ECLAG Comments on Belgian Presidency Proposal for Risk Categorisation
March 2024

This feedback was drafted by ECLAG* Steering Group following consultation with members of the whole coalition.

* The European Child sexual abuse Legislation Advocacy Group (ECLAG) is a coalition of child rights NGOs joining forces to fight to protect children from sexual violence and abuse. ECLAG brings together over 65 European and global NGOs. It supports the #ChildSafetyON campaign to call for laws and policies to ensure children are safe online. We cannot wait any longer! EU policymakers must act decisively to end online child sexual abuse once and for all. ECLAG Steering Group is formed by Brave Movement, ECPAT International, Eurochild, Missing Children Europe, Internet Watch Foundation, Terre des Hommes and Thorn.

—

ECLAG welcomes the Belgian Presidency's commitment to tackle all forms of child sexual abuse in all services, including end-to-end encrypted ones. We understand the Belgian Presidency proposal as a step to respond to find a compromise within the Council. We are committed to ensure that the discussions around this proposal won't compromise on child safety online.

Introduction

- ECLAG welcomes the commitment of the Belgian Presidency to ensure the protection of children from **all forms of child of sexual abuse (CSA) in all online spaces, including end-to-end encrypted (E2EE) environments**.
- ECLAG recalls the importance to protect children from against **all form of child sexual abuse** (grooming, known and unknown child sexual abuse material). All children who experience sexual abuse deserve protection. Grooming represents a growing threat for children with a 82% increase in 2022.
- ECLAG welcomes the **inclusion of E2EE services within the scope of the Regulation**. While offenders widely use E2EE platforms to ensure impunity for their crimes, it is crucial to mandate detection and innovation to protect children in encrypted spaces.
- ECLAG stresses that the **targeted detection of suspects** does not allow to tackle the scale of the phenomenon online and should be limited to platforms with negligible risk.
- ECLAG welcomes the inclusion of a possibility for online service providers to ask for authorisation to detect child sexual abuse on their services. ECLAG has long been calling for a

clear legal basis for a **voluntary detection framework** to avoid significant protection gaps and allow innovation, if Europe does not want to fall below today's effort.

- ECLAG calls on the Belgian Presidency to also include a **legal basis for the EU Centre to conduct pro-active detection in online public spaces** - building up on the existing work and expertise in conducting public proactive detection. This would help reduce the availability of CSA online and contribute to building the expertise of the EU Centre in fighting CSA online.

Risk categorisation

General principles

- The **notion of risk** should be defined in accordance with the **objectives of the Regulation**: to protect children from all sexual abuse and to remove all instances of child sexual abuse from online spaces in order to prevent the revictimisation of the child and future harm. Research demonstrates that the viewing and sharing of CSAM is strongly correlated with seeking online and offline contact with children. Platforms must therefore be assessed not only on the risk that they are used to groom children but also to exchange and share child sexual abuse materials.
- **Risk analysis and risk management is a science** and any selected methodology needs to be **evidence-based** and stressed-test before adoption. It should draw from the expertise of relevant stakeholders, including online service providers and civil society organisations.
- The selected methodology should also acknowledge the **continuous and iterative nature** of risk assessment and risk management processes.
- Risk assessment should be **service-specific**, i.e. conducted on a **case-by-case basis**, taking into account the specificities of a service and how criteria interact with each other on that service. Risk assessment should recognise that there is no one-size-fits-all approach to combat child sexual abuse; what may be suitable and sufficient for one service, may not be for another.
- As the digital environments and the manifestation of CSA rapidly evolve, it is crucial to ensure **future-proof methodology and criteria** to avoid leaving out of detection scope potential high risk services in the future. Accordingly, **any criteria listing should remain indicative** as it will be subject to evolving.
- **Risk assessment should be mandated on a regular basis** to ensure its relevance and to avoid offenders moving to platforms labelled "low or negligible risk"where they know detection is not conducted.

Recommended methodology

- ECLAG recommends a **methodology that combines both an objective risk analysis (factors/environment-based) and a real risk analysis (evidence-based)**.
- The **risk factor/environment analysis** should assess the level of risk of the services based on multiple objective criteria, including:
    - Categories of services;

    - Architecture and functionalities of services (notably a safety by design architecture, user identification, age verification/assurance functionalities, end-to-end encrypted spaces)
    - Existing policies (taking into account the limited impact of policies in preventing risk) ***(Please find detailed comments on the specific criteria for each methodology in Annexe).***
- **The weight attributed to these criteria differs in theory and in practice**, notably depending on how they interact with each other. Their impact also differs depending on the risk/objective considered (e.g. age verification mechanism may help reduce grooming but has no impact on the dissemination of CSAM). This balancing exercise cannot be formalised in wording and will require expertise from the coordinating authorities and the EU centre (see below).
- This risk factor analysis must be complemented with **real risk analysis** based on:
    - Tendencies and statistics;
    - **Evidence of the service or evidence stemming from comparable services having been used in the past 12 months** and to an appreciable extent for the dissemination of CSAM or the solicitation of children (see Article 7.5: known CSAM, Article 7.6: new CSAM, Article 7.7: solicitation of children of the proposed CSAR).
- To gather evidence, platforms would ideally be entitled to temporarily and to a limited extent use detection technology to get a real assessment of the misuse of their platforms for CSA. Besides, other type of evidence could contribute to assess the real risk of the platform including:
    - evidence stemming from the implementation of the DSA (e.g. reports from trusted flaggers, transparency report and transparency database)
    - the users' reports collected by hotlines and helplines;
    - results from proactive search by the EU Centre in public online spaces - as recommended by the European Parliament.
- ECLAG recalls that analysing metadata alone is not sufficient to assess the risk due their limited effectiveness in tackling the dissemination of CSAM.
- Any evidence and sample-based analysis should address the risk that platforms select the evidence they are presenting to influence the categorisation of their services. Insights from platforms in designing these criteria would however be particularly relevant.
- Without a real risk analysis, providers will lack the necessary insight into the actual dangers posed by their services. It would lead to ineffective or misguided mitigation strategies, wrong risk categorisation and impossibility to use detection technologies, ultimately leaving children unprotected.

Implementation concerns

- While ECLAG welcomes greater detail on the risk assessments and categorization, we stress that **successful implementation will require expertise and resources** - be it at the Coordinating authorities or the EU Centre level.

- ECLAG argues that **establishing a central body within the EU Centre to process risk assessment** will be pivotal to success. The EU Centre should:
    - ensure a qualitative and harmonised approach across EU Member States,
    - alleviate the burden for countries with large presence of online service providers;
    - build up expertise thanks to its global overview of the mitigation measures adopted by providers - expertise which will best position the EU Centre to draft guidelines on risk assessment and risk mitigation measures.
- While this inevitably means additional resources for the EU Centre, ECLAG believes that if not the EU Centre, these resources would need to be allocated to the national authority in charge. By mandating the EU Centre, some economies of scale can be expected.

Risk mitigation and scope of detection orders

- ECLAG stresses that '**standard detection orde**r' implemented in high risk services must allow for the effective detection, report and removal of child sexual abuse **at scale** - including in E2EE services.
- ECLAG expresses concerns regarding **'limited detection order' in medium risk services** and how that would be implemented in practice. **Targeted detection of suspects** does not allow to tackle the scale of the phenomenon online and **should be limited to platforms with negligible risk.**

ANNEXE - Comments on possible risk categorisation criteria

Please find below ECLAG comments on Belgian Presidency's Annexe. As mentioned above, ECLAG recommends a combination of these approaches, in addition to a real risk analysis - partly missing from the current Proposal.

1. Based on the category of services

- Many platforms will offer more than one of these services and we can't really separate one service from its interaction with the other services. Besides, ECLAG recalls that all platforms and services are (likely to be) used by offenders, even if they are diverted from their intended purpose. Recently, CSAM were found on secondhand marketplace or music listening platforms. This is why ECLAG recommends to consider this categorisation of risk only in combination with other (architecture/features, polices but also real risk analysis).
- *Adult services*: this seems both vague and limited definition and should be made more explicit if it concerns the sharing of sexually explicit content. It's unclear how this is a category of service and whether this should be rather approached under the architecture categorisation.

2. Based on the core architecture of the service

- ECLAG recommends to group both architecture and design functionalities as one risk categorisation since these can hardly be distinguished. For instance, access for children to the

service as well as user identification functionalities go hand in hand with age verification/assurance functionalities. User identification functionalities are 'functionalities' which both impact the level of interaction between users and help to protect child users.

- *User identification functionalities:* sharing content anonymously is relevant but the consideration of anonymity might not be limited to the sharing of content only: 'to use the platform anonymously' is equally relevant.
- *Possibilities on user communication:* ECLAG stresses that any of these would most likely trigger high risk categorisation - certainly if combined with full anonymity.
- ECLAG also suggests to include:
    - Storage functionalities (in particular how the information is stored, for how long, for what reason and how law enforcement authorities can have access to what type of stored information)
    - Existence of download/save/screenshot/screen video functionalities
    - Existence of Child Rights Impact Assessment tools

3. Based on policies and safety by design functionalities in place

- As mentioned above, this listing should be merged with the 'architecture' listing.
- *Effectiveness of CSA Risk Policies:* this would require more clarity on how effectiveness is measured and what is a CSA Risk Policy.
- *Age verification:* this could be renamed to include age assurance and age verification measure. While ECLAG recognises the importance that such measures be proportionate, effective and privacy preserving, their impact on the risk may differ - also depending on the risk/objective concerned (eg. age verification mechanism may help reduce grooming but has no impact on the dissemination of CSAM).
- Many of the measures and functionalities proposed (including notification of CSA) should also be assessed for their age appropriateness and in line with the evolving capacities of children.
- *Efficiency in Handling Notified/flagged Potential Child Sexual Abuse*: this should be further defined, taking into account the review processes, its length and the action adopted (including action with hotlines and law enforcement authorities)
- ECLAG also recommends to assess:
    - Content moderation process (when, how, what is and what is not detected)
    - Definition of CSA in Terms of Services,
    - Transparency reports on CSA,
    - Safety by design measures such as public/low privacy profiles, visible location data, recommender systems and in-game and in-app gifts.
- The suggested criteria for scoring as effective and comprehensive are focused on reporting, age-appropriate interface and education. While those are important, they are far from enough to tackle CSA at scale. Detection is required for effective functionalities together with child safety by design features (e.g. high privacy settings by default, not allowing to search/ engage easily with children, etc.).

- *Measures for Promoting Users' Media Digital Literacy and Safe Usage Scoring System:* although these measures are crucial, more evidence is needed to determine how these (in particular media literacy) impacts the level of risk on a platform.
- *Alignment of Business Model, Governance and Systems with CSA risk mitigation:* more evidence is needed to determine how these impact the level of risk on a platform.
- *Functionalities enabling users to Share Potentially Harmful Content*: more clarity is needed on whether harmful or only illegal content should be considered.
- *Functionalities Assessment of Potential Dissemination Risks*: more clarity is needed on the definition of this criteria.