Open Letter from Public Interest Technologists in relation to the European Commission's proposed Regulation on Child Sexual Abuse (CSA)

10 May 2023

Dear Members of the European Parliament and Representatives of the EU Member States:

We, the undersigned public interest technologists and academics, are writing to express our concerns as leading experts in our respective fields regarding the serious risks that the European Commission's proposed Regulation on Child Sexual Abuse (CSA) poses to the privacy and security of all communications, and to the overall health of the internet and information ecosystem.

Child sexual abuse is one of the most horrific crimes imaginable. It requires a robust all-of-society response. Yet, the proposal on the table would jeopardise the safety of everyone, including some of the most vulnerable: children. In a democracy, we have laws that are safeguards against mass surveillance because we understand the harm to all of society it would bring.

A major concern is that the European Commission's documentation regarding the proposal states in several places that experts have been consulted, and yet there is no transparency regarding who these experts are and what were their inputs to the process. Furthermore, the conclusions of the experts who appear to have been consulted are at odds with computer science and current research findings with regard to the impact of the draft Regulation on end-to-end encryption, and with regard to the effectiveness, efficiency, and indeed the very feasibility of the technologies proposed. In sum, the formal consultation which led to the first proposal appears to be divorced from an evidence-based scientific approach.

The Current Proposal Breaks End-To-End Encryption

As currently written, the draft law would effectively oblige all hosting services and providers of interpersonal communications to scan all content. It would do so in two ways. First, the risk assessments are so broadly framed that hosting services and providers would be compelled to scan all communications to avoid detection orders and/or liability. This is because the proposal stipulates that providers of applications will face legal consequences unless they report "*any information indicating potential online child sexual abuse*." Second, the proposal allows for detection orders to be launched covering a service (e.g., an entire social network such as Facebook or a messaging service such as Whatsapp or Signal), implicating thousands if not more individuals without reasonable suspicion of wrongdoing. It would mean that individuals using these services could be subject to bulk scanning of all of their private messages and images (effectively conducting surveillance of their communication) just because they happen to use a certain service, irrespective of whether

or not there is any reason to believe that they have been involved in criminal activity. Private companies would then determine what to hand over to law enforcement.

This client-side scanning would by its nature create serious security and privacy risks for all of society while the assistance it can provide for law enforcement is at best problematic. There are also multiple ways in which client-side scanning can fail, be evaded, and be abused. Filtering technologies are notoriously imprecise and will result in the misidentification of innocuous content as child exploitation, and fail to identify real cases of child exploitation because evasion is easy and those willing to exchange CSAM will change their inputs so as not to be detected. Providers will face enormous incentives to err on the side of over-reporting, resulting in potentially disastrous results for people mistakenly identified as engaging in the sexual abuse of children (false positives).

The risk of false negatives, whereby cases of abuse are missed, should also be an issue of profound concern. As the EU Data Protection Board explains: “Recital 26 of the Proposal places not only the choice of detection technologies but also of the technical measures to protect the confidentiality of communications, such as encryption, under a caveat that this technological choice must meet the requirements of the proposed Regulation, (i.e., it must enable detection). This supports the notion gained from Articles 8(3) and 10(2) of the Proposal that a provider cannot refuse the execution of a detection order based on technical impossibility.”

Traditional backdoors or client-side scanning built on privacy-preserving computation using secure enclaves or through cryptography (e.g., homomorphic encryption), are complex methods proposed for content moderation in end-to-end encrypted environments. However, these technical approaches share one crucial thing in common: they attempt to detect whether the content of messages is prohibited, and, no matter the methods used, this effectively inserts pervasive monitoring into the confidential communications of billions of people, putting everyone worldwide, including nation states and children, at risk.

The potential outcomes of this imprecise scanning of all communications content create significant privacy risks: Either services will use these filters to automatically detect, remove, and report content, jeopardising the privacy and even safety of those falsely accused, or services will implement human review of flagged content, exposing people's private communications to scrutiny by third parties.

Our assessment and consensus as technical experts is that each of these proposed content moderation methods for end-to-end encrypted communications introduces an inherent vulnerability, breaks the privacy and security promises to users, and weakens strong cryptographic standards universally, which amounts to breaking end-to-end encryption. This is why it is crucial that no proposed EU legislation should ever oblige providers to break end-to-end encryption or engage in general monitoring of communications.

The Shortcomings of Current Technological Solutions

The technology regarding the probabilistic detection of new child sexual abuse material or of "grooming" communications is far from being reliable enough to be safely deployed in the high-risk and sensitive context of child protection online. Furthermore, the nature of this technology, which provides estimates of the likelihood that novel content meets some pre-defined criteria, inherently requires some tolerance for errors. Indeed, even the technology for known CSAM carries a high risk of error and can be easily circumvented by determined parties, which should be considered.

Perceptual hashing - shortcomings with Photo DNA

The proposal suggests that the EU Centre should hash known CSAM images and videos and distribute those hashes to providers in the manner of the United States National Center for Missing and Exploited Children (NCMEC). Whilst perceptual hashing can be effective in identifying known images and videos, even those that feature slight variations which may occur either naturally or from attempts to circumvent detection, it still has many serious limitations which have not been sufficiently considered by the proposal. The error rates cited by the European Commission in relation to Photo DNA technology, for example, are at odds with research findings on this topic. Furthermore, a freedom of information request in relation to the European Commission's assertion of a 99.9% accuracy rate has been revealed to be based on unverified industry claims.

Machine Learning for Unknown CSAM and Grooming

It is even less clear what is intended for the "indicators" of new CSAM or solicitation content in the draft proposal; 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" 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.

Moreover, Article 44(3) describes an incoherent process for developing a classifier of any type, instructing that the classifiers are to be 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." Machine learning classifiers typically need to be trained on data sets that include both violating and non-violating content in order to learn how to differentiate between the two. For the Centre to even attempt to develop a CSAM classifier, it would need to amass a vast data set that included art, educational materials, adult pornography, private communications, and innocuous images of children; this would further

compound the threat to individuals' privacy from generalised data collection by a law enforcement agency.

Indeed, automated models are repeatedly shown to fail in situations they have never encountered in their design or training. A tool trained on clear and distinct images may struggle when presented with blurry images or photos taken through a window or filter. Machine learning tools are also vulnerable to adversarial attacks, where distortions or perturbations in the image that are imperceptible to the human eye nevertheless interfere with the tool's ability to classify the image correctly. The robustness of the tools underlying automated content analysis—or the ability to not be fooled by minor distortions in data—is a constant and unsolved problem.

We urge that laws proposed in such sensitive subject matter areas rely on scientific and evidence-based approaches. As outlined above, this has not been the case with this proposal. We therefore urgently call on EU decision makers to consider our collective and serious concerns with regard to the proposal.

Signatories

Prof. Dr. Srdjan Čapkun - Professor, Swiss Federal Institute of Technology (ETH Zurich)

Prof. Dr. Peter Y. A. Ryan - Professor of Applied Security, University of Luxembourg

Prof. Dr. Holger Karl - Chair of Internet-Technology and Softwarization, Personal Capacity

Prof. Dr. Anja Lehmann - Hasso-Plattner-Institute, University of Potsdam

Prof. Dr. Dominik Herrmann - Professor, Otto-Friedrich-Universität Bamberg, Germany

Prof. Dr. Cas Cremers - Faculty, CISPA Helmholtz Center for Information Security

Prof. Dr.-Ing. Tibor Jager - Professor, University of Wuppertal

Prof. Dr.-Ing. Paul Rösler - Professor, Friedrich-Alexander-Universität Erlangen-Nürnberg

Prof. Martin Albrecht – Chair of Cryptography, King's College London

Prof. Carmela Troncoso - Professor, École Polytechnique Fédérale de Lausanne (EPFL)

Prof. Bart Preneel, KU Leuven

Dr. Elizabeth Farries - Director, University College Dublin, Centre for Digital Policy

Dr. Hamed Haddadi - Associate Professor and Chief Scientist, Imperial College London and Brave Software

Dr. Dan Bogdanov - Security and Cryptography Researcher, Personal capacity

Dr. Kris Shrishak - Cryptography Researcher and Senior Fellow, Irish Council for Civil Liberties (ICCL)

Douwe Korff - Emeritus Professor of International Law, London Metropolitan University

Nigel P. Smart - Professor/Fellow of IACR, KU Leuven

Claudia Diaz - Associate Professor, KU Leuven.

Aleksandra Kuczerawy - Senior Research Fellow, KU Leuven Centre for IT & IP Law

Pierre Dewitte - Researcher in Law, KU Leuven Centre for IT & IP Law

Stefano Calzavara - Associate Professor, Università Ca' Foscari Venezia

Rikke Bjerg Jensen - Reader, Information Security Group, Royal Holloway University of London

Konstantinos Komaitis, Internet policy expert and author, personal capacity

Sofía Celi - Cryptography Researcher, Brave

Raphael Robert - CEO, Phoenix R&D

Julian Mair - Head of Operations & Project Development, Phoenix R&D

Elina Eickstädt - Spokesperson, Chaos Computer Club

Runa Sandvik - Founder, Granitt

Michael Kreil - Data Journalist, Personal capacity

Alec Muffett - Consultant Security Researcher, Personal capacity

Karthikeyan Bhargavan - Researcher, Personal capacity

Giancarlo Pellegrino - Faculty, CISPA Helmholtz Center for Information Security

This Open Letter has been coordinated by the Centre for Democracy & Technology Europe. For further information, and to be put in touch with signatories, please contact eupress@cdt.org.