ECLAG
Ending Child Sexual Abuse Online

October 2023

Myth Busting & Facts on the proposed Regulation on Child Sexual Abuse: addressing the privacy concerns with data and facts

This document provides an overview of the main myths and concerns that have been used to reprove the proposed Regulation to combat and prevent child sexual abuse (CSA). This paper also brings together a set of facts to address these concerns, aiming to draw a research- and data-based assessment of the proposed Regulation.

The Myths

Myth 1: The Regulation would unleash mass surveillance and 'read' all messages

This claim builds on misinterpretation of the process established by the proposed Regulation and misunderstanding of the technology at hand.

Under the proposed Regulation, detection would only happen after a thorough process of risk assessment, reviews, multiple checks and a court order, making it **virtually impossible for the detection technology to be misused** (see further below Fact 2: The Regulation will establish strong oversight and ensure privacy).

Indeed, the proposed Regulation mandates all online service providers to assess the risk that their service is being used for the distribution of child sexual abuse material (CSAM) or grooming of children and to adopt preventive measures (such as safe design or user reporting) to mitigate this risk. If, despite these measures, there is still evidence of a significant risk, a national court or independent authority will determine **on a case by case basis** the **necessity and proportionality** of the use of specific detection tools before mandating their use through a detection order, taking into consideration the impact on users' privacy (see also Myth 2: The slippery slope of technology). Detection would only happen in a specific part of the service (e.g. specific types of channels or specific users) which presented significant risk of being used to abuse children and for a limited period of time. Under the proposed Regulation, the technology deployed must be **reliable** with the least margin of error and must be the **least intrusive** in terms of the impact on the users' rights. It cannot extract any other information than strictly necessary to detect CSAM.

In addition, this claim builds on **unfounded fears and on a misunderstanding of the technology** at hand. Detection technology is built for the sole purpose of detecting CSAM and only recognise grooming patterns indicating this. It doesn't "read" or understand messages. It looks for matches. It either compares digital fingerprints of images via hash-matching to a database of known and verified CSAM - or it uses an AI-based machine learning (i.e. classifier) to flag content that is suspected to be CSAM, which is trained to make the difference between CSAM and innocent imagery. Even in this last case, the content would then undergo a multi-step process to get verified as CSAM, i.e., human review. In short, **it operates in the same way as a metal detector, which can only detect metal and does not know or flag anything else** that may be underground.

Myth 2: The slippery slope of technology - i.e. governments could use it to surveil political opponents or human rights defenders

Detection technology is **built for the sole purpose of detecting CSAM** or to only recognise grooming patterns. It is extremely difficult and costly to repurpose and abuse CSA detection technology.

Detection technology has been **deployed for over a decade** and is built to only detect CSA to a high level of accuracy. Over [200 companies](#) have already deployed advanced technologies to safely detect, report and eliminate child sexual abuse.

In fact, **the Regulation puts in place safeguards that would effectively prevent misuse of detection technologies** (see further below [Fact 2: The Regulation will establish strong oversight and ensure privacy](#)). Only detection technologies that meet the requirements of the Regulation (in terms notably of efficiency, reliability and scope) and assessed as safe and privacy-preserving by a new independent EU Centre would be allowed.

Detection technologies would only be used:

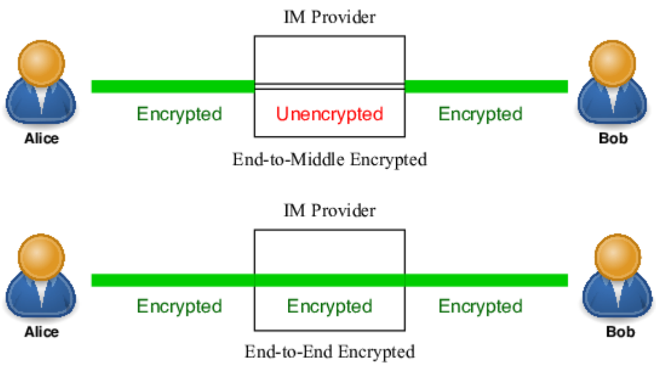
1. In a specific part of the service presenting a high risk of being used to abuse children
2. After mitigation measures failed
3. Upon request of a judicial court
4. With technologies assessed as safe and privacy-preserving by an EU Centre
5. For a limited period of time (see further below [Fact 2: The Regulation will establish strong oversight and ensure privacy](#)).

In addition, the databases of indicators which providers will use to detect CSA (known CSAM, new CSAM or grooming) will be created and maintained by the EU centre itself - not the providers, nor the national law enforcement authorities.

This framework sets a high bar and ensures checks and balances to avoid misuse of detection technology.

Unfortunately, surveillance technology already exists and has already been used by governments through spyware such as Pegasus, whether or not CSA detection tools are deployed.

Myth 3: Undermining end-to-end encryption



In End-to-end encryption (E2EE), only the sender and receiver of encrypted communications have the key to access it. With standard encryption, service providers also hold the key to the encrypted message but can only use it in certain circumstances.

Image source: [ResearchGate](#).

The proposed Regulation does not include any provision on E2EE. The Regulation is **technology neutral,** meaning that it does not require any specific technology to detect child sexual abuse, but sets specific criteria for such technology to be met, including that it ensures the respect for privacy, before it can be ordered to be deployed. This is important to ensure the law can adapt to and include developing technologies.

Public authorities have the obligation to ensure children are protected from sexual abuse **in all environments**, even through the most private forms of personal communication. **Two-thirds of children** who received sexually explicit material online did so through private messaging, mostly on their personal mobile. Predators use the method of **off-platforming**, meaning of moving the conversations with children to E2EE services in order to avoid detection of the abuse. Our societies cannot allow the creation of a black hole where all crimes go undetected.

The technology to detect child sexual abuse in E2EE while respecting privacy and not undermining encryption already exists, using similar technology as for malware (see Myth 4: Client side scanning breaks encryption). WhatsApp, an E2EE service, already deploys advanced technology to detect malware and viruses without affecting E2EE.

Myth 4: Client-side scanning breaks encryption

Client-side scanning consists in **scanning the message before it is sent to the encrypted channel**. Therefore, it does not touch the encryption whatsoever. Client-side scanning can operate on device only or with the support of an external database to ensure the match against CSAM. The EU Centre would ensure that any database used for client-side scanning only contains confirmed CSAM or approved classifiers.

Client-side detection shows promise in filtering CSAM before a message enters an encrypted environment in a privacy preserving manner. **This technology has already been deployed effectively at scale**. This is how, for example, WhatsApp prevents the spread of malicious URLs on its encrypted messaging service without affecting E2EE, and how browsers like Chrome and Edge warn users of malware on https. Recently, Apple launched their 'Sensitive Content Warning' and 'Communication Safety' tool. The tool scans messages locally on children's devices to flag sent and received content containing nudity.

Detecting for CSAM within end-to-end encrypted environments can also be done in a privacy-forward way through homomorphic encryption, multi-party computation or secure enclaves. And, there may be more ways we have yet to discover: it will take a multitude of solutions from industry to tackle the problem so that they can be used by a variety of companies of different sizes and scales. By being technology neutral, the proposed Regulation will encourage **innovation** in this area.

Myth 5: Detecting new CSAM will lead to many false positives

The tools used to combat online child sexual abuse and exploitation have been **used for over a decade**.

Known CSAM, i.e. that which has already been flagged as CSAM and added to a database, is detected using 'hashes' with compare two images and flag (almost) identical matches. Detection technology to detect **new/unknown CSAM** and grooming use 'Classifiers' that are trained on confirmed CSAM, adult pornography and legal images in order to make the difference between CSAM and innocent 'baby in the bathtub' pictures. Companies which deploy them can set the threshold for detection accuracy to be extremely high to avoid false positives.

Once content is flagged by detection technology, **human review -** with analysts trained to identify illegal content **-** will confirm that the content is indeed criminal, ensuring that only criminal material is acted upon by law enforcement authorities. To avoid false positives, specific threshold and accuracy requirements could be set up by the **EU Centre** to ensure that a high standard is met. Under the proposal, the EU centre will assess the reports received to ensure only not manifestly unfounded reports are shared with law enforcement authorities.

Ultimately, false positive rates are a trade-off between precision rates (how much of all the flagged content is CSAM) and recall rates (how much of the CSAM on a platform is being detected). These two rates are adjusted by the technology developers when training them depending on what they want the technology to be more efficient at. In practice, detection methods are tuned to have extremely high precision rates to ensure that all children suffering sexual abuse are effectively protected, justifying the extremely low risk of false positives.

Many currently deployed technologies, such as road radars, have false positives. Nevertheless, societies have opted to deploy them as the objective of reducing road accidents was considered important enough to accept a low number of false positives. There is no zero-error technology, but protecting children from online abuse is a legitimate and crucial objective proportional to the use of detection technologies.

Myth 6: A new Regulation is not necessary, extending the interim regulation is enough.

The interim Regulation was adopted in 2021 derogating some provisions of the e-Privacy directive to allow number-independent interpersonal communications service (NI-ICS), such as webmail or chat services, to continue detecting child sexual abuse material in their platforms on a voluntary basis. The extension of the temporary derogation alone will not be sufficient to address the scale of the situation. For instance, this extension **would not apply to online service providers who start operating after 2 August 2021** and would not cover private communications. This would leave out many of the apps children use nowadays and where they are exposed to sexual abuse. Moreover, tackling child sexual abuse should not rely solely on the own initiative of online service providers. Transparency and accountability are key for the fight against child sexual abuse online. Children have the right to be protected equally on all the online platforms they use.

Even with a new Regulation in place, there must be **a clear legal basis for voluntary detection** to ensure **no gaps** in child protection. This could happen, for instance, in the case an online service provider has to wait to “fail” the risk assessment and mitigation process to receive a detection order in order to have a legal basis to detect. **Voluntary detection is a risk mitigation** tool and is complementary to the detection orders system proposed by the Regulation. Online service providers cannot identify the risks of child sexual abuse on their services without detecting them, to understand the scale of the issue on their platforms. Moreover, to avoid gaps in child protection and ensure the long-term feasibility of the proposed Regulation, mandatory and voluntary detection must coexist.

The Facts

Fact 1: Detection is effective and essential in preventing the spread of CSAM. Public reporting will never be sufficient

The effectiveness of detection is evidenced by the fact that pausing detection correlates directly with falling statistics on the total amount of CSAM being reported and removed. This was visible during the legislative gap in 2021 when Facebook was forced to stop detecting in the EU for 10 months resulting in a 58% reduction in CSAM being found and removed.

Public reporting will never be sufficient due to the significant barriers to reporting. Education and awareness about the value of reporting by ‘’bystanders’’ can help improve reporting but will not entirely resolve the under-reporting issue. Child victims are often not likely to report. According to a prevalence study, “**83%** of young people aged 11 to 17 years old who had been sexually assaulted by a peer had **not told anyone**”. Victims may not know their abuse has been recorded, some victims are too young to speak out, older children often don’t report because of shame, fear or threats from the offender not to report.

Data shows that the **proactive detection** of CSAM leads to a substantially higher volume of identified and removed CSAM. In 2021, the 50 INHOPE hotlines processed 928,278 URLs reported by the public, while the IWF, the UK hotline, handled 361,062 CSAM items alone, of which 66% resulted from its proactive search. The Canadian Project Arachnid’s automated web crawling detection tool of known CSAM or close matches processed **158 billion+** images between 2017 and March 2023.

Mandatory company reporting to the NCMEC Cybertipline amounted to 85 million files. And while public reporting is crucial to discover known and previously unknown material, proactive search is able to do so at a much higher scale that meets the **volume of CSAM in circulation**.

Preventative measures, such as the risk assessment and mitigation measures, are crucial to build a digital environment that is safe-by-design for children. However, **prevention measures alone will not stop the proliferation of child sexual abuse online**. Prevention and detection are complementary mechanisms, both play their part in effectively protecting children from re-victimisation and ongoing abuse.

Fact 2: Detecting new CSAM and grooming saves lives

Behind every image and video of child sexual abuse there is a child in danger. Detecting new CSAM and grooming is crucial to stop ongoing abuse and protect children from imminent danger.

At the moment, online service providers voluntarily use detection technologies to find and report child sexual abuse to The National Center for Missing and Exploited Children (NCMEC). NCMEC then refers these reports to the relevant national law enforcement agencies, who can open investigations to arrest the perpetrators. New CSAM detected and reported enable the law enforcement to save children and arrest offenders every day. In the

UK alone, an estimated **1,200 children safeguarded** and 800 suspected child sex offenders arrested on average **every month.**

It is also crucial to prevent re-victimisation. The redistribution of CSAM means that, for victims, the abuse not only stays in their memory but it's re-lived constantly. It is widely acknowledged that the **trauma** of having child sexual abuse material recirculating is extremely damaging, creating difficulties for victims to heal from their abuse due to the ongoing nature of the trauma. **Children exposed to grooming and sexually explicit content report similar levels of trauma symptoms (i.e. clinically diagnosable PTSD) to victims of penetrative offline sexual offences.**

Detecting new CSAM is key to allowing known CSAM to be removed (last year, about 1 million new CSAM were added to the list of 6 million known CSAM).

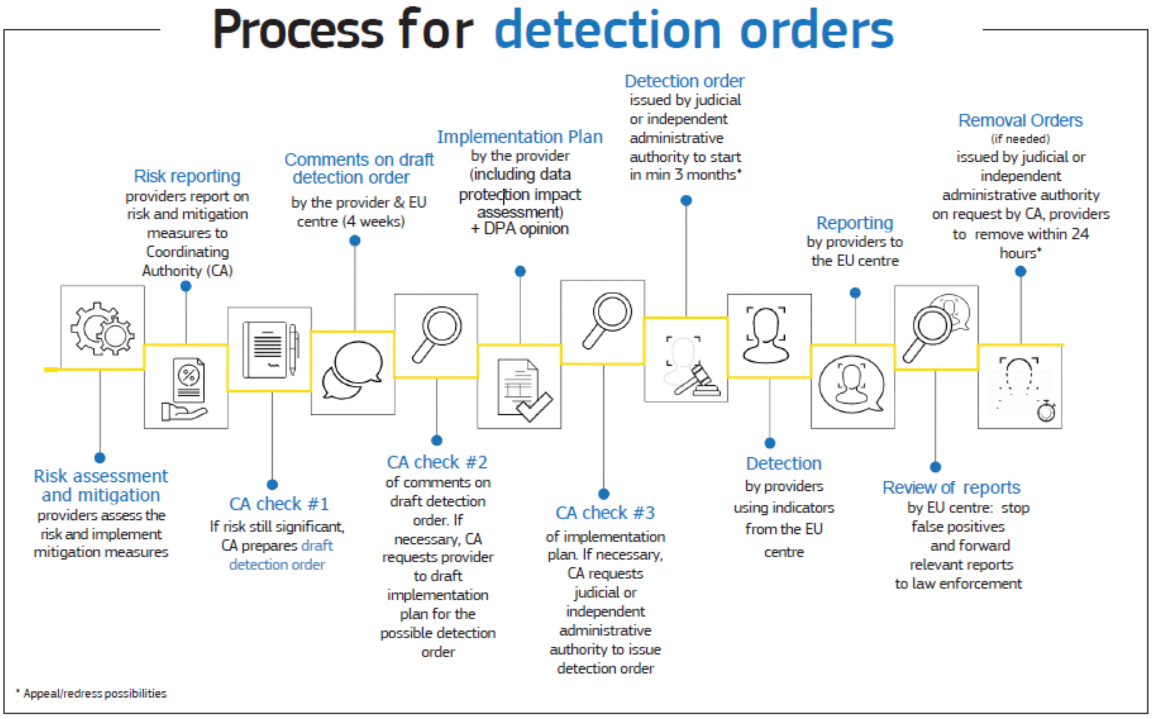
Fact 3: The Regulation will establish strong oversight and ensure privacy

The unsolicited contact of an adult to a child with sexual intent and the dissemination of images and videos depicting the sexual abuse of a child are breaches of the child's right to privacy. This Regulation will ensure **the right to privacy of children** is protected.

For the rest of the users, the Regulation does not allow indiscriminate scanning of private messages. In fact, the Regulation will establish strong **safeguards and a long review process** before any detection is authorised to ensure that **no indiscriminate detection** of illegal material is carried out and to minimise any invasion of privacy. These safeguards include:

1. Detection technologies to be used will have to be authorised and provided by the EU Centre established in the Regulation. Online service providers will not be able to use detection technology that infringes the minimum standards of security and privacy established by the Centre and the Regulation.
2. Detection orders will be necessarily issued by a national judicial or administrative authority, in line with national and EU law on data protection and fundamental rights.
3. National Coordination Authorities will review and give feedback on the risk management and on the implementation plan of a detection order.
4. Data Protection Authorities can also provide recommendations in the detection process.

As shown in the graphic below, **multiple checks**, including from the EU Center and a Data Protection Authority, transparency reporting and oversight, are foreseen to ensure that the detection will conform to EU law, including the GDPR, and respect the privacy of users.



Process for detection orders

Risk assessment and mitigation
providers assess the risk and implement mitigation measures

Risk reporting
providers report on risk and mitigation measures to Coordinating Authority (CA)

CA check #1
If risk still significant, CA prepares draft detection order

Comments on draft detection order
by the provider & EU centre (4 weeks)

CA check #2
of comments on draft detection order. If necessary, CA requests provider to draft implementation plan for the possible detection order

Implementation Plan
by the provider (including data protection impact assessment) + DPA opinion

CA check #3
of implementation plan. If necessary, CA requests judicial or independent administrative authority to issue detection order

Detection order
issued by judicial or independent administrative authority to start in min 3 months*

Detection
by providers using indicators from the EU centre

Reporting
by providers to the EU centre

Review of reports
by EU centre: stop false positives and forward relevant reports to law enforcement

Removal Orders
(if needed)
issued by judicial or independent administrative authority on request by CA, providers to remove within 24 hours*

* Appeal/redress possibilities

Source: European Commission.

One must bear in mind that:

- ➔ All legislation in the EU must comply with other existing laws, including the **GDPR**, which strictly regulates the control and processing of personal data by private companies.
- ➔ Filtering has been accepted by the Court of Justice of the European Union in cases of high accuracy (for example in IP protection).
- ➔ The proposed legislation places the responsibility of balancing fundamental rights **with independent authorities, rather than by individual companies**.
- ➔ Multiple checks will take place to ensure that only illegal material is removed. Any content detected by the technology will be checked to ensure that they indeed constitute illegal material.
- ➔ When signing into a platform, Internet users must consent to the platform's Terms of Services in order to benefit from the use of the platform.
- ➔ The proposed Regulation will mandate **transparency and accountability** of the platforms so that users know when using a platform what it is doing to prevent and remove illegal material.
- ➔ The databases of indicators which will be used by providers to detect each type of CSA (known CSAM, new CSAM or solicitation of children) will be created, maintained and operated by the EU centre itself - not the providers, nor the national law enforcement authorities.

Fact 4: Technology already exists to tackle child abuse while respecting privacy

Technologies already exist that effectively detect CSAM with high accuracy rates. Those include PhotoDNA, YouTube CSAI Match, Facebook's PDQ and TMK+PDQF for known

CSAM and Thorn's Safer Tool, Google's Content Safety API and Facebook's AI Technology, for new/ unknown CSAM and grooming. **Those technologies are already deployed at scale with no issue of misuse or privacy concerns.**

As mentioned above, **client side scanning is already deployed at scale in E2EE** for various legitimate purposes, such as viruses and malwares. Some online service providers, such as Apple, already use it to flag CSAM and grooming conversations in their messaging apps. Detecting CSA in E2EE could be done in the exact same manner, using the same technology (see Myth 5: Client-side scanning breaks encryption and Myth 3: Undermining end-to-end encryption).

The Regulation provides **a framework to double check** that the technology used to detect CSA and grooming will do so in a manner that **minimises any privacy intrusion** through the intervention of experts' opinion and judicial courts, while ensuring that the detection is targeted and effective (see Fact 3: The Regulation will establish strong oversight and ensure privacy) Because of these requirements, companies will have a powerful **incentive** to further develop privacy preserving technologies that can be deployed in any platform, including E2EE environments.

Privacy of all users, including children, is essential. **The right to privacy of children is infringed when pictures and videos of their abuse are shared online** without their consent and when they receive unsolicited contact from adults**. Privacy concerns of child victims of CSA should be equally regarded** by privacy rights organisations and data protection authorities.

Fact 5: Most child sexual abuse occur in private messaging

CSAM and grooming mostly occur through the use of private messaging. **Two-thirds of children** who received sexually explicit material online did so **through private messaging,** mostly on their personal mobile.

Therefore, tools targeting private messaging are key to help detect and remove images and videos and to flag potentially grooming conversations. Detection technologies would not be able to 'read' the messages just to predict the probability of grooming happening in a conversation (see Myth 1: The Regulation would unleash mass surveillance and 'read' all the messages). Detecting CSA in private messages thus plays a crucial role in keeping children safe and disclosure of their abuse on behalf of the victim of sexual abuse.

A common tactic used by perpetrators is called **'off-platforming'**, meaning that perpetrators initiate contacts with children from public platforms, then entice them to move the conversations into applications that use end to end encryption or ones that don't have detection tools so that they can better obtaining CSAM from their victims **undetected**. If we do not include private messaging in the scope of this Regulation, we risk private communications becoming a safe haven for perpetrators to abuse children.

Fact 6: Citizens overwhelmingly back the EU Regulation

In 2023, analysts at the Internet Watch Foundation reviewed 101,988 webpages hosted in the EU containing child sexual abuse material between January and August alone. The EU continues to be the larger hub to host this material, with over 60% of CSAM reported in 2023 being traced to an EU country. Europeans are aware that child sexual abuse is a rising problem in their countries and there is overwhelming support for online service providers to proactively fight against this crime.

The recent Eurobarometer survey shows that 78% of respondents approve the Commission legislative proposal to prevent and combat child sexual abuse and 96% see the ability to detect child abuse equally as important or more important than the right to online privacy. Moreover, between **84% and 89% support that service providers use tools to automatically detect** images and videos of already known CSAM (89%), new images and videos (85%) and grooming (84%), even if those tools may interfere with the privacy of users.

Similarly, a recent ECPAT and NSPCC poll proved that 81% of European respondents support obliging online service providers to detect, report, and remove child sexual abuse online. According to 86% surveyed Europeans, children are increasingly at risk of child sexual abuse and exploitation online, and data reveal that the majority of polled EU citizens see online service providers as one of the most important actors in preventing and protecting children from sexual abuse and exploitation online. **95% say it is key that there are regulations to prevent online child sexual abuse.** These findings, in alignment with the Eurobarometer results, underscore a critical message: European citizens are deeply concerned about child sexual abuse online.

More than half of all Europeans surveyed declare that the issue of child sexual abuse and exploitation online will **influence how they vote at a future election.** There is a clear and urgent demand for decisive action to address this issue. With the European Parliament elections on the horizon, MEPs (Members of the European Parliament) face a duty and a moral responsibility to enact meaningful legislation for child safety online.

More resources

Fact-check: Top 9 claims made on the Regulation to fight Child Sexual Abuse