Council of the European Union
General Secretariat

Interinstitutional files:
2022/0155 (COD)

Brussels, 27 March 2024

WK 3036/2024 REV 1

LIMITE

JAI	**FREMP**
ENFOPOL	**TELECOM**
CRIMORG	**COMPET**
IXIM	**MI**
DATAPROTECT	**CONSOM**
CYBER	**DIGIT**
COPEN	**CODEC**

This is a paper intended for a specific community of recipients. Handling and further distribution are under the sole responsibility of community members.

WORKING DOCUMENT

From: To:	Presidency Law Enforcement Working Party (Police)
Subject:	Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse - updated possible criteria and classification methodologies for the risk categorisation of services

The Presidency provides delegations in the Annex with a working document outlining updated possible criteria and classification methodologies for the risk categorisation of services to facilitate discussions at the meeting of the Law Enforcement Working Party (Police) on 3 April 2024.

ANNEX

Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse – updated possible criteria and classification methodologies for the risk categorisation of services.

Following the LEWP discussions on 1 and 19 March 2024 and on the basis of the written comments sent by the MS, the Presidency decided to adapt the working document which is intended to serve as a reference to feed Annex XV (methodology and criteria for the risk categorisation of services) and the related Annex XIV (template for the self-assessment of providers) which are referred to in Article 5 of the draft legislative text.

Article 5 lays down the process and the key principles for categorising services into 3 categories (high, medium, low). Annex XV will include the criteria and parameters guiding the categorisation and a concrete and precise method of classification. The presidency therefore proposes the criteria outlined below as a basis for the further discussion.

As stressed by some delegations, it is important to have an agile and flexible approach that is future-proof and takes account of consultations with businesses and civil society. To this end, this document should be considered as a basis for defining the methodology and criteria in ANNEX XV and the related ANNEX XIV, which may be adapted via a delegated act by the Commission.

We draw the attention of delegations to the fact that a delegated act is only feasible if the amendments that may be made by COM represent only adaptive/ complementary amendments. It is therefore necessary to regulate the essential/ core aspects of this categorisation procedure and the key principles of the methodology and the criteria to be applied in the operative part of the text. Everything that is related to limiting fundamental rights cannot be delegated and needs to be duly implemented in the operative text.

Based on the results of the discussion to define the categorisation procedure and delineate the criteria as well as an adequate taxonomy, it will be the next step to define the main and unalterable aspects of this content in order to extract the key principles that will be introduced in Article 5 (2a).

I. Possible risk categorization criteria

1) Based on the category of services

A first approach could be to distinguish the potential risk if the service provides one or more of the following service types. The level of risk of these categories could be based on accessible statistics, including the number of CSA reports that these types of services submit, statistics from law enforcement or other public authorities, or from other sources (e.g., academia, researchers, etc.).

- Social media platform (services that connect users and enable them to build communities around common interests or connections)
- Electronic messaging service (services that are typically centred around allowing users to send messages that can only be viewed or read by a specific recipient or group of people)
- Online gaming service (services that allow users to interact within partially or fully simulated virtual environments)
- Adult service[1] (services that are primarily used for the dissemination of user-generated adult content). For instance, adult services could be one or more of the following:
    - Camming Services: These platforms facilitate live streaming or webcam performances by individuals typically engaged in adult-oriented activities such as explicit conversations, striptease, or sexual acts for an audience.
    - Pornographic Websites: These are platforms that primarily host or distribute sexually explicit videos, images, or other adult content for viewing or download.
    - Adult Gambling Services: These services involve online betting or gambling activities that are explicitly geared towards adults and may include adult-themed games or gambling content.
    - Escort Services: Escort services connect individuals with escorts or companions for adult-oriented activities, which may include companionship, intimacy, or sexual services in exchange for payment.
    - Adult Social Networking Sites: These are platforms similar to mainstream social networking sites but cater specifically to adults interested in connecting with others for adult-oriented interactions, such as dating, casual encounters, or discussions about sexual topics.
    - Adult Dating Apps: These mobile applications focus on facilitating connections between adults interested in casual or intimate relationships, often emphasizing physical attraction and sexual compatibility.
    - Adult Content Subscription Services: These platforms offer access to exclusive or premium adult content through subscription-based models, providing users with a variety of adult-oriented media such as videos, images, or stories.

1 An "adult service" typically refers to an online platform or service that primarily deals with or facilitates the dissemination of adult content. This content may include, but is not limited to, explicit imagery, videos, or text that is intended for mature audiences and may contain nudity, sexual content, or explicit language. Adult services encompass a wide range of platforms, including adult websites, adult social media networks, adult chat rooms, adult streaming services, and adult dating or hookup platforms. These platforms are designed to cater to individuals seeking adult-oriented content, entertainment, or interactions. Note that adult services may vary in terms of the types of content they offer, the audience they target, and the services they provide. However, they share a common characteristic of providing access to adult-oriented material and often require users to confirm their age before accessing such content.

- Discussion forum or chat room service (services that allow users to send or post messages that can be read by the public or an open group of people).
- Marketplace or listing service (services that allow users to buy and sell their goods or services)
- File-storage and file-sharing service (services whose primary functionalities involve enabling users to store digital content and share access to that content through links).
- Online dating sites and dating apps (services that facilitate connections between individuals seeking romantic or interpersonal relationships, typically through profile creation, matching algorithms, and messaging features).
- Web and server hosting services (services that provide individuals or organisations with the infrastructure and technology needed to host websites or web applications on the internet, including server space, bandwidth, and technical support).
- Services defined as "VLOPs" (= very large online platforms)
- Services defined as "VLOSEs" (very large online search engines)[2]

2) Based on the core architecture of the service

This section assesses the level of interaction which is possible between users on a service.

a) Access for children to a part or the entirety of the service[3].

b) User identification functionalities
   - to display information through a user profile that can be viewed by others (e.g. images, usernames, age).
   - To use the platform anonymously.
   - To share content anonymously (e.g. anonymous profiles or access without an account).
   - Functionality preventing user to access to website(s) in another geographic region where the legislation is less strict.
   - To have a multi-factor authentication and user signup information gathering (phone number, email address, or other identifiers).

c) method of connecting users to reach other users[4]
   - Possibility to form closed groups and/or send group messages.
   - Possibility to search for user by specific categories (place, gender, hobbies, etc.)

[2] cfr Art 33 DSA – classification is made by calculating the number of average monthly active recipients of the service in the Union) + Art. 34 DSA - Risk assessment.
[3] Should take into account not just whether children can access the site but whether they do access the site.
[4] 'User connections' (UK Safety Act): A user-to-user service functionality that allows users to follow or subscribe to other users. Users must sometimes be connected to view all or some of the content that each user shares.

d) Possibilities on/ of user communication[5]
   1. Communication via livestream.
   2. Communication via direct messaging/ephemeral direct messaging.
   3. Communication via encrypted messaging and functionalities for "Opt-in/opt-out"[6].
   4. Posting/sharing images or videos (either open or closed channels).
   5. Reposting and forwarding content (either open or closed channels).
   6. Sharing of content via hyperlinks and plain-text URLs.
   7. Commenting on content (open and/or closed channels).
   8. Posting/sharing location information. Visible location data.
   9. Searching for user-generated content.

e) Possibility for users to post goods/services for sale[7]
   - Possibility to use cryptocurrency to buy service/material (promotes anonymity)
   - Existence of gift-card-related transactions

f) Possibility for allowing payments through the service.
   - Existence of download/save/screenshot/screen video functionalities.
   - Possibility to limit the number of downloads per user to reduce the distribution of illegal content.

g) 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

h) Existence of recommendation algorithms (algorithms that recommend content similar to that already viewed may potentially expose users to inappropriate content if they have already been exposed to child pornography)

i) Functionalities preventing users from making recordings and screenshots of shared content or saving a local copy of shared content (might be on opt out basis might be based on age verification - i.e. content shared by minors not being able to be saved, etc.)

[5] These criteria have been presented ranked to help for the future scoring system (to be developed). This ranking places activities involving direct real-time communication (livestreaming, messaging) at the highest risk due to their immediate and potentially unfiltered nature. Encrypted messaging follows closely due to privacy concerns and the potential for misuse. Posting and sharing of multimedia content are also high-risk activities, as they can easily disseminate harmful material. Reposting, forwarding, and sharing via hyperlinks carry a moderate risk, while commenting, sharing location information, and searching for user-generated content are deemed lower risk, though they still warrant attention in terms of potential risks.

[6] Making design choices such as whether the use of E2EE is opt-in by default, rather than opt-out which would require people to choose E2EE should they wish to use it, therefore allowing certain detection technologies to work for communication between users that have not opted in to E2EE.

Links to encrypted services are often shared on unencrypted online spaces to facilitate the exchange of CSAM.

[7] 'Posting goods and services' (UK Safety Act): a user-to-user service functionality allowing users to post content dedicated to offering goods and services for sale. This does not include paid-for-advertisement but may serve the function of allowing users to promote goods and services.

Potential perpetrators may try to promote illegal goods or services by posting them for sale using this functionality. Often illegal items such as drugs and firearms are posted for sale using code names. In certain contexts, the ability to post goods or services for sale, such as through user-generated advertisements, also enables potential perpetrators to advertise and broadcast the sexual services of adults in exploitative environments. The risk of harm can be increased if your services also allow users to make online payments directly.

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

This section assesses the level of measures taken by the providers to protect child users. 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.

a) Effectiveness of CSA Risk Policies[8]
b) Measures for Promoting Users' Media Digital Literacy and Safe Usage Scoring System [9]
c) Definition of CSA in Terms of Services
d) Strength of Prohibitions and Restrictions
e) Functionalities for Age Verification [10]
    - Privacy Protection: The age verification system protects user privacy, ensuring data is not disclosed or processed for any purpose other than age verification.
        - Minimal Data Collection: Only minimum of data is collected, adhering to a minimal data collection approach for age verification.
        - Data Retention: Personal data related to age verification is not retained after the verification process is completed.
        - Proportionality: The age verification system is proportionate to the risks associated with the product or service, ensuring a balanced approach.
        - Remedies and redress: Adequate remedies and redress mechanisms are provided for users whose age is wrongly identified.
    - Selective Disclosure: Users have the option for selective disclosure of attributes during the age verification process.
    - Zero-Knowledge Protocol: The system requires the user to provide a 'token' to confirm that the minimum age requirement is met.[11]
    - Anonymous accounts: Users have the option to use anonymous accounts for age verification, enhancing privacy protection.

[8] For instance, this criterion would be analysed on this basis:
The platform's approach to addressing child sexual abuse (CSA) risks varies across different levels of policy implementation. At the basic level, explicit policies specifically targeting CSA risks are absent. Moving toward effectiveness, policies related to CSA risks exist but lack regular updates, and users may find them somewhat unclear. Progressing to comprehensiveness, the platform demonstrates clear and explicit policies addressing CSA risks, which are regularly updated and generally well-understood by users. Finally, at the comprehensive level, the platform boasts explicit, user-friendly policies on CSA risks that are not only regularly updated but also enforced in a manner easily comprehensible to users.

[9] Duties about children's access assessments:
-A provider must carry out the first children's access assessment within one year
- The provider must carry out a children's access assessment of the service
*before making any significant change to any aspect of the service's design or operation to which such an assessment is relevant,
* in response to evidence about reduced effectiveness of age verification or age estimation that is used on the service
*in response to evidence about a significant increase in the number of children using the service.

[10] A provider is only entitled to conclude that it is not possible for children to access a service, or a part of it, if age verification or age estimation is used on the service with the result that children are not normally able to access the service or that part of it.

[11] Such a token is produced once a person's age is verified or estimated by an age assurance provider and allows the online service to confirm that the age requirement is met without viewing or collecting users' personal information. The token can be stored in a device's digital wallet or browser to be reused for a period of time to access services requiring the same level of assurance.

- Non-Identification Requirement: The age verification system does not require the identification of each user of a service.
- No Biometric Data Processing: The system does not require the processing of biometric data during the age verification process.

f) Functionalities for Parental Control Scoring System
g) Functionalities for Notifying/flagging Online Child Sexual Abuse[12]
h) Efficiency in Handling Notified/flagged Potential Child Sexual Abuse[13]
i) Statistical Information availabilities, completeness and Relevance
j) Functionalities preventing users to make recordings and screenshots of shared content or saving a local copy of shared content
k) Possibility to use peer-to-peer downloading (allows direct sharing of content without using centralized servers)
l) Functionalities Assessment of Potential Dissemination Risks
m) Possibility to delete shared content for all users it has been shared to[14]
n) Systems for selecting and presenting advertisements

4) Based on user tendencies and statistics

This section assesses the user tendencies and trends based on a statistical analysis of users.

a) Assessing User Patterns
   - Service's Popularity Among Different Age Groups
   - Existing level of confirmed risk on the service provider platform
b) Analysis of grooming Risks Based on User Mapping
   - "Cyber Flashing" tendency (Unsolicited Intimate Messages)
   - Analysis of Solicitation Risks Based on User Mapping
   - Creation of Private Group or Chatboxes
   - Moving Public Conversation to Private Channels
   - Identity Verification Tools for Opening Accounts
   - Use of Unsecured Public WIFI Hotspots
   - Obfuscation of IP Addresses
   - Use of Anonymous Account
   - Fake or Imposter Accounts
   - Pseudonymity

[12] Article 16 of the DSA already sets a legal standard for flagging and processing notices on potentially illegal content such as CSAM. Article 16 applies to all providers of hosting services (e.g. online platforms).
[13] For point f, "Efficiency in handling notified/flagged potential Child Sexual Abuse," the criteria can be further defined by considering the review processes, their duration, and the actions taken. This entails evaluating the promptness and thoroughness of the platform's response upon receiving notifications or flags regarding potential instances of Child Sexual Abuse (CSA). Efficiency encompasses the timeliness of initiating investigations, the depth of examination conducted during reviews, and the swiftness of implementing appropriate actions based on the findings. Additionally, it involves assessing the transparency of the platform's communication regarding the outcomes of these reviews and the measures taken to mitigate identified risks. This comprehensive evaluation ensures that platforms demonstrate not only a commitment to addressing CSA but also an effective and expedient approach to handling reported cases, thereby safeguarding users from harm.
[14] For scoring purposes it might be a positive or negative point depending on the implementation, it might be positive for the potential victim (able to delete extorted indecent pictures) but might be misused by perpetrators too. To score it as purely positive, the service provider should backup delete content for a reasonable amount of time (at least for accounts of minors or based on risk scoring).

- Frequent Changing Accounts of Profile Details
- Unmatching or Defriending Victims on Social Media Accounts
- Switching Between Private and Public Platforms
- Temporary Accounts
- Consecutive and Repetitive De- and Re-Activation of Accounts

c) Data related practices of the provider

5) Related to Company Policy on User Safety

This section assesses the measures implemented by the service to ensure the safety of its users.

a) Usage of Premoderation functionalities
   - Use of text moderation (specific term, hashtags, emoji, abbreviations)

b) Usage of Delisting Content System

c) Usage of Image Masking

II. Possible scoring methodologies

The risk categorisation system would be based on a set of parameters for which different scoring methodologies could apply, such as binary questions, hierarchical criteria (Absent / Basic / Effective / Comprehensive), or sampling as currently proposed in Article 47a of the latest compromise text. The procedure could, if relevant, also integrate a combination of these solutions:

1) Binary methodology

Scoring based on simple yes/no questions related to the core architecture of the service. A yes/no response could be given a +/- score respectively (meaning more/less risk) resulting in a final score.

For example: Does the service have a livestream system? A "yes" will represent a positive value whereas a "no" will represent a negative value. A higher total score implies a higher risk. The potential range of possible scores will then be divided into 4 categories accordingly.

2) Multi-class scoring with 4 hierarchical criteria methodology

Scoring based on the extent to which policies and functionalities are in place to address the risk of child sexual abuse material being disseminated or grooming practices taking place on the service. Scoring could be based on 4 levels: Absent / Basic / Effective / Comprehensive. Each level would represent a score from 4 (high risk) to 1 (low risk) resulting in a final score.

For example:

Functionalities for notifying/ flagging Online Child Sexual Abuse

- Absent/very limited
  - The platform lacks or has very limited functionalities enabling users to flag and report online child sexual abuse in the sense that they are ineffective.
- Basic
  - The platform provides basic flagging tools, but accessibility and age-appropriateness need improvement. Need to enhance the accessibility of reporting tools to ensure users can easily locate and utilise them. Might improve the interface to make reporting tools more age-appropriate, especially for younger users.
- Effective
  - The reporting tools are effective, offering users a straightforward and age-appropriate means to flag and report online child sexual abuse. Ensure reporting tools are easily accessible within the platform, promoting quick and efficient reporting. Maintain an age-appropriate interface for reporting tools, catering to users of all age groups. Provide ongoing educational resources to keep users informed about recognising and reporting online child sexual abuse.
- Comprehensive
  - The platform excels by providing comprehensive tools for notification of online child sexual abuse, ensuring a swift and effective response to flagged content.
  - Collaborate with external organisations and law enforcement agencies to enhance the efficiency of the reporting and response process.
  - Regularly update reporting tools based on user feedback and technological advancements.

3) Sampling methodology

For certain evaluation criteria and parameters, it may also be possible to implement a compartmentalization system based on the sampling and analysis of specific data. This is a relatively technical method which, if used, requires specification of the types of data to be sampled, the collection procedures, the compartmentalization mechanisms, and so on. It is the same type of method that is defined in article 47a and article 7(2) of the latest compromise text proposed under the former Spanish Presidency (12611/23).

For example:

This type of methodology could be used to analyze CSAM data itself in order to assess the risk, or metadata such as data relating to user accounts to assess the frequency of occurrence of certain risk-relevant points. This can be useful for calculating, for example, the extent of anonymous account use, the use of fake accounts, the frequency of account changes, the use of VPNs, etc. It would therefore be necessary to collect the data from providers, process it, and conclude trends based on these samples.

Use of Anonymous Account:

- *frequent use of anonymous accounts*
  - *Over 25% of accounts lack identifiable information.*
- *Moderate instances of anonymous accounts.*
  - *From* 26 to 60% *of anonymous accounts.*
- *Minimal or no use of anonymous accounts*
  - *Majority of account have identifiable information* (from 61% to 100%).