DOT europe
Digital • Online • Tech

December 2022

DOT Europe position paper on the Regulation laying down rules to prevent and combat child sexual abuse

ISSUE PAPER 1

Risk assessment and mitigation

Risk assessment and mitigation
Detection orders
Technologies
Scope
EU Centre and global reporting framework

The proposal requires hosting service providers and interpersonal communication providers to conduct a risk assessment in order to evaluate the risk of their respective services being used for the purpose of Child Sexual Abuse (CSA). Once this assessment is completed, providers are required to adopt mitigation measures to tackle the risks identified in their assessment. We welcome this approach which provides an incentive for providers to identify the specific CSA risks associated with their services and address them effectively.

DOT Europe first notes that the DSA also requires very large online platforms (VLOPs) to identify, analyse, assess and mitigate systemic risks their services pose to the protection of children. Recital 8 of the draft Regulation considers this Regulation as *lex specialis* in relation to the generally applicable framework set out in the DSA. Nevertheless, VLOPs would need to consider how their approach to compliance can be aligned under these two instruments in order to avoid potential overlaps.

Furthermore, as highlighted by the European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS)[1], the proposal as it currently stands does not provide a clear understanding of what constitutes "significant risk". This is however a fundamental element to underlying both risk assessment and risk mitigation. Additionally, the "effectiveness of mitigation measures" (Recital 18) and the criteria under which Coordinating Authorities will determine the “appreciable extent” to which a service is used for the dissemination of CSAM (Article 7) needs clarification. The proposal lacks precision on how Coordinating Authorities will move to a detection order; it remains unclear whether services having successfully passed the risk assessment and mitigation phase can avoid detection orders or not.

The voluntary use of technologies to prevent online CSA and to detect child sexual abuse material (CSAM) with appropriate safeguards is an important mitigation measure. While the text of the proposal indicates that hosting services may continue to voluntarily detect CSAM, providers of interpersonal communications services (ICSs) would only be allowed to use detection technology upon failing a risk assessment and receiving a detection order, which will constitute a legal basis to process communications data, traffic data or location data a limited time.

[1] EDPB-EDPS, Joint Opinion 4/2022 on the Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse, Adopted on 28 July 2022

The fact that the proposed text does not recognise voluntary efforts, including prevention measures, as possible risk mitigation measures for ICSs is a **step backwards** compared to the current voluntary regime under the interim ePrivacy Directive derogation[2], which allows for the voluntary detection of CSAM subject to safeguards. The text also does not acknowledge the volume and quality of work already undertaken voluntarily, which has led to important innovation in the fight against CSA. This work involves significant historical and ongoing investment, to the benefit of consumers and law enforcement.

DOT Europe believes that the important voluntary work carried out by ICS providers should still be possible to carry out once the interim ePrivacy Directive Derogation[3] reaches its sunset clause. We therefore believe that **the Regulation should include an express legal basis** for ICS providers to use technology for the voluntary detection of CSAM and for the processing of communications metadata for the purposes of prevention and detection, all of this with appropriate safeguards.

Voluntary detection measures and other mitigations conducted by ICSs could be supervised by the relevant Coordinating Authority and Data Protection Authority. This would provide the appropriate safeguards and avoid creating a scenario where companies cannot effectively mitigate the risk in their services because they cannot implement voluntary measures which they then are asked to implement later by a detection order, creating a gap in companies' ability to detect and thereby give perpetrators room to continue to spread this heinous content online.

Last but definitely not least, the proposal introduces possible new obligations on age verification/assurance for ICS providers in need to mitigate solicitation risks on their service. This could de facto result in an obligation for many individual services to each develop their own age verification solutions, without being able to rely on a clearly agreed industry approach on this important issue. As this is an area where the Commission expects to make progress as part of the Better Internet for Kids (BIK+) strategy[4] and the code of conduct[5] that will emanate from it, the proposed legislation should avoid legislating on a contested and evolving matter and instead **defer to the BIK+ strategy for the development of appropriate age verification approach and mechanisms**. The proposal should encourage a multistakeholder discussion in the context of the upcoming code to develop effective solutions. Indeed, the proposal's provisions could de facto result in an obligation for many services to each develop their own age verification solutions, without being able to rely on a clearly agreed industry approach on this important issue. The debate surrounding age verification is technologically complex and will have broader consequences fundamental rights, data, security and privacy. In order to consider all possible outcomes, prior to introducing any new obligation entailing a specific group of

[2] Regulation (EU) 2021/1232 of the European Parliament and of the Council of 14 July 2021 on a temporary derogation from certain provisions of Directive 2002/58/EC as regards the use of technologies by providers of number-independent interpersonal communications services for the processing of personal and other data for the purpose of combating online child sexual abuse, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021R1232&qid=1667993362830
[3] Regulation (EU) 2021/1232 of the European Parliament and of the Council of 14 July 2021 on a temporary derogation from certain provisions of Directive 2002/58/EC as regards the use of technologies by providers of number-independent interpersonal communications services for the processing of personal and other data for the purpose of combating online child sexual abuse
[4] A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2022:212:FIN
[5] "A comprehensive EU code of conduct on age-appropriate design", see point 5.1 of the BIK+ Strategy

digital service providers, DOT Europe sees the value in a **multistakeholder discussion** to develop sustainable technical and technological approaches.

Examples of current practices

Discord's proactive content moderation includes:

- Scanning images uploaded to the platform using industry-standard PhotoDNA to detect matches to known child sexual abuse material.
- In the course of Discord's proactive content moderation efforts, when Discord discovers data suggesting that a user is engaging in illegal activity or violating their policies, Discord investigates their networks, their activity on Discord, and their messages to proactively detect accomplices and determine whether violations have occurred.
- Discord also strives to inform its users about best practices to set up a safe server and a safe account through information disseminated in their Safety Center and Policy and Safety Blog.

Meta developed a multi-layered approach to safety on its private messaging services, which focuses on: working to prevent abuse from happening in the first place; giving people more controls to help them stay safe and; responding to reports on potential harm. This is bespoke and takes a slightly different, but complimentary, approach to how the company keeps people safe on its public services - like Facebook - using signals from those public spaces to help keep private messaging safe.

Prevention is at the core of the work Meta does to protect safety, with its main objectives being:

- Preventing people from being exposed to harmful contact or harmful content;
- Preventing potential offenders from contacting potential victims;
- Preventing offenders from contacting each other;
- Contribute to the broader public safety effort, including by continuing to respond to law enforcement requests and providing reports to NCMEC.

TikTok works with families and youth safety advocates on a holistic approach to keeping children safe. TikTok works to educate families on the safeguards available to help them manage their TikTok experience at the Youth Portal, Safety Centre, and in-app videos. In addition to detection and the use of PhotoDNA, TikTok has taken an upstream, safety by design approach to prevention: it does not permit off-platform images or videos to be sent via direct messages. TikTok also disabled direct messaging for registered accounts under 16.

Google and **Youtube** have invested heavily in fighting child sexual abuse and exploitation online and developed cutting edge technology to deter, detect, remove and report offences on their platforms.

- They use hashing technology to identify, remove, and report copies of known images. This technology allows them to automatically detect illegal content that matches the content that has previously been identified as CSAM.
- In 2015, YouTube engineers created CSAI Match, world-leading technology that can be used to scan and identify uploaded videos that contain known CSAM. YouTube makes this technology available to other platforms and NGOs free-of-charge.
- In 2018, Google engineers created the Content Safety API which, based on machine learning technology, helps identify content that is likely to contain abuse, at scale, more quickly by helping reviewers to prioritise the most likely CSAM content for review.

Google and YouTube support others fulfilling their commitments in the fight against CSAM by offering our cutting-edge technology free-of-charge for qualifying organisations to make their operations better, faster and safer.

Microsoft takes a range of actions to protect children across its consumer hosted services, including the use of technology to detect online child sexual exploitation and abuse and the provision of family safety controls for Windows and its Xbox gaming services. Microsoft's Digital Safety Content Report covers actions that Microsoft has taken in relation to child sexual exploitation and abuse imagery.

Snap Inc has built extra protections for teenagers from the beginning. On Snapchat, teens have to be mutual friends before they can start communicating with each other, by default, friend lists are private and teens are not allowed to have public profiles. Snap recently introduced its Family Center, a tool designed to offer parents, caregivers and other trusted adults insight into who their teens are communicating with on the app, while at the same time protecting teens' privacy, autonomy and growing independence. Adults can view their teens' friends' lists, who they communicated with in the last 7 days and report to Snap accounts that may be of concern to them. Snap employs a multi-pronged strategy, which includes investing in and deploying the latest technologies, leading, collaborating and engaging with others in industry and across sectors and raising awareness among and educating its community about online risks.

Twitter has developed a #ThereIsHelp prompt for child sexual exploitation (CSE) which is designed to help users when looking up terms associated with various manifestations of CSE and to provide them with information about Twitter's zero tolerance policy encouraging users to report such content. The prompt is also intended to connect users to local partners that offer intervention and prevention programs in the national languages. This feature is currently available in 6 countries worldwide (Germany, India, Indonesia, Philippines, Taiwan and Thailand) and in 7 languages.

DOT Europe's recommendations

- In order to have a workable and coherent text, the proposal should not duplicate risk assessments, where a provider is subject more than one regime. For example, the DSA also requires very large online platforms (VLOPs) to assess and mitigate systemic risks their services pose to the protection of children. To avoid any risk of duplication with the DSA requirement for VLOPs to assess and mitigate systemic risks their services pose to the protection of children, one risk assessment should satisfy both obligations, taking the DSA as a baseline.
- The proposal needs to provide a clear understanding of what constitutes "significant risk" and "the likelihood that the service is used to an appreciable extent" (Recital 21). Coherence of understanding on this concept among all Competent Authorities is essential.
- We suggest including a recommendation to clarify the language in the text around the "effectiveness of mitigation measures" and the criteria under which Coordinating Authorities will determine the "appreciable extent" to which a service is used for the dissemination of CSAM. Currently it remains unclear whether services having successfully passed the risk assessment and mitigation phase can avoid detection orders.
- Voluntary efforts to fight CSA should be expressly authorised for all providers in scope of the Regulation. The Regulation should explicitly recognise voluntary efforts as part of the package of risk mitigation strategies available to all service providers under Article 4.
- In addition, the Regulation should create an express legal basis for ICS providers to process communications metadata for the purposes of prevention and detection, with appropriate safeguards. This would ensure continuity with the principles of the interim ePrivacy Directive derogation rather than focusing solely on mandated detection as the basis for detection. The Regulation fails to recognize the importance of prevention, when relevant, and existing voluntary efforts. The proposal should also incentivise the development and use of new detection technologies by permitting them as a mitigation measure. Such technologies play a vital role in detecting new CSAM and expanding the volume of high quality hashes to the benefit of both industry and law enforcement.
- Regarding the risk mitigation measures, Recital 17 clarifies that service providers remain free to assess their respective CSA risks and select which technology and risk mitigations are appropriate for their services. Language of Recital 17 should be reflected in the actual text of the proposal, which should ensure the mitigation measures include prevention, detection and/or other measures that meet the requirements of Article 4.
- Similarly, Recital 18 and Article 4 provide some degree of flexibility when it comes to considering the mitigation measures suggested by the proposed text. DOT Europe underlines the need for the Regulation to maintain the possibility for providers to take into account relevant differences, including between content types and services, when selecting appropriate and proportionate mitigation measures. This will enable providers to ensure mitigation measures are carefully tailored, in line with the intent of Article 4(2). Additionally, the Regulation should require Coordinating Authorities to take these differences into account in evaluating the mitigation measures a provider has taken on its services.
- DOT Europe strongly encourages a multi-stakeholder coordinated discussion via the foreseen age-appropriate design code to first identify the best way forward and accordingly develop sustainable technical and technological approaches before any obligation is laid down in the Regulation.