

# Balancing Cybersecurity, Privacy and Child Protection in detecting Child Sexual Abuse in End-to-End Encrypted Spaces

## Advocacy paper

As online platforms transition to end-to-end encryption (E2EE), user privacy and cybersecurity are being strengthened against threats like identity theft and scams. However, this shift brings a critical unintended consequence: a reduced ability to detect and stop the proliferation of child sexual abuse online. This reduced ability has an impact on children's risk of experiencing online sexual abuse and on law enforcement authorities' capacity to fight child sexual abuse and save children. Reporting of child sexual abuse by Facebook has [dropped by 40%](#) since the implementation of end-to-end encryption on the platform in the summer of 2024 - this led to 6.9 million fewer reports in 2024.

**Excluding E2EE from the scope of detection of child sexual abuse online would create huge loopholes that will be exploited by perpetrators.**

In the fight against child sexual abuse, online platforms must not enable safe havens which can be exploited by sexual offenders to commit their crimes with impunity. We need to build and implement solutions to balance privacy and child protection and ensure that E2EE spaces cannot serve as a shield allowing offenders to commit their crimes.

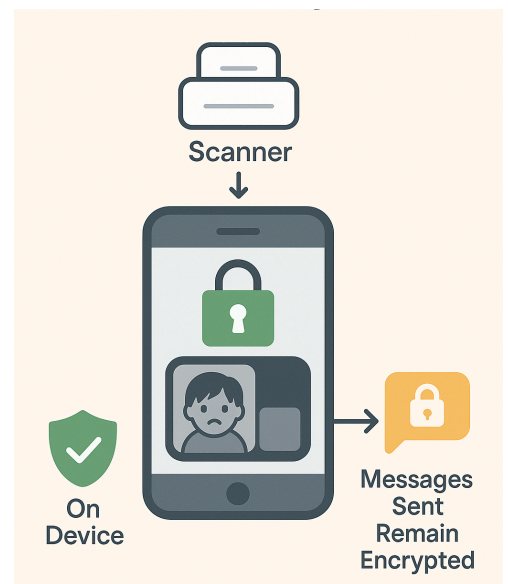
Child sexual abuse constitutes a violation of multiple children's rights, as outlined in the UN Convention on the Rights of the Child, especially their right to protection from all forms of abuse and to privacy. Finding balanced solutions to detect child sexual abuse in E2EE spaces is part of the responsibility of EU Member States to guarantee children's rights online.

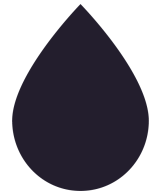
**Solutions to balance cybersecurity, privacy and child protection exist.** This paper explains them and calls on decision-makers to allow for their deployment and use in the context of the proposed Regulation to fight child sexual abuse.

**E2EE and child protection can go hand in hand — with the right safeguards in place, we can create digital spaces that are both secure and safe for children.**

## Techniques to safely detect child sexual abuse in E2EE services exist

Contrary to popular belief, E2EE does not make child sexual abuse (CSA) detection impossible. Detection tools can operate without compromising the security and





privacy of encryption. These tools enable action before content is sent or reaches the recipient (before it is encrypted), often without requiring access to the message content itself. Techniques that can be used to detect child sexual abuse on E2EE platforms are -

- **Matching Techniques:** Compare hashed images to known CSAM databases.
- **Content Classifiers:** AI detects CSAM patterns — crucial for unreported cases.
- **Behavioural Detection:** Identifies grooming via interaction patterns.
- **Metadata Analysis:** Already standard for spam and scams; flags suspicious activity.
- **Online Social Network Analysis:** Maps offender networks.
- **Message Franking:** Enables reporting without breaking encryption.
- **Multi-Indicator Reporting:** Combines signals to ensure high-precision alerts.

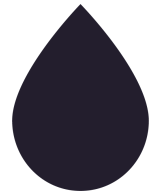
Techniques can take different approaches to detection based on the reporting and escalation of the material (some techniques can be used for both approaches). These approaches fall broadly into two categories:

1. **Prevention-based approaches**, client-side solutions that act directly on a user's device. They do not share the information with a third party (such as law enforcement or platform moderators) - the intervention is only between the users by, for example, **blocking** harmful content before it is sent, or sending automated messages to **warn** potential victims or **deter** potential offenders, directing victims to relevant support services or options to report. This approach offers stronger privacy protection, but limits opportunities for prosecution and victim safeguarding. Hash matching techniques, content classifiers, behavioural detection, metadata analysis, online social network analysis, message franking, safeguarding and multi-indicator reporting can be deployed for prevention.
2. **Reporting-based approaches**, where the detected material is reported to the platform moderators and potentially to law enforcement. While more effective for legal enforcement and facilitating the rescue of child victims, these approaches carry additional privacy and security considerations demanding robust oversight and transparency. Matching techniques, content classifiers, metadata analysis, online social network analysis and message franking can be deployed for reporting.

Choosing one approach over another involves **trade-offs** that must be considered when making the decision on the most appropriate solution for a platform.

## **Detection in E2EE spaces should be guided by key principles and risk mitigation measures to balance cybersecurity, privacy and child protection**

Addressing CSA in E2EE environments requires a multi-faceted approach — **no single solution** fits all platforms or offender profiles. Techniques must be tailored to the specific design and functionality of each platform and proportional to the risk.



A combination of techniques is necessary to match the complexity and scale of the problem. We know that offender profiles and techniques in this space are evolving at a rapid pace and we need technological solutions that can be scaled, implemented, and have the ability to evolve at the same pace.

Effective implementation of these tools must be guided by **key principles**: measures must be **ethically** sound, **proportionate**, **transparent**, **feasible**, and **monitored** by independent experts. Stakeholder engagement — including input from child rights advocates, cybersecurity experts, and law enforcement — is also essential to ensure legitimacy and public trust.

The cybersecurity and privacy concerns tied to CSA detection tools are **no more significant** than those already accepted for a wide array of technologies deployed in E2EE, including to counter spam, scams, malware, or terrorist content moderation. With proper oversight and safeguards, privacy and protection can — and must — coexist.

We acknowledge that risks exist as with any technology. In the case of the detection of child sexual abuse, the main risks are:

- **Data Breaches**
- **Poisoning Attacks** (covert altering of a hash database to achieve another objective, like surveillance)
- **Model Misuse or Manipulation** (models are retrained to find things other than CSAM)
- **Using hashes to search and view CSAM** (offenders hacking the database to find CSAM)

The likelihood of most of the above risks to occur are extremely low in practice. In addition, there are ways to mitigate each of the cybersecurity risks<sup>1</sup> that may arise when implementing CSA detection, including through:

### 1. Cybersecurity Expert Involvement to Assess Detection Tools

Cross-disciplinary task forces with cybersecurity professionals, child rights experts, technologists, law enforcement, and relevant authorities must be established. These bodies should assess and audit CSA detection tools to ensure alignment with privacy and child protection goals. Youth advisory panels should inform this process. Enhanced transparency and oversight will help build trust in the tools.

---

<sup>1</sup> Risk: Data Breaches; Mitigation: Client-side detection, secure processes, limiting data retention, and anonymisation mitigate exposure.

Risk: Poisoning Attacks; Mitigation: Legislation, oversight, automated checks and third-party verification prevents database tampering.

Risk: Model Misuse or manipulation; Mitigation: Adversaries have easier alternatives than misusing CSA tools. Automated checks can further mitigate the risk.

Risk: Reverse Engineering hashes to view CSAM; Mitigation: Secure hash databases, use private hashing techniques. Offenders have much easier ways of finding CSAM.



## 2. Detection on Child Devices Only

Limiting CSA detection and blocking to verified child accounts or devices significantly reduces risk exposure. This client-side approach ensures privacy for adult users while safeguarding young users from grooming and CSAM.

However, an important **trade-off** to consider is that this approach does not prevent sharing of CSAM among adult users. A limited, prevention approach to detection on adult devices could be envisaged, leading to blocking or deterring the sharing of CSAM.

## 3. Automated Third-Party Hash Database Verification

Hash databases should be subject to regular integrity checks and statistical testing by independent bodies. Changes to the database should be logged and verified by a third party. This protects against manipulation and fosters public confidence.

## 4. Independent and Accountable Moderation Bodies

Neutral third-party moderators — like the proposed EU Centre — must be introduced to vet CSAM reports before escalating to law enforcement. This builds user trust and ensures accurate, proportionate responses. Transparency and oversight from international organisations are key.

## 5. Alternative Child-Safe E2EE Platforms

We must pilot open-source, child-safe E2EE services that integrate CSA detection and blocking tools on a small scale. This allows testing, refinement, and demonstration of feasibility, building trust in the tools.

However, this solution has an important **trade-off** as it would not prevent sharing of CSAM among adult users, which is equally problematic.

## Child Protection and Privacy Must Advance Together

The mitigation measures presented in this Paper involve trade-offs for privacy or child protection that must be assessed on a case-by-case basis for each platform using the key principles mentioned above. **It is no longer acceptable to frame E2EE and child protection as opposing forces.** Tools and frameworks already exist that preserve E2EE while enabling the detection of CSAM and grooming behavior.

With informed policy and expert collaboration, we can ensure that E2EE environments are not safe havens for abusers. Rather, they can be transformed into safer spaces that protect both the privacy and safety of all users — especially children.

**Policy must evolve to protect the most vulnerable without compromising the rights of all. The tools are ready, solutions are possible — what's needed now is the will to act.**

