European Digital Rights

# Submission to the European Commission's post-adoption consultation on Combating online child sexual abuse – amending temporary derogation from certain provisions of Directive 2002/58/EC

*8 February 2024*

EDRi is Europe's biggest network of digital human rights groups, representing over fifty national and international civil society organisations. We advocate for a safe and secure internet for all.

As we emphasised in our advocacy on the original temporary derogation from certain provisions of Directive 2002/58/EC (aka "the temporary ePrivacy derogation"), any limitation on fundamental rights must be strictly necessary and proportionate for a legitimate aim. Any law to implement this limitation must also contain clear and effective safeguards.

The general and indiscriminate scanning of people's private messages, regardless of whether or not this is done on a "voluntary" basis, entails a particularly serious interference with the fundamental rights to privacy, data protection, free expression and more, of a very high number of people. In fact, when such measures are only voluntary, rather than mandatory under EU law, this may undermine claims that the rules are necessary.

That's because if the co-legislators agree that certain measures are strictly necessary for a legitimate aim, then rules in Article 52(1) of the EU Charter of Fundamental Rights would imply that such measures cannot be left to the discretion of commercial entities. To be legally coherent, measures which aim to balance the important aim of protecting children from CSA with the wide-ranging intrusion into the digital private lives of people who are not suspected of any wrongdoing cannot be left to the whims of Facebook, Microsoft, Apple, Thorn or any other tech company.

**Data from the implementation report on the original ePrivacy derogation**

To date, the European Commission has not provided sufficient evidence that the voluntary measures under the interim ePrivacy derogation, which are now proposed for extension, are necessary and proportionate, given the serious negative consequences for the protection of several fundamental rights.

The Commission's implementation report on the temporary derogation explains that as EU figures were not provided by the companies who performed scanning under the derogation, "it is not possible to draw any conclusions for the purposes of this report" (§ 2.1.1. and 2.1.2.).[1]

Furthermore, the implementation report explains that providers relied on a crude metric of the number of accounts reinstated after a complaint as a rough approximation of the accuracy of their technology (§ 2.1.6). This does not, however, provide useful information on the technical accuracy of the technology.

We do not, therefore, have robust or credible evidence in support of a link between the mass scanning of digital communications and an improvement in child safety online. We also lack evidence about how scanning measures translate to investigations, arrests and/or convictions.

Whilst EU Member States were supposed to provide at least some of this information under the ePrivacy derogation, the implementation report explains that this information is not known (§ 2.1.1.). Instead, the report provides a total number of reports of online child sexual abuse in some Member States, without stating the proportion that come from scanning, nor the proportion that led to action (table 6). Where data about the number of children identified are provided, the report acknowledges that these data are not detailed enough to know to what extent this correlates with scanning practices (§ 2.2.2.).

Whilst the report claims that "it can be inferred from the data that a significant number of victims have been identified with the help of voluntary reporting in accordance with the Interim Regulation," (§ 2.2.2.), this assertion is not backed up by the data provided. With the exception of two convictions in Estonia, the provision of data on perpetrators of CSA who have been convicted includes all CSA cases for a given year, including offline offenses, and cases which pre-date the temporary ePrivacy derogation (§ 2.2.3.). This does not provide any insight into the extent to which online scanning contributed to any of those convictions.

However, the EDRi network's position paper on the draft CSA Regulation, 'A Safe Internet for All', provides evidence of how mass scanning has led to high rates of false accusations of abuse, and likely unlawful data retention by the police in Ireland, to give one example.[2]

1 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2023%3A0797%3AFIN. References throughout.
2 https://edri.org/wp-content/uploads/2022/10/EDRi-Position-Paper-CSAR.pdf, pp. 53-56.

**Discussion of error rates in the implementation report**

The unverified claim by Hany Farid that PhotoDNA has fewer than one in 50 billion false positives, which was given as 'evidence' for the accuracy of scanning tools in the impact assessment for the long-term CSA Regulation (2022/0155(COD)), is repeated in the implementation report (§ 2.3.1.). However, PhotoDNA users such as Cloudflare have show that it is relatively easy to force the system to generate false positives, casting doubt on the veracity of Farid's claim.[3]

The Commission's report also makes the (again, unverified) claim that Thorn's scanning technologies produce only 0.1% false positives (§ 2.3.2.). However, as revealed in January 2024 by journalist Alexander Fanta (Follow The Money), when set at this precision rate, **false negatives (i.e. when actual CSAM is missed by the scanning) occur at a rate of one in five pieces of content (20%)**.[4] The oft-repeated claim of Thorn's 99.9% precision is therefore deeply misleading when trying to assess the real-world accuracy of scanning technology.

For 2021, LinkedIn provides information about actual false-positive error rates which conforms to the reporting requirements of Article 3(1)(g)(vii) of the ePrivacy derogation (as the only provider to do so). LinkedIn's use of automated detection technology flagged 75 items, but human review could only confirm 31 items as CSAM. This is an **actual false-positive error rate of 59%** for a mix of known and unknown CSAM. The LinkedIn sample is small (75 items), but large enough to conclude (in terms of statistical significance) that the automated detection tools **do not have an accuracy of 99.9%**, as claimed by the Commission, based on unverified industry information.

The inaccuracy of automated detection will increase even more when machine-learning indicators are used to predict the presence of grooming. Yet the figures and accuracy statistics put forward by the Commission for the draft CSA Regulation inflate the accuracy, and downplay the false positives and negatives, entailed by the use of these scanning technologies. More analysis of the inherent limitation of mass AI-based scanning can be found in EDRi's aforementioned position paper.

**Political integrity and good administration**

As reported by Follow The Money, the European Commission may also have buried the inconvenient truth about the accuracy of scanning technologies by rejecting legitimate freedom of information (FOI) requests on the subject. This refusal is part of a systematic refusal of the Commission's DG HOME unit to disclose relevant documents related to the CSA Regulation, as seen in two recent decisions against the Commission by the European Ombudsman.[5]

3 https://blog.cloudflare.com/the-csam-scanning-tool/
4 https://www.ftm.eu/articles/ashton-kutcher-s-anti-childabuse-software-below-par
5 https://www.ombudsman.europa.eu/en/case/en/65012;
https://www.ombudsman.europa.eu/en/decision/en/176658

With the proposed extension of the ePrivacy derogation – as with the draft CSA Regulation - it seems that the European Commission is systematically failing to meet their obligations to good public administration, transparency and political integrity. This can undermine public trust in the lawmaking process and must be urgently rectified, especially when issues as important as protecting children from CSA, and everyone's privacy rights and data protection rights, are on the line.

**Concerns about claims of proportionality and effectiveness**

The ePrivacy derogation implementation report makes several concluding claims which lack any basis in the data provided there, or anywhere else publicly. The first claim is that "there are no indications that the derogation is not proportionate" (§ 3.). The second is that "it can be concluded that voluntary reporting contributed significantly to the protection of a large number of children, including from ongoing abuse, and it appears that the Interim Regulation is effective" (§ 3.).

Under EU fundamental rights law, proportionality must assess not only the benefits of a rights-restricting measure, but also possible risks and threats. Yet the implementation report admits that it has not collected data on whether scanning measures were the least-privacy invasive measure. The threats to the rights of innocent internet users have been framed in the implementation report only in terms of wrongful content take-down or account closure. Without having undertaken a full proportionality assessment and analysis of the negative consequences of mass scanning practices, it is inappropriate for the report to claim that there are no indications of the temporary ePrivacy derogation not being proportionate.

What's more, the data provided are not detailed or dis-aggregated enough to make a connection with the effectiveness of the ePrivacy derogation. The "voluntary reporting" that the implementation report mentions is not limited to reports of material/content identified through voluntary scanning. Therefore, it is neither possible, nor credible, to draw a conclusion about whether the ePrivacy derogation is effective.

In fact, it is deeply concerning that after stating throughout the report that the data provided cannot allow conclusions to be drawn, the Commission's implementation report nevertheless ends by claiming that the temporary ePrivacy derogation appears to be effective.