26 May 2022

Briefing on Key Concerns Relating to a Proposal for Regulation laying down the Rules to Prevent and Combat Child Sexual Abuse (CSAM)

Further to the publication of the proposed CSAM Regulation (the Regulation) by the European Commission on 11 May 2022, CDT Europe provides this briefing as an overview of some of the most pressing human rights concerns in the proposal. Child sexual abuse is an abhorrent crime that demands a robust and holistic response. It is essential to recall that security of communications and privacy protections are of vital importance to the protection of childs' rights, and therefore also need to be protected in the context of this Regulation.

The draft Regulation has such a broad scope, applying to both hosting services and providers of internal communications, and mandating scanning of all communications, that, in its current draft, it very concerningly amounts to giving a mass surveillance mandate to law enforcement authorities. The measures envisaged by the Proposal would constitute a disproportionate interference with the fundamental rights to respect for private life, the right to free expression and association and data protection of all users of electronic communications services.[1]

Key Recommendations

In the pursuit of drafting the Regulation, legislators must be cautious not to undermine already well established safeguards and principles aimed at keeping checks on surveillance and privacy. Most importantly, the regulation should avoid:

- providing a **mass surveillance** mandate to the EU Centre, Europol, nor compelling private actors to carry out such surveillance.
- the very real risk of the Regulation being **struck down** by the European Court of Justice due to non-compliance with existing **EU privacy laws and fundamental rights.**
- introducing a law that would compromise end-to-end encryption, given how important **privacy safeguards** are, not only to protect children and minors, but also journalists, activists and civil society actors.
- **undermining the data minimisation principle** and the existing rights infringements that Europol has perpetuated in its processing of huge datasets.

General Monitoring & Mass Surveillance Risk

The draft Regulation creates a filtering mandate in a number of different ways. Art. 7.1 and Art. 10 outright require hosting services and providers of interpersonal communications services to

[1] Including social media services, cloud-based file storage, instant messaging platforms, applications, and Internet access providers

scan all communications for an open-ended set of potential illegal content. Articles 3 and 4 require providers to conduct risk assessments and implement mitigation measures against the risk that their services will be abused for online child sexual abuse, and Art. 7 empowers the new Coordinating Authority to seek Art. 7 detection orders against services that, in the Coordinating Authority's view, do not adequately mitigate these risks. Moreover, Art. 49 empowers the new EU Centre to conduct searching and scanning of providers' services, using its own filtering technologies, to determine whether providers are adequately complying with detection and removal orders. In short, the entire framework of the Regulation presumes that filtering technologies and general monitoring obligations are core to achieving its purpose.

Art. 7 poses a threat to the rule of law, as it provides the Coordinating Authority with a mandate to request either Courts or 'administrative bodies' in the given member state to issue a detection order. Providers will also face a legal obligation to report "any information indicating potential online child sexual abuse" discovered via these detection orders or other quasi-voluntary filtering the provider undertakes, leaving it to private companies to determine when the detailed information required in Article 13 should be transferred to the EU Centre. This effectively establishes a system of mass surveillance, through general monitoring. But mass surveillance is incompatible with human rights and surveillance of any kind should be subject to checks and balances, and in particular should be overseen by a tribunal within the meaning of Art. 47 of the EU Charter of Fundamental Rights.

Risks of Abuse of Power

As drafted, the Regulation would provide a real risk of abuse of power by government and law enforcement agencies. It could allow for government agencies to order a hosting service or a provider of interpersonal communications services to hand over communications of investigative journalists or human rights defenders on the pretence of searching for CSAM. Even though paragraph four of Art. 7 lists criteria that should be met in the advance of the issuance of such a detection order, this list cannot compensate for the potential lack of independence of the coordinating authority or administrative authority, nor can it supersede the role of a Court. Furthermore, the criteria listed under Art. 7 (4) are not robust enough to meet the standard of proportionality as set out by both the European Court of Human Rights and the European Court of Justice. It will be at the discretion of EU member states as to which authority shall be the national coordinating authority for the purpose of this Regulation. International human rights law is clear that only independent Courts should determine whether content is illegal; however the draft Regulation under Art. 14 would give an EU-wide blocking mandate on content to the coordinating authorities in member states irrespective of whether it is a Court or not.

End-to-End Encryption

The proposed Regulation would enable the issuance of orders that compel online platforms, including those offering end-to-end encrypted messaging, to scan users' content to detect both

known and unknown CSA images and to detect conversations and behaviour regarding solicitation of children. They could be compelled to report such data to public authorities and delete them from their platforms. These requirements are fundamentally incompatible with end-to-end encrypted messaging because platforms that offer such services cannot access communications content to detect such information. This has been confirmed by experts around the world who showed that the belief that scanning communications is consistent with end-to-end encrypted systems is mythical. Other such experts issued a detailed report describing how, in particular, client-side scanning, "can fail, can be evaded, and can be abused."

As previously outlined by CDT and other civil society actors, the ability to communicate securely via encryption is essential to the protection of democracy and human rights, that importantly include children's rights. It is particularly important for at-risk children to have the ability to communicate confidentially and securely using end-to-end encryption. The Draft Regulation does not sufficiently account for the benefit to human rights that encryption provides. Instead, it indicates, for example in Article 7 para 4(b), that when determining whether to request a detection order, the Coordinating Authority shall consider the "negative consequences for the rights and legitimate interests of all parties affected…" without explaining that such negative consequences may accrue to all users of a system that must compromise encryption in order to meet the requirements of a detection order.

Perils of Processing Large Datasets by Law Enforcement

The processing of data about individuals in an EU law enforcement database can have deep consequences on those involved. This is why Europol[2] is already subject to restrictions on how it processes data. According to Article 28 of the Europol Regulation, Europol is responsible for compliance with the principle of data minimisation for all personal data processed by it. Europol has recently been subject to an inquiry by the European Data Protection Supervisor (EDPS) that raised concerns that the Agency was not compliant with these very restrictions due to the manner in which it was found to be processing large data sets. Given that the draft Regulation would provide for a very broad monitoring mandate, and given the access that Europol would then have to these huge sets of data, there is a real cause for concern that the draft Regulation would undermine existing safeguards and exacerbate the current problems regarding data minimisation and proportionate processing of personal data by Europol.[3] Furthermore, the ECJ has ruled that information and evidence obtained by means of the general and indiscriminate retention of traffic

[2] The EU Centre has a separate legal personality, but will be housed and share resources with Europol, and is mandated (Art. 45(4)) to further pass on data obtained from scanning communications to Europol directly.

[3] Furthermore, in the Schrems II case, the ECJ determined that the potential for government access to bulk collection of data resulted in a failure to afford data subjects the privacy rights provided by the General Data Protection Regulation (GDPR and the Charter of Fundamental Rights).

and data in breach of EU law,[4] so there would even be a question as to whether evidence obtained through this system would be permissible in Court.

Rights of Individual Users

The draft Regulation refers rightly to the right to information of victims, however it neglects to refer to the right of all users to information regarding the scanning of their communications.[5] Given the hugely broad mandate for scanning all communications, it cannot be argued that it would be 'necessary and proportionate' under the public interest clause in GDPR Art. 6. Indeed the EDPS[6] has been clear that it is not just *any* processing carried out to combat illegal activities that may be considered as lawful under Article 6(1)of GDPR.

In addition, the scanning and evaluation of all communications poses a dire threat to individuals' freedoms of opinion, expression and association. Filtering can be understood as a form of prior restraint or prior censorship that exposes every utterance to review and approval before it can be spoken.[7] Filtering obligations are a disproportionate burden on freedom of expression and treat every post or communication as a potential violation of law. This also greatly increases the chance that an individual's speech will be erroneously classified as unprotected, and expose that individual to scrutiny by law enforcement. This risk of exposure and scrutiny can have a chilling effect on lawful, protected speech. Moreover, mandated filtering systems, like other systems of prior censorship, can be nearly impossible for individuals to perceive in operation, in direct conflict with the principle of legality that individuals must be able to know what restrictions will be applied to their speech. This risk is compounded by the draft Regulation's instructions for the EU Centre to develop its own "indicators" to identify potential new CSAM or solicitation activity (Art. 44). Such indicators, which the Centre will use to evaluate compliance with detection and removal orders (Art. 49), risk becoming opaque technical encodings of legal prohibitions against speech that are illegible to the people bound by them, in contravention of the right to freedom of expression.

Technical Challenges of Using Filtering to Scan Large Data Sets

The technical proposal at the core of the draft Regulation also presents significant human rights risks. Article 44 states that the EU Centre will create, maintain and operate a database of 'indicators' of online child sexual abuse. It is proposed that the relevant indicators consisting of digital identifiers should enable the detection of known and new CSAM and the solicitation of children. This may refer to the concept that the Centre should hash known CSAM images and

[4] Judgments in Case C-623/17, Privacy International, and in Joined Cases C-511/18, La Quadrature du Net and Others, C-512/18, French Data Network and Others, and C-520/18, Ordre des barreaux francophones et germanophone and Others; Joined Cases C-293/12 and C-594/12

[5] See Art. 20 of the draft Regulation.

[6] See Opinion on the proposed temporary derogation from Directive 2002/58/EC for the purpose of combatting child sexual abuse online

[7] See Llansó, No Amount of "AI" Will Solve Filtering's Prior Restraint Problem,Big Data & Society (April 23, 2020) https://journals.sagepub.com/doi/full/10.1177/2053951720920686.

videos and distribute those hashes to providers, in the manner of the United States' National Center for Missing and Exploited Children (NCMEC). (However, Article 46(2) specifies that the Centre may only make access to these hashes available to providers during the period of time they are covered by a detection or removal order, in significant contrast to the voluntary and persistent access to its hash database that NCMEC enables.) It is less clear what is intended for the "indicators" of new CSAM or solicitation content; it is not possible to hash new CSAM before it has been detected for the first time, which implies that these so-called "digital identifiers" (Art. 44(2)(a)) necessarily include other technical approaches that do not identify specific violative content. Rather, Article 44 likely means to instruct the Centre to develop machine-learning classifiers that will predict the likelihood that new content is CSAM or a solicitation communication. The use of such classifiers can pose significant risks of both overbroad and underinclusive flagging of content, risks which are particularly borne by already vulnerable and marginalised populations.[8] This risk may be exacerbated if, as Article 44(3) instructs, the classifiers are trained "solely on the basis of the child sexual abuse material and the solicitation of children identified as such by the Coordinating Authorities or the courts," as machine learning classifiers typically need to be trained on data sets that include both violating and non-violating content, such as art, educational materials, adult pornography, and innocuous images of children, in order to learn how to differentiate between that lawful content and CSAM.[9]

Conclusion

The draft Regulation's provisions are wide-ranging, and would compel companies to scan all communications, including encrypted communications, which could then be accessed by law enforcement. This would constitute an interference with fundamental rights which could affect practically the entire European population and so cannot possibly be considered necessary nor proportionate. The Pegasus scandal proved, once again, how our right to communicate securely underpins the pillars of democracy, including press freedom, the presumption of innocence, privacy and freedom of expression and association. That is why CDT is appealing to EU lawmakers to urgently revise the approach.

For further information, contact CDT Europe Director Iverna McGowan at imcgowan@cdt.org

[8] See CDT, Mixed Messages: The Limits of Automated Social Media Content Analysis (Nov. 2017) https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/; CDT, Do You See What I See? The Capabilities and Limitations of Automated Multimedia Content Analysis (May 2021) https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/.

[9] It is worth noting that the finite volume of known CSAM on which computer vision machine learning can be trained presents a natural limit to the efficacy of the technique (e.g. the machine learning algorithm can't be improved without more real examples). Because these "indicators" developed by the Centre will be made available to providers who are covered by a detection or removal order (Art. 46) and will be used by the Centre in evaluating whether providers are, at the Coordinating Authority's request, adequately mitigating risk or implementing a detection or removal order (Art. 49(1(b)), they will become the de facto technical and legal standard, across providers, for designating new CSAM and solicitation communications, which will exponentially amplify the risks to privacy and free expression that they pose.