the formal partnership which has been established with StopNCII where the focus is on adult related harms. Engagement with a child protection focus remains ongoing, in particular in relation to technology and moderation lead protections and access controls. It’s important to note that engagement is not neatly compartmentalised into adult harms and child protection given common crossover in interests (e.g. through OffLimits the focus is on both CSAM prevention and tools to deter (and provide helpline services to) adults who are potentially engaging in harmful behaviour).

Current Partnerships and Engagements:

- **StopNCII / SWGfL:** Ongoing cooperation focused on the use of StopNCII’s hash tool for identifying non-consensual intimate image abuse. NKL has been engaging with SWGfL since mid-2024 and is now in the final stages of the process of being covered by the StopNCII tool under a contract signed in mid-2025. The collaboration also involves participation in regular partnership meetings and conferences arranged by StopNCII.
- **OffLimits:** NKL maintains an active partnership with OffLimits, utilising its CSAM hash lists to identify and remove illegal material. The cooperation extends to work on incorporating a helpline service and deterrence messaging targeting adults who are potentially engaging in harmful behaviour, alongside OffLimits’ partners in Belgium and the UK.
- **Interpol:** NKL has participated in in-person and online meetings on child safety, AV technologies, and the relative merits of site- versus device-level approaches and is engaged in the development of a partnership process.
- **WeProtect Global Alliance:** NKL maintains an active relationship with the Alliance through direct contact with its CEO. The cooperation focuses on information exchange focused on adult content regulation and child protection.
- **Canadian Digital Governance Standards Institute:** The Regulation Director has been actively involved since April 2024, including as a member of the drafting team for a national age verification standard which has recently been published.
- **UK Expert Panel on Age Verification:** The Regulation Director serves as a member of this UK Government-sponsored panel, which meets quarterly brings together a range of experts from different organisations to share knowledge on regulation, policy and technology and to help inform government.
- **Trust & Safety Professional Association:** The Regulation Director is an active member, regularly engaging with peers on online safety.
- **Age Verification Providers Association / euCONSENT:** Multiple meetings have taken place since 2024, focusing on AV methods, efficacy, interoperability, and regulatory frameworks.
- **Insafe / INHOPE Network:** NKL has participated in both online and in-person meetings on child safety, AV technologies, and the relative merits of site- versus device-level approaches.

Strengthening age assurance mechanisms

NKL maintains a continuous process for monitoring and engaging with developments in age assurance technologies. This includes direct collaboration with trade and professional bodies, individual companies providing age-verification services and other experts through collaborative working, including parties closely watching the UK experience. Direct engagement includes participation in expert panels and industry associations that closely follow the evolution of AV tools.

The Regulation Director was directly involved in the development of a standard with the Canadian standards body and engages directly with the Age Verification Providers Association and the Age Check Certification Scheme. On the other hand, NGOs generally provide less expertise on technology and standards than the professional and trade bodies. Further, reports are reviewed, including the recent report to the eSafety Commissioner in Australia, and development of standards is monitored (if not actively engaged in).

Moreover, the technical team supporting XNXX are engaging with the European Commission on testing their interim age verification solution. XNXX's own experience with the application of age verification at the Platform level for UK located users provides valuable insight. The deployed age verification system requires every new user in these jurisdictions to complete the age verification process upon their first visit to the Platform before they can view any adult video content. Currently, a biometric 'selfie' check or credit card method is applied. The process duration varies depending on the method and user, as verification has been seen to fail, despite users being adult, requiring the user to retry or select an alternative method. Metrics from the UK are being compiled and examination of these remains ongoing. NKL continues to monitor this space closely.

Key interactions include:

- **Interpol:** NKL has participated in in-person and online meetings on child safety, AV technologies, and the relative merits of site- versus device-level approaches.
- **Canadian Digital Governance Standards Institute:** The Regulation Director has been actively involved since April 2024, including as a member of the drafting team for a national age verification standard.
- **UK Expert Panel on Age Verification:** The Regulation Director serves as a member of this UK Government-sponsored panel, which meets quarterly brings together a range of experts from different organisations to share knowledge on regulation, policy and technology and to help inform government.
- **Age Verification Providers Association / euCONSENT:** Multiple meetings have taken place since 2024, focusing on AV methods, efficacy, interoperability, and regulatory frameworks.
- **Insafe / INHOPE Network:** NKL has participated in both online and in-person meetings on child safety, AV technologies, and the relative merits of site- versus device-level approaches.

Develop a standardized user-counting methodology

Following the second iteration of the risk assessment, this action point has been classified as a process-oriented initiative rather than a measure directly mitigating a specific risk. Nevertheless, to ensure continuity with the previous action plan, it is noted that the development of a standardized user-counting methodology remains ongoing. The team has received and reviewed materials outlining the European Commission's methodological guidance, which are currently being compared with the Platform's internal calculation approach. The finalization of the methodology is anticipated by March 2026.

Compliance management and auditing

The Compliance Team is continuing the implementation of the Compliance Management System with a specific focus on the requirements of the DSA. The team continuously monitors regulatory developments related to the DSA, maps existing Platform processes, and coordinates with the respective internal teams responsible for implementing the

relevant obligations, including areas of moderation, notice and action mechanism, recommender system, transparency and advertising. Accordingly, the development and update of internal documentation are currently in progress.

External audit

The external audit conducted pursuant to Article 37 of the Digital Services Act is currently in progress and is expected to be finalized by 13 November 2025. Following receipt of the Independent Audit Report, an Implementation Report will be prepared outlining the audit findings and planned follow-up actions.

Continuous review of studies related to systemic risks and mitigation measures

As part of the annual risk assessment process, the Compliance Team conducts a review of external studies and expert materials relevant to the identification and evaluation of systemic risks and corresponding mitigation measures. The team maintains an internal library of these sources, which is available to the European Commission upon request.

6.3 Action Plan 2025-26

In accordance with the predefined risk response strategies, mitigation measures assessment outcomes, and Action Plan 2024-25, the current Action Plan for the current risk assessment cycle comprises of the following action points:

Action point	Description
Enhance and align transparency of all notice mechanisms	Align feedback process across all reporting channels with the abuse reporting form workflow to ensure that reporters can track the status of their submitted notices and receive clear follow-up information on the outcome of the content review.
Collaboration with the regulators	Maintain communication with regulatory authorities in line with legislative priorities and legal obligations for protection of minors. Engages with regulators primarily in jurisdictions of strategic importance, including Italy, while continuing dialogue with authorities in the United Kingdom, France, and the Czech Republic.
Monitor EU Guidance on Interface Design Practices	Continuously monitor the publication of European Commission guidelines concerning manipulative interface designs (dark patterns). Once such guidance becomes available, assess its relevance to the Platform's interface and take appropriate steps to ensure consistency with the recommended practices.
Continuously monitor development of Inherent Risk Values	Monitor the evolution of inherent risk values to identify any early signs of deterioration or emerging threats (increases in probability or severity). If such changes are observed, reassess the adequacy of existing controls and, where necessary, introduces additional or enhanced mitigation measures.
Continuously monitor development of Risk Driver Values	Monitor risk driver values. If monitoring indicates a decline re-evaluate Residual Risk Values and, if necessary, introduce targeted improvements or supplementary actions.
Collaboration with NGOs and experts on online protection of minors	<p>Liaison with reputable NGOs to fund and implement programs aimed at raising awareness about the importance of online safety and security for minors. These programs focus on educating minors and their parents or guardians on risks associated with internet use and how to mitigate them.</p> <p>Engage experts on age verification and minor protection to ensure that the NKL's risk mitigation strategies consider the latest best practices and innovative solutions. This collaboration would help fine-tune policies that affect vulnerable groups.</p> <p>Monitor the industry recent developments of solutions that effectively prevent underage access and respect fundamental freedom and privacy as well as safety as well as business considerations, such as operational efficiency.</p>
Define KPIs for mitigation measures effectiveness assessment	In cooperation with the relevant teams, define KPIs for individual measures, where applicable. These KPIs will serve as the basis for the quantitative assessment of the

effectiveness of mitigation measures and support consistent monitoring of their performance over time.

Compliance Management System implementation (DSA)	Continued implementation of the CMS ensuring it aligns with the latest regulatory requirements under the DSA and other relevant regulations. The CMS should be designed to manage and mitigate risks related to content moderation, user privacy, and illegal content dissemination.
Internal audit	Evaluate the adherence to internal policies and external regulatory frameworks. These audits are essential for identifying potential non-compliance areas and implementing timely corrective actions.
External audit	External audits of CMS assess the system's effectiveness in managing compliance risks, identifying gaps, and recommending improvements. It helps ensure that internal processes are robust and meet the requirements set by authorities.
Review of external sources related to systemic risks and mitigation measures	As part of the annual risk assessment process, conduct a review of external studies and expert materials relevant to the identification and evaluation of systemic risks and corresponding mitigation measures.
Risk monitoring	Adhere to ongoing and periodic monitoring of systemic risks. The process includes regular reviews of internal data and reports, continuous tracking of industry and regulatory developments, and proactive identification of potential risk indicators. Findings from the monitoring process are used to update the risk and mitigation measures register and inform adjustments to existing action plan.
Data Management for Reporting and Compliance	Develop a standardized user-counting methodology, supported by regulators, which incorporates insights from public and regulatory consultations to ensure accuracy in reporting.

Each action point has been assigned to a designated owner to ensure clear accountability and transparent implementation. Internal deadlines have been established for all actions to secure timely execution and prevent the measures from remaining merely declarative.

According to the risk response strategy outlined in this Section, for each risk scenario rated as “Medium” have been assigned corresponding additional measures described in the Action plan. Such comprehensive overview of all risk scenarios with residual risk “Medium” and corresponding measures is provided in the Annex to this report.

Annex 1

Action Plan

Below are listed risk scenarios with residual risk value “Medium” accompanied by additional mitigation strategies. These strategies are described in detail in the Section 6.3 Action Plan 2025-26.

ID	Risk Scenario	Response Strategy	Measures to Implement
SR-IC-01	Users upload or share non-consensual intimate images or videos (NCII), including revenge porn, deepfakes, leaks behind paywalls, and doxing links	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-IC-02	Users upload prohibited sexual or exploitative material, including extreme pornography or coerced/trafficked content	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-IC-03	Users publishing comments that promote or solicit sexual services, including self-prostitution, self-advertising of sexual activities, or the offering of escort services	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-IC-05	Users distribute child sexual abuse material (CSAM) in the form of uploads (e.g., images, videos)	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-IC-06	Users share URLs, QR codes, or link shorteners in the comments that redirect to externally hosted CSAM	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values

SR-IC-09	Users use comments to spread malware or phishing links that compromise accounts or devices	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-IC-10	Users share copyrighted or IP-protected content (e.g., pirated videos, stolen creator content, counterfeit brands)	Maintain / Monitor	Enhance and align transparency of all notice mechanisms . Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-IC-12	Users coordinate harassment, stalking, threats, coercive behavior, or mobbing campaigns via comments	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-IC-18	Users post comments that promote violence, hatred, or discrimination against protected groups	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-HD-01	Users post degrading or dehumanising sexual content (e.g., extreme humiliation, racist or misogynistic abuse) that normalises the violation of dignity	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PF-01	Users upload or share non-consensual intimate images or videos (“revenge porn”), including leaked paywalled content, with the intention of damaging the private life of the targeted individual	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values

SR-PF-02	Users doxx creators or victims by publishing personal data (addresses, contact details, family information) linked to explicit content	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PF-03	Users upload and disseminate private content or information in order to threaten or blackmail individuals, deliberately causing harm to their private and family life	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PF-04	Users misuse manipulated images of real individuals (e.g., ex-partners, private civilian photos, or public figures), including deepfakes or photoshopped material, publishing such content to harm reputations and family relationships	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PF-05	Sensitive personal data (location, metadata, contact details) tied to adult content is exploited by offenders, threatening private life and family life	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-DP-01	Large-scale exposure of intimate content and identifiers leads to blackmail and harassment	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-DP-02	Sensitive sexual identity or preference data are used for profiling, leading to discrimination or unequal access to the Platform	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values

SR-DP-03	Pervasive tracking chills lawful sexual expression and undermines user autonomy and freedom of association/expression	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-DP-04	Sensitive data are processed at scale without valid consent, eroding autonomy and enabling exploitation or unlawful secondary use	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-DP-05	Intimate material or identifiers become linked to victims' families or communities, damaging dignity and relationships	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-ND-03	Systemic discrimination arises where users in non-English regions receive weaker protection from harmful content, facing higher risks than English-speaking users	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PM-01	Minors access pornographic content on the service, resulting in exposure to harmful or illegal sexual material	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Collaboration with NGOs and experts on online protection of minors Collaboration with the regulators Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PM-02	Users upload, or share material depicting sexualized minors (CSAM) or images of minors in sexualized contexts	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values

SR-PM-03	Minors are systematically exposed to explicit pornographic content via the platform's content delivery	Maintain / Monitor	<p>Collaboration with NGOs and experts on online protection of minors</p> <p>Collaboration with the regulators</p> <p>Continuously monitor development of Inherent Risk Values</p> <p>Continuously monitor development of Risk Drivers Values</p>
SR-PM-04	Minors encounter harmful sexual content through discovery features (e.g., search suggestions, autocomplete, trending tags)	Maintain / Monitor	<p>Enhance and align transparency of all notice mechanisms</p> <p>Collaboration with NGOs and experts on online protection of minors</p> <p>Collaboration with the regulators</p> <p>Continuously monitor development of Inherent Risk Values</p> <p>Continuously monitor development of Risk Drivers Values</p>
SR-PM-07	Minors are exposed to sexually explicit or pornographic advertisements, including ads for live sex cams, adult dating apps, sex toys, and sexual enhancement products	Maintain / Monitor	<p>Collaboration with NGOs and experts on online protection of minors</p> <p>Collaboration with the regulators</p> <p>Continuously monitor development of Inherent Risk Values</p> <p>Continuously monitor development of Risk Drivers Values</p>
SR-PM-08	Minors' well-being is impaired by exposure to pornography, leading to anxiety, confusion, and distorted views of relationships and sexuality	Maintain / Monitor	<p>Collaboration with NGOs and experts on online protection of minors</p> <p>Collaboration with the regulators</p> <p>Continuously monitor development of Inherent Risk Values</p> <p>Continuously monitor development of Risk Drivers Values</p>
SR-PM-10	Weak or bypassable age verification measures allow minors to access pornographic content, exposing them to harmful or illegal sexual material	Maintain / Monitor	<p>Enhance and align transparency of all notice mechanisms</p> <p>Collaboration with NGOs and experts on online protection of minors</p> <p>Collaboration with the regulators</p> <p>Continuously monitor development of Inherent Risk Values</p> <p>Continuously monitor development of Risk Drivers Values</p>
SR-PM-11	Intrusive or unsafe age verification systems (e.g., biometric scans, ID uploads) retain or process sensitive identity data without sufficient safeguards, exposing minors and adults to privacy violations	Maintain / Monitor	<p>Collaboration with NGOs and experts on online protection of minors</p> <p>Collaboration with the regulators</p> <p>Continuously monitor development of Inherent Risk Values</p> <p>Continuously monitor development of Risk Drivers Values</p>

SR-CP-01	Users are systematically left uncertain about their rights, protections, or obligations on the Platform, leading to diminished trust and reduced ability to exercise consumer rights	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-CP-02	Users face persistent barriers to flagging illegal/harmful content, appealing moderation decisions, or resolving disputes, resulting in widespread denial of effective remedies	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-CP-03	Platform uses manipulative interface designs (dark patterns) that pressure users into subscriptions, or sharing personal data	Maintain / Monitor	External audit Monitor EU Guidance on Interface Design Practices Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-CP-05	Lack of accessible and responsive user-Platform communication mechanisms (e.g., reporting, appeals, customer support) systematically deprives users of protection	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-CP-06	Unequal enforcement of consumer rights and protections across Member States results in systemic disparities, i.e., some users face greater exposure to harm or weaker redress	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-GB-02	The visibility and normalisation of misogynistic, sexist, or toxic content on the Platform reinforces harmful gender stereotypes and perpetuates gender-based violence	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values

SR-GB-04	Stolen, coerced, or fabricated intimate images are used to extort money or sexual acts from women, causing severe mental, emotional, and physical harm	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-GB-05	Non-consensual sexual content, including AI-generated deepfake pornography that targets women, is uploaded and shared, violating privacy, dignity, and well-being	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-GB-08	The Platform publishes or allows advertisements that depict or normalise degrading portrayals of women (e.g., humiliation, coercion, violent sexual acts), reinforcing harmful stereotypes and legitimising gender-based violence	Maintain / Monitor	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PH-01	Users/advertisers upload/promote content/advertisements that encourages unprotected sex or risky sexual behaviours without disclaimers, normalising unsafe practices	Accept / Disclose	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PH-03	Exposure to content that promotes unrealistic or unhealthy body standards contributes to self-esteem issues, or disordered behaviours	Accept / Disclose	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PW-02	Prolonged Platform use leads to compulsive consumption that harms users' health	Accept / Disclose	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values

SR-PW-03	Sharing or distribution of non-consensual intimate content poses serious risks to victims, including psychological trauma, heightened anxiety, and depression	Maintain / Monitor	Enhance and align transparency of all notice mechanisms Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values
SR-PW-04	The ad system boosts compulsive porn consumption through constant push ads (e.g., “live now”, “exclusive access”), encouraging addictive behaviour and degrading users’ mental health	Accept / Disclose	Continuously monitor development of Inherent Risk Values Continuously monitor development of Risk Drivers Values