CDT Europe Feedback on the Belgian Presidency's proposal on the Regulation on Child Sexual Abuse Material.

31 May 2024

CDT Europe has been engaged in the development of the *Regulation laying down rules to prevent and combat child sexual abuse*, counting on our over 30 years of legal, policy and technical expertise to advise co-legislators. We recognise the importance of addressing the many harms caused by child sexual abuse material (CSAM) and the difficulties of trying to reach a common approach to advance the negotiations. However, the latest proposal by the Belgian Presidency does not resolve the underlying problems stemming from the Member States' obligations under Article 7 and 8 of the Charter of Fundamental Rights. In addition, it creates a number of inconsistencies with existing EU legislation, particularly the General Data Protection Regulation (GDPR) and the Digital Services Act (DSA).

Upload moderation undermines fundamental rights

The Presidency's proposal relies on upload moderation technologies for the detection of known and unknown CSAM. New Article 10a creates an obligation for providers of interpersonal communications services to *"install and operate technologies to detect, prior to transmission, the dissemination of known child sexual abuse material or of new child sexual abuse material."*

While the terminology may vary, the filtering functionalities of upload moderation technologies are fundamentally identical to those of client-side scanning technologies. Hence, the concerns regarding the disproportionate impact to fundamental rights and the associated security vulnerabilities that have been widely documented by the international privacy and cybersecurity community are equally applicable to both types of technology.

Threat to freedom of expression

CDT Europe has repeatedly raised concerns about the scanning methods under consideration in the context of the CSAM proposal. Scanning and evaluating all communications for potentially illegal imagery poses a serious threat to individuals' freedoms of opinion, expression, and association. These filtering obligations impose a disproportionate burden on freedom of expression, treating every post or communication as a potential legal violation. Universal filtering using a technology that, as discussed below, is inherently unreliable also greatly increases the chance that an individual's speech will be erroneously classified as unlawful and expose that individual to scrutiny by law enforcement. Fear of such consequences creates a dangerous chilling effect on lawful, protected speech.

Threat to user's online privacy

We strongly reject the Presidency's decision to maintain end-to-end encrypted services within the scope of the detection orders, in complete disregard of the recommendations to refrain from doing so by civil society, academia, the United Nations High Commissioner for Human Rights, and the European Data Protection Supervisor and Board, amongst others. The proposed changes in the language describing providers' obligations regarding encryption protection are of particular concern. In earlier versions, the text explicitly stated that the regulation aimed to uphold the integrity of encryption without weakening, undermining, or circumventing it. However, the current text states only that providers are not obliged to prohibit, break or decrypt end-to-end encrypted communications in order to comply with their obligations under the CSAM regulation. This omission, coupled with the reporting requirements in the Belgian Presidency's proposal, is a tacit acknowledgment that the proposed technology undermines the fundamental principle of strong end-to-end encryption: that in a system secured by end-to-end encryption, only the sender and the intended recipients can access the contents that are encrypted. Therefore, once again, we call on the Member States to reconsider this approach.

Machine learning technologies and delayed reporting approach to detect unknown CSAM

The Presidency proposes that machine learning technologies be used to identify unknown CSAM (images that are likely to be CSAM but that have not been previously determined to be such). These technologies are inherently unreliable. Indeed the proposal concedes as much by proposing to establish a system of "delayed reporting" to mitigate the rate of false reports inherently associated with such technologies. The Presidency's approach entails that providers will only be able to report potential new CSAM to the EU Center after specific content has been flagged twice by the technologies or once by a user. In addition, the material should undergo pseudonymization prior to human verification at the EU Center. Once the material is determined to be not-manifestly unfounded, the provider is required to resubmit the report including all the identifiable personal data that was initially pseudonymized.

The way in which this system is supposed to operate in practice requires further clarification from the Presidency. Particularly, the Presidency must clarify how the proposed approach interacts with the data protection obligations of service providers, who under GDPR qualify as controllers. Controllers must be aware of the data they control in order to execute their GDPR obligations. Questions arise in regards to the requirement that the first hit be stored away without service providers getting knowledge of or control over that information until a second hit is received. It is unclear how the service provider, as controller, can be expected to operate this technology and store the hit in a secure manner necessary for such sensitive content while not being made aware of it. The information that a particular image has received one hit that is

associated with the account of a specific user can be stored on the client. However this "one hit" will more likely be stored with the service provider because it is more durable when stored there. If, for example, the user gets a new cell phone, the hit can be associated with their account when stored with the provider, but not when stored on the client. As a result, the provider will have this data about the user account and could ostensibly share the fact that one hit has been reached by any user.

Relying on consent as a legal basis for scanning content data through upload moderation.

CDT Europe has serious doubts about the feasibility of relying on consent to authorise scanning of content to detect CSAM, as is contemplated in the Belgian Presidency's proposal. The GDPR sets a high standard for consent by imposing stringent conditions for it to be considered valid as a legal basis for processing of personal data. Article 4(11) of the GDPR defines consent as: "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." The conditions for valid consent as a legal basis for personal data processing are set out under Article 7 GDPR and have been interpreted at length by the EDPB in its dedicated opinion providing concrete guidance on the necessary requirements controllers must fulfil in order to rely on user's consent for their data processing operations. A basic read of the opinion shows that the proposal to rely on a user's consent to enable generalised scanning of images and videos via upload moderation and/or client side scanning would be **incompatible with the GDPR.**

Freely given

The term "free" requires genuine autonomy and authority for individuals over their data. Consent could not be considered "freely given" when individuals lack genuine choice, feel compelled or pressured into giving consent, or face adverse outcomes for refusing it. The new proposal makes the ability of users to share images and videos conditional on the user consenting to the said materials to be scanned for CSAM upon upload. Should the user refuse consent, they will not be able to share images and videos online and therefore not allowed to make full use of the services offered by the service providers. In essence, refusing to give consent will result in negative consequences for the user, as they will not be able to access the full range of services offered by a said platform or provider, unlike those who do consent.

In addition, it is easy to imagine that the inability to share images and videos on platforms specifically designed for that purpose, which are very popular among younger people, will put undue pressure on users to give consent in order to be part of a community and engage in social activities typical of their age group. In other words, peer pressure might compel some users to

consent to upload moderation and client-side scanning without giving proper consideration to the consequences to their fundamental rights online.

Moreover, recital 43 of the GDPR states that "(i)n order to ensure that consent is freely given, consent should not provide a valid legal ground for the processing of personal data in a specific case ***where there is a clear imbalance between the data subject and the controller,*** in particular ***where the controller is a public authority*** and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation." If consent is to be truly informed and specific, users will have to be made aware that their images will be processed by law enforcement authorities, the coordinating authorities and the EU Center, potentially leading to undue pressure on users to comply fearing negative inferences being drawn by law enforcement and the other authorities from the refusal to consent.

Ability to withdraw consent

Moreover, the requirement that consent be "freely given" entails that users must be able to **withdraw consent** as easily as they give it, at any point, **without consequences.** This means that any data collected and processed on the basis of consent must be immediately deleted when consent is withdrawn:

> *EDPB opinion, para 117. As a general rule, if consent is withdrawn, all data processing operations that were based on consent and took place before the withdrawal of consent - and in accordance with the GDPR - remain lawful, however, the controller must stop the processing actions concerned. If there is no other lawful basis justifying the processing (e.g. further storage) of the data, they should be deleted by the controller.*

Consequently, any identifiable data, images or indicators stored in the various databases under this regulation will have to be extracted and permanently erased, which may impair the functioning and accuracy of the technologies and databases and hinder further prevention objectives under the proposed regulation. Furthermore, consent withdrawals could compromise the effectiveness of the proposed delayed reporting for the detection of unknown CSAM. For instance, a user whose content has been flagged once and stored away for up to 12 months may withdraw consent before a second hit is received and a formal report can be submitted to the EU Center. As a result, the opportunity to collect a second hit is lost because the data flagged by the system in the first hit will have to be deleted following the user's withdrawal of the legal basis for processing under GDPR.

In addition, to be able to rely on consent the controller will have to demonstrate that it is possible to refuse or withdraw consent without detriment or clear disadvantage for those who withdraw consent.

> *EDPB, para 48: If a controller is able to show that a service includes the possibility to withdraw consent without any negative consequences e.g. without the performance of the service being downgraded to the detriment of the user, this may serve to show that the consent was given freely.*

As mentioned above, the inability of users to share images and videos online after they withdraw consent has a detrimental impact on the user. It is itself sufficient to determine that consent is not a valid basis to render the far-reaching processing proposed under the CSAM proposal lawful.

Consent of minors

When it comes to assessing the appropriateness of relying on children's consent in the context of the CSAM proposal **special considerations must be given to the vulnerability of children.** The GDPR includes **special protections of children and their personal data**, as children are assumed to be 'less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data' (Recital 38). Hence, the assumption of the Presidency's proposal that children who have reached the age of digital consent can indeed give 'free' consent to personal data processing under the CSAM proposal, is questionable. Even though children who have reached the age of digital consent might display stronger cognitive skills, they are at the same time more prone to peer pressure and impulsive decision making. **As they are otherwise prohibited from accessing one of the key functions of such services, i.e. sharing photos and videos, children will inherently pivot towards consenting to the scanning measures**, as they want to enjoy the full scope of communication services, regardless of whether such scanning will have potential adverse impacts in the future.

In this context it is also questionable **whether their consent is *informed* as they will most likely not understand the possible far-reaching impact their consent can have on their live**s, such as self-incrimination and forced disclosure of sexual activity or even sexual orientation. In particular for cases of unknown CSAM, it is very likely that the technology will flag lawful nudity or consensual sexual material as potential CSAM, as the technology will not be able to understand the context in which this imagery is sent.. As neither the provider, nor the EU Centre or the coordinating authority will have sufficient information to assess the context and hence whether the sharing was consensual, this imagery will be referred to law enforcement for investigation.

Lastly, the clear incentive for children to consent to the processing of their personal data in order to access the image/video sharing function, combined with the lack of ability to foresee the various adverse impacts their consenting might bring in the future, **might even impact the**

fairness of the personal data processing altogether. Similar challenges might arise if parents provide consent on behalf of their children, if the children have not reached the age of digital consent yet. Also the parents' consent might not be valid as they will lack an understanding of the vast adverse impacts the scanning might have on their children, as pointed out above.

Effectiveness of consent

Given the strict requirements outlined above, Member States must question the effectiveness of this approach altogether. If the proposed legal framework were to be implemented, it is difficult to imagine users voluntarily and knowingly agreeing to relinquish their privacy to service providers and law enforcement agencies. Moreover, it seems unlikely that criminals and child abusers would willingly consent to online surveillance as they engage in criminal activities. Ultimately, the risk to fundamental rights much outweighs the potential benefits of the measures and obligations imposed on service providers.

Interference with the Digital Services Act

The proposal for delayed reporting of unknown CSAM raises questions about the effect given to user notices of illegal content, which would fall under Article 16 of the DSA Notice and Action Mechanism. As a first point, Article 16 obliges the provider to only inform the user of the outcome of their notification. Under this new proposal, should the notification concern CSAM material, there would be an additional obligation for the provider to inform the EU Centre. Alongside this, if the notice from a user fulfils the requirements of Article 16(2) of the DSA in which a provider of a hosting service is able to identify the illegality of the relevant content without a detailed legal examination, providers are considered to have obtained *'actual knowledge*', therefore bringing into scope Article 6 of the DSA -the conditional intermediary liability principle. Specifically, the DSA states that "*in order to benefit from the exemption from liability for hosting services, the provider should, upon obtaining actual knowledge or awareness of illegal activities or illegal content, act expeditiously to remove or to disable access to that conten*t". A question is therefore raised about how the Presidency's proposal for delayed reporting and authorisation from the Coordinating Authority for removal will operate in tandem with this existing regime given that providers are highly unlikely to risk their liability protections.

No general monitoring obligations

Article 1(3)(b) of the Presidency's proposal states that the Regulation does not affect the rules laid down in the E-Commerce Directive nor the DSA. Recital 31 further states that "*this Regulation should not be understood as affecting the requirements regarding removal orders or the rules on no general monitoring or active fact-finding obligations set out in (the DSA)*." Indeed, Article 8 DSA is clear that intermediary services (i.e. platforms and interpersonal

communication services) are explicitly protected by the “no-general monitoring obligation” and therefore cannot actively seek facts or circumstances indicating illegal activity. The proposal on detection orders, and arguably several mitigation obligations under Chapter II of the proposal, would fundamentally undermine this principle and conflict with the no-general monitoring obligation established in the E-Commerce Directive, and subsequently reaffirmed by the DSA.

Recital 30 DSA states that 'providers of intermediary services should not be, neither de jure, nor de facto, subject to a monitoring obligation with respect to obligations of a general nature.' The Presidency argues that the CSAM proposal does not impose a general monitoring obligation and that detection orders would only be applicable to services categorised as high-risk. However, the risk-categorization methodology proposed by the Presidency would result in most of the largest platforms and services being categorised as high-risk and maintained within the scope of detection orders, subjecting hundreds of millions of users to a de facto general monitoring obligation under the flawed pretext of user consent.

Therefore, we urge Member States to reconsider their position and reject the latest Presidency’s proposal.