EDRi

A safe internet for all

Upholding private and secure communications

EDRi network position paper on the "Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse" 2022/0155(COD) (CSA Regulation)

This paper has been co-authored by the following civil society organisations: **Access Now**, **the Asociatia pentru Tehnologie si Internet** (ApTI, Romania), **Bits of Freedo**m (BoF, the Netherlands), **the Chaos Computer Club** (CCC, Germany), **epicenter.works** (Austria), **European Digital Rights** (EDRi), **the Foundation for Information Policy Research** (FIPR), the **Irish Council for Civil Liberties** (ICCL), the **IT-Political Association of Denmark** (IT-Pol) and **Vrijschrift** (the Netherlands).

It builds on the invaluable work of the dozens of other EDRi members with whom we have collaborated, as well as on discussions with many other organisations across various human rights topics, and reflects the position of the EDRi network.

Index

Glossary and acronyms

AI: artificial intelligence;

Age of sexual consent: there is no harmonised age of sexual consent in the EU. Depending on the Member State, it varies from 14 to 17 years. Some EU Member States have specific laws decriminalising consensual sexual acts by older adolescents;

Backdoor: an intentionally built-in mechanism that allows an actor to bypass the security measures of a system in order to gain access to that system or its data;

Charter: the Charter of Fundamental Rights of the European Union 2012/C;

Child: under the EU Child Sexual Abuse Directive 2011/93 and following the UN Convention on the Rights of the Child, a child is a person below the age of 18;
CJEU: the Court of Justice of the European Union;

CSA: refers to child sexual abuse and exploitation, the crime(s) of engaging in sexual acts with or soliciting sex from a child. Note that the exact definition will vary from country to country;

CSS: client-side scanning, a technology that allows on-device analysis of data before being encrypted;

CSA Directive: the EU's Child Sexual Abuse Directive 2011/93;

CSAM: child sexual abuse material (most commonly photos and videos);

CSAR: or CSA Regulation, the EU's proposed child sexual abuse regulation 2022/0155;

"Child pornography": child rights groups discourage the use of this term, as it does not capture the gravity of the offence of child sexual abuse. However, the use of the term is sometimes required for legal accuracy, for example definitions and wording in the CSA Directive and related national laws;

DSA: the Digital Services Act 2022;
Dissemination: this refers to the crime of spreading CSAM, for example by sending it to or sharing it with someone else, usually online;

Encryption: a security technique that conceals data by applying mathematical algorithms, so that it can only be decrypted by parties that hold the correct key;

ePD: the ePrivacy Directive 2002/58/EC;

GDPR: the General Data Protection Regulation;

Grooming: a term for the solicitation of children, when an adult makes contact with a child, including via the internet, with the intention of committing child sexual abuse or producing material of sexual abuse;

Hotlines: the national organisations that operate web-based and/or telephone lines for reporting online child sexual abuse. After a check, these organisations may initiate the removal of the material, and/or forward reports to the relevant investigative services and internet providers or platforms for action;

IA: Impact Assessment (for the CSAR proposal);

Lex specialis: a law which builds on and particularises another law;

NCMEC: the US National Center for Missing and Exploited Children;

OCSA: online child sexual abuse, a term referred to in the CSAR covering the creation or sharing of CSAM as well as the solicitation of a child;

Sexting: the exchange of text, images or videos of a sexual nature via a digital message service;

Service provider: this term is used for a broad range of providers of services on the internet, such as hosting providers (e.g. cloud and web hosting services), online platforms (e.g. app stores, social media platforms) and intermediary platforms (e.g. those offering internet access or network infrastructure);

Solicitation: the legal term for the crime of grooming children for sexual purposes;

TCO: Terrorist Content Online (Regulation); The Temporary Regulation: EU Regulation 2021/1232 on a temporary derogation from certain provisions of Directive 2002/58/EC [ePrivacy Directive] [...] for combating online child sexual abuse.

Executive Summary

The EU's proposed Child Sexual Abuse Regulation (CSAR) threatens the safety, security, privacy and free expression of everyone that uses the internet globally – including the very children that it aims to protect. Instead of focusing its efforts on pursuing criminal investigations against genuine suspects, the CSAR treats every internet user as a potential child abuser, which is disproportionate under EU law and contradicts the presumption of innocence.

The CSAR is not compatible with fundamental rights

In Chapter 2 and throughout this paper, our analysis shows that the CSAR proposal fails to meet the key human rights principles of necessity and proportionality, and will likely constitute a large-scale violation of the fundamental rights of all internet users, including potentially infringing the essential core of the right to privacy.

We argue that the proposal likely contradicts several existing EU laws including the Digital Services Act, lacks a sufficient legal basis, privatises the protection of children – a state responsibility – and may not meet the EU requirement of subsidiarity.

The statistics which the European Commission has put forward to justify the proposal's intrusive measures are opaque, misleading and lack independent review, instead taking vague supplier claims at face value. The CSAR likely also violates the EU prohibition of general monitoring. In particular, detection orders cannot be implemented in a way that is sufficiently targeted, effective or, therefore, lawful.

The CSAR is not technically or practically feasible

In Chapters 3 and 4, we raise serious concerns about the technical and practical feasibility of this overly-complicated and bureaucratic proposal, along with procedural concerns that its proposed solutions could make it harder for law enforcement agencies to investigate and prosecute perpetrators of child sexual abuse (CSA).

These proposed measures are likely to be not only ineffective, but even counterproductive. The CSAR could also hamper the removal of child sexual abuse material (CSAM), for example by incentivising the use of blocking orders.

Further, the proposal provides limited information about how EU Member States will be able to implement its

rules, especially as the majority of the enforcement burden will fall to Ireland and the Netherlands. The proposal also relies on a harmonised definition of CSA, and it is unclear how this could be enforced, given the fragmented national rules across Member States concerning the legal age of sexual consent.

Scanning technologies, which all providers offering online services on the EU market could be forced to implement in order to comply with the CSAR, cannot be implemented safely and securely. They are inaccurate for all types of CSAM, and especially flawed for "new" material and grooming detection. They have high rates of false alarms, and would undermine end-to-end encryption (E2EE), a vital human rights tool.

The proposed risk assessment and mitigation measures will incentivise providers to take the most intrusive steps possible in order to comply with the legislation, which will require them to have **knowledge of the content of everyone's private digital lives all the time.**

The CSAR will lead to serious harm across society, including to children

Throughout our analysis, we emphasise that these measures are likely to have profound consequences for anyone that relies on digital tools to stay safe.

In particular, the proposal is likely to deprive child sexual abuse survivors, as well as women trying to leave abusive relationships, of safe digital spaces, whilst also making their personal devices vulnerable to hacking by their abusers and other malicious actors. It will break trust in digital communications and remove the possibility of online anonymity, making the work of journalists, human rights defenders, political dissidents, protesters and activists more difficult and less safe.

And it could lock some people out of digital communications services, for example those who face high levels of exclusion, such as undocumented people and Roma people, or others with low levels of digital literacy.

The proposed law will also catch large amounts of legitimate communications in its broad net. This includes teenagers lawfully exploring their sexual self-identity, which could particularly affect LGBTQ+ teenagers, as well as adults consensually sharing lawful material of a sexual nature, who will find their pictures and conversations sent to the police. It will lead to high numbers of dangerous false accusations and the possibility of unlawful retention of people's data even after they are confirmed as innocent.

Child rights experts also warn that pervasive online surveillance can cause psychological harm to children and hamper their free expression and development. Crucially, the proposal fails to sufficiently engage with important preventive and societal measures which could stop the problem from existing in the first place.

The EU must withdraw the CSAR and pursue alternative measures

The abhorrent crime of child sexual abuse, in all its forms, requires effective action from national governments and EU institutions. However, we warn that it is unlikely that the CSAR will be effective or efficient at achieving its aims.

Based on our analysis, EDRi urges the EU's co-legislators to reject the proposed CSAR. Instead, we call on EU and national authorities to pursue evidence-based approaches that are more likely to be effective in the fight against CSA, whilst ensuring respect for fundamental rights.

This includes societal measures such as increasing access to welfare, mental health and other support services, as well as reforming judicial institutions and law enforcement authorities. Crucially, it also includes empowering children and teenagers to make sensible and informed decisions about how they act online by educating and empowering them.

National and EU institutions, services and authorities must enable this by ensuring that children and young people are supported and believed when reporting abuse, and that cases are pursued swiftly and with sensitivity to the young person's needs, which are currently barriers to justice for survivors.

There are also many measures in existing legislation, particularly the 2011 CSA Directive, its upcoming revision, and the 2022 Digital Services Act, which will positively contribute to tackling CSAM, but which have not been (fully) implemented yet. The EU should also reinforce the network of national hotlines already leading the way in the fight against CSA, by ensuring that they have a legal basis for their work and more resources to carry it out.

Low-tech measures, such as ensuring that internet users can easily report abuse, can further help in the fight against online CSA. Implementing evidence-based prevention strategies will ensure that the EU's approach tackles the roots of CSA, not just the symptoms. And by bringing all the right stakeholders to the table – children's rights groups, digital rights groups, experts in tackling CSA, other human rights groups, and survivors – the EU will be able to develop sustainable measures that can protect fundamental rights, including children's rights, and ensure a safe internet for all.

1. Introduction to the CSA Regulation in context

On 11 May 2022, the European Commission put forward a draft law that threatens the safety, security, privacy and free expression of internet users globally – including children. EDRi and over a hundred other human rights and civil society groups have called on the EU to reject this misguided proposal, which, despite its important goal, puts forward measures that are likely to be dangerous, ineffective and unlawful.[1]

The "Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse," or CSAR, is a law mandating the monitoring and partial restriction of virtually all public and private digital communications.

National authorities can require digital service providers, such as WhatsApp or Signal, to conduct even more intrusive censorship or surveillance, including in encrypted environments. Many experts have confirmed that this is technically unlikely to achieve its stated aims, but that the means proposed to attempt it will fundamentally undermine encryption, which is a vital human rights tool.[2]

The goal of fighting child sexual abuse (CSA) and keeping children safe online and off is critically important, and the EU and its Member States have a serious responsibility to respect, protect and fulfil children's rights. Sexual abuse can cause lifelong harm to victims and their families.

It is particularly important, given the gravity of the problem, that policy and legal responses are based on solid research, evidence and a proper assessment of the facts. Based on the analysis in this paper, we conclude that the CSA Regulation as proposed fails on all of these counts.

The proposal is aimed mainly at the crime of disseminating CSAM online. It does not focus on strengthening investigative capacity into CSA crimes, or on increasing prosecutions or convictions. Neither does it tackle the original (largely) offline acts of abuse, nor the complex factors that lead to offending. It does not put forward sufficient evidence that its proposed measures will be effective in achieving its aims.

On the contrary, the significant volume of false alarms that will inevitably arise from these new rules could make investigations into perpetrators even harder. These likely ineffective measures will also constitute a severe interference with human rights on a mass scale.

Child protection organisations confirm that CSA is most frequently committed by family members or other persons well known to the victim. This is reportedly true for 80-90% of CSA cases.[3]
As child protection organisations EPCAT International and the Global WeProtect Alliance state in a 2022 report based on interviews with hundreds of case workers and survivors:

> **"While commonly held perceptions tend to frame sexual abuse both online and offline in terms of 'stranger danger', in reality children face more frequent risk of harm from people within their circles of trust."**

Children's rights groups also note that most CSA crimes do not have an online component. And the Child Rights International Network (CRIN) emphasises that:

> **"Sexual violence is one of the worst crimes against children as it violates so many of their rights, but it will continue if the root causes that allow it to exist in the first place are not challenged."**

Despite this, the CSAR focuses almost entirely on regulating online service providers. This causes us to question whether such rules are the right ones to tackle CSA, especially as the new law will make already overly-powerful Big Tech companies legally responsible for scanning and analysing the content of the most private conversations of every person that uses the internet. This will also harm children and their rights to privacy, data protection and free expression online.

In 2021, some Members of the European Parliament pointed to the "moral blackmail" which accused them of not caring about children when they tried to question legal and procedural issues with the CSA Regulation's predecessor.

This facilitated the rushed adoption of the "Temporary Regulation", a law which allowed service providers to continue a legally-questionable practice of scanning private communications. Since that point, the European Commission's Directorate-General for Home Affairs and Migration (DG HOME) has worked on the proposal for a long-term replacement. Even within the Commission, the new proposal – the CSA Regulation that is the subject of this paper – has been controversial.

At its first attempt, the proposal failed to pass the Regulatory Scrutiny Board (RSB), a Commission body responsible for assessing whether a legislative proposal is necessary and proportionate according to human rights law.

At its second attempt, the draft CSAR was approved by the RSB with "reservations" about its "significant shortcomings".

The RSB pointed out that parts of the proposed law would likely amount to generalised surveillance, which contravenes the EU prohibition of general monitoring – a risk that has also been emphasised by the United Nations High Commissioner for Human Rights.

The RSB also suggested that DG HOME seemed to have a preconceived idea about the rules they wanted to impose on service providers, and produced a legislative proposal and impact assessment to justify that, rather than starting with the problem and assessing the best methods to tackle it. The RSB also questioned the efficiency and proportionality of the proposal.

In this paper, the EDRi network argues that the proposed CSA Regulation lacks a sufficient legal basis; contradicts EU law – in particular fundamental rights law; adds significant complexity to existing processes which could hamper national efforts to remove CSAM; and is technically impossible for service providers to implement in a way that respects rights and effectively achieves its stated aims.

We urge the co-legislators to take the issue of CSA seriously by ensuring that laws mandating the use of digital technology are realistic, achievable, lawful, rights-respecting and actually effective. The CSA Regulation does not meet these criteria and must be replaced with sustainable, effective alternatives.

1 https://edri.org/our-work/european-commission-must-uphold-privacy-security-and-free-expression-by-withdrawing-new-law

2 **For example:** https://arxiv.org/abs/2110.07450; https://www.globalencryption.org/2022/05/joint-statement-on-the-dangers-of-the-eus-proposed-regulation-for-fighting-child-sexual-abuse-online/; https://privacyinternational.org/sites/default/files/2022-09/SECURING%20PRIVACY%20-%20PI%20on%20End-to-End%20Encryption.pdf

3 https://learning.nspcc.org.uk/media/1710/statistics-briefing-child-sexual-abuse.pdf

4 https://ecpat.org/wp-content/uploads/2022/01/05-01-2022_Project-Report_EN_FINAL.pdf

5 https://learning.nspcc.org.uk/media/1710/statistics-briefing-child-sexual-abuse.pdf; https://ecpat.org/wp-content/uploads/2022/01/05-01-2022_Project-Report_EN_FINAL.pdf

6 https://home.crin.org/issues/sexual-violence

7 https://www.politico.eu/article/european-parliament-platforms-child-sexual-abuse-reporting-law/

8 https://edri.org/our-work/internal-documents-revealed-the-worst-for-private-communications-in-the-eu-how-will-the-commissioners-respond

9 https://home.crin.org/issues/sexual-violence

10 https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report

2. Analysis of the legal framework

Chapter Summary

2.1 Legal basis
The CSAR would make digital service providers responsible for child protection, a state competence. It lacks a sufficient and legitimate legal basis, and may not be compliant with the principle of subsidiarity;

2.2 Fundamental rights
The CSAR unnecessarily and disproportionately limits a wide range of fundamental rights for a large proportion of the population;

2.3 General monitoring
The EU prohibits general monitoring, meaning that the CSAR is very likely in contradiction of this rule;

2.4 The Temporary Regulation
We have many fundamental rights and transparency concerns about the CSAR's predecessor, the ePrivacy temporary derogation. The Commission has not put forward any evidence of its efficiency or effectiveness, yet even broader measures are put forward under the CSAR;

2.5 The ePrivacy Directive
The CSA Regulation's derogation from the ePrivacy Directive may contravene the essence of the original Directive's goals and purpose;

2.6 The 2011 CSA Directive
Several Member States continue not to meet their child protection obligations more than a decade after the CSA Directive's entry into force;

2.7 The Terrorist Content Online Regulation
The TCO has incentivised the over-removal of legitimate content on a purportedly "voluntary" basis, with the CSAR creating similar risks;

2.8 The GDPR
The CSA Regulation may incentivise the processing of personal data in ways that are not compliant with the GDPR, in particular risk assessment and mitigation;

2.9 Children's digital rights
The CSAR largely fails to recognise that children are digital citizens who need respect for their privacy and data protection, too.

2.1 The legal basis of the CSA Regulation

According to the European Commission's proposal, the legal basis of the proposed CSA Regulation is Article 114 of the Treaty on the Functioning of the European Union (TFEU).

Like the Temporary Regulation that it is intended to replace, the CSA Regulation is a proposed derogation from the 2002 ePrivacy Directive. Unlike its predecessor, it is also lex specialis to the Digital Services Act (DSA). Article 114 TFEU allows the co-legislators (the European Parliament and the Council of the European Union) to set laws on the "functioning of the internal market", predominantly by removing barriers to trade.

Whilst part of the CSAR proposal is indeed lex specialis to the Digital Services Act (DSA), a single market regulation, other parts of the CSAR clearly relate to the practices of law enforcement, despite no legal basis to this effect.

The blurring of law enforcement competencies and internet regulation in the CSAR could therefore lead to a harmful privatisation of the protection of children, which is and should remain a law enforcement responsibility.

Rules which relate to practices of law enforcement, and not the harmonisation of obligations for private service providers, would require a specific proposal with an appropriate legal basis. We further argue that the creation of a new legal basis for the processing of personal data under the GDPR is not acceptable given that, as Chapter 2.8 explains, the CSAR is probably incompatible with the GDPR.

Article 114.3 requires that any new internal market law must respect "consumer protection" and take account "in particular of any new development based on scientific facts". As this paper will demonstrate, the proposed CSAR legislation would weaken consumer protection.

What's more, the draft CSAR also does not take into account the scientific facts that have been repeatedly put forward by internet and cybersecurity experts around the world, most specifically regarding the technical tools that would be required to implement the proposal, which will be explored in Chapters 3.2, 3.4 and 4.1.

In addition, one of the justifications for the proposal has been the rise in online activity as a result of COVID-19 lockdowns. However, as lockdown conditions are not currently in force, this justification is – at best – debatable.

The requirement for all EU laws to respect subsidiarity also creates problems for the CSAR.[11] As the CSAR's Explanatory Memorandum outlines, the principle of subsidiarity requires that an EU law in an area of "shared competence", like the digital single market, can only be pursued if it "can be better achieved at Union level" than at national level.[12]

As the 2021 annual report of the global network of child protection hotlines, INHOPE, demonstrates, the prevalence of CSA crimes, the types of service providers, as well as the number of dissemination methods all vary significantly between Member States.[13]

The proposed one-size-fits-all approach in the CSA Regulation could thus risk interfering with national efforts, particularly those of hotlines, and overshadowing vital national context and knowledge as a result of its attempts to standardise approaches to online CSA.

The centralised processes and rules of the CSA Regulation may therefore be non-compliant with the principle of subsidiarity. The lack of harmonised national penal codes and definitions (e.g. of the age of sexual consent) (see Chapter 3.2) further emphasises the subsidiarity challenge that faces the CSAR.

Given the inconsistency between the CSAR's legal basis and the rules that it puts forward, the risks of the privatisation of law enforcement duties, and the risk of undermining national efforts to tackle CSAM, there are intractable problems with the legal basis of the proposed CSAR.

11 https://www.europarl.europa.eu/factsheets/en/sheet/7/the-principle-of-subsidiarity

12 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0209

13 https://inhope.org/media/pages/articles/annual-reports/8fd77f3014-1652348841/inhope-annual-report-2021.pdf

2.2 The EU Charter of Fundamental Rights

The Charter of Fundamental Rights of the European Union ("the Charter") is the EU's commitment to recognise, preserve, protect and develop "the indivisible, universal values of human dignity, freedom, equality and solidarity".[14]

In particular, this includes several fundamental rights which will be directly and – as we will argue – disproportionately limited by the proposed CSA Regulation: Liberty and security (Article 6), Respect for private and family life (Article 7), the Protection of personal data (Article 8), and Freedom of expression and information (Article 11).

As a result of the limitation of these rights, the CSAR is likely to impose limitations on other fundamental rights whose exercise relies on these rights, in particular: freedom of thought, conscience and religion; freedom of assembly and association; freedom to conduct a business; non-discrimination; the rights of the child; the rights to health care, consumer protection, good administration and an effective remedy; the rights to a fair trial; and the presumption of innocence.[15]

We point out that it is not just adults that have a right to the privacy of digital communications, but children too, a concern which has been largely neglected in the Commission's proposal. These rights also all have counterparts in international instruments to which the EU and/or EU Member States are parties, such as the European Convention on Human Rights (ECHR) and international human rights law.

In particular, Article 52.1 of the Charter requires all limitations of fundamental rights to be lawful, legitimate in a democratic society, necessary, proportionate, and sufficiently safeguarded.

This does not mean that rights such as privacy can never be limited, nor is that the argument made in this paper. Rather, any limitation must meet the Charter's

criteria. It is a vital facet of human rights that rights cannot be limited arbitrarily, and that states seeking to limit them must bear the burden of demonstrating that doing so is necessary (meaning that the measure is effective and that there is no alternative viable option) and proportionate (meaning that the gravity of the infringement relates to the gravity of the issue).

In this paper, we argue that the measures put forward in the CSA Regulation are manifestly unnecessary and disproportionate, entailing impermissible limitations on a wide range of fundamental rights of people not just in Europe, but potentially across the world. The serious goal of preventing child sexual abuse deserves a far more robust approach – one that is lawful and legitimate according to EU law.

2.2.1 Understanding the CSAR's proportionality assessment

Whilst the gravity of the crime of CSA is an important consideration in assessing the necessity and proportionality of the CSAR, the pursuit of even very serious crimes does not mean that any measure is permissible. It is especially important, therefore, to interrogate the widely-reported figures about the prevalence of online CSAM and why such figures cannot be taken at face value.

In 2021, the US National Center for Missing and Exploited Children (NCMEC), which coordinates the global reporting of suspected abuse material, said that there were 29.3 million online CSAM reports in 2021; 8 million more reports than in 2020.[16]

This does not necessarily mean that more CSAM has been disseminated, however. As Dutch hotline EOKM explains, increased rates of dissemination can be a result of more widespread awareness and reporting, changes in the use of automated detection tools, or other reasons for more abuse being brought to light.[17]

These figures also include reports of suspected CSAM which subsequently turn out not to be CSAM, as well as repeat or "viral" content. This makes it very difficult to properly assess the scale of online CSAM, as pointed out by Netzpolitik.[18]

The Ireland case study discussed in Chapter 3.5 provides a real example of how statistics like the 29.3 million 2021 reports translate into reality, showing that the scale of CSAM cannot be sufficiently understood based on high-level numbers.

This issue is also highlighted in research from Meta – whose platforms report the vast majority of online CSAM to NCMEC – that "copies of just six videos were responsible for more than half of the child exploitative content we reported" in a sample period in 2020.[19] This suggests that the millions of reports may relate to a comparatively small number of pieces of content.

This does not mean that the repeat sharing of offending content should not be prevented; it is still a crime which causes serious harm, and must be tackled. But the figures put forward by the Commission to justify the CSAR's highly intrusive measures do not accurately

portray the reality of the situation, which causes us to question the accuracy of the European Commission's assessment of the new law's proportionality.

14 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

15 **Articles 10, 12, 16, 21, 24, 35, 38, 41, 47 and 48 respectively.**

16 https://ec.europa.eu/commission/presscorner/detail/en/QANDA_22_2977

17 https://www.weprotect.org/wp-content/uploads/EOKM-Annual-report-2021.pdf, p.8

18 https://netzpolitik.org/2022/ncmec-figures-explained-how-the-spectre-of-millionfold-abuse-haunts-european-policy-makers/

19 https://about.fb.com/news/2021/02/preventing-child-exploitation-on-our-apps/

> **"it is not just adults that have a right to the privacy of digital communications, but children too, a concern which has been largely neglected in the Commission's proposal."**

2.3 The EU prohibition of general monitoring

2.3.1 Case law of the CJEU

The Court of Justice of the European Union (CJEU) has repeatedly held that combatting serious crime cannot justify a general and indiscriminate obligation to retain traffic data and location data (metadata) for electronic communications services. Serious crime related to CSA was explicitly considered in one of those judgements.[20]

The CJEU ruled that the IP address assigned to the source of an internet connection only (no other metadata) could be retained for the purpose of combatting serious crime, in particular CSA.[21]

However, insofar as content is concerned, the Court ruled that legislation permitting public authorities to have access to the content of electronic communications on a generalised basis must be regarded as compromising the essence of the fundamental rights to privacy as enshrined in Article 7 of the Charter.[22]

As we show in Chapter 3.4, the CSAR's detection orders, and possibly other measures, go even further than this, making the CSAR likely to constitute unlawful general monitoring, as well as violating the very essence of the right to privacy.

However, there seems to be no consideration of these legal issues in the CSAR proposal. The only reference to the La Quadrature du Net judgement in the Explanatory Memorandum and Impact Assessment accompanying the CSAR mentions paragraph 126 about the positive obligation to combat serious crime under Article 7 of the Charter.[22]

This is highly misleading when it comes to the conclusion of the overall judgement, which applied only in the special case of source IP addresses. The draft CSAR proposal goes much further than just IP addresses, without a proper legal analysis of compliance with the Charter.[23]

What's more, the CJEU has also defined narrow limits for the use of automatic detection systems by service providers. With regard to the prohibition of general monitoring, the use of filtering technologies is only permitted if the filters have such a low error rate that an independent assessment of the content by the service provider is not needed.[24]

The essential reliance on context to determine whether a piece of material is CSAM or not, as well as the low practical accuracy of all CSAM detection technologies (Chapter 3.2), means that any scanning, whether of public-facing or private digital communications, and regardless of whether it is for known material, new material or grooming, would be very unlikely to meet this threshold.

Whilst the Explanatory Memorandum to the CSAR notes that measures have been taken to ensure compatibility with the prohibition on general monitoring, the accompanying legal analysis does not explain what these measures are, nor how the CSAR achieves this.

2.3.2 The Digital Services Act

The proposed CSA Regulation is meant to complement existing EU rules, in particular the Digital Services Act (DSA).

The DSA reasserts the EU prohibition of general monitoring obligations (Article 8) under the e-Commerce Directive, and further reinforces the EU's limited liability system (Articles 4-7). This means that digital service providers cannot be held responsible for illegal content about which they have no knowledge or control.

Article 1(3)(b) of the CSAR states that it shall not affect the rules laid down in the DSA. Yet in its currently-proposed form, our analysis suggests that the CSAR would fundamentally undermine the DSA.

The DSA's prohibition of general monitoring forbids general obligations from being placed on service providers which would compel them to actively search for illegal material. The ban on mandated general monitoring is a core protection against censorship, which enjoys general protection under the EU Fundamental Rights Charter. This is important because it protects users' freedom of expression and privacy, and avoids harmful profiling.[25]

However, the CSAR's detection obligations (see Chapter 3.4) would expressly contradict the DSA's prohibition of general monitoring obligations and the limited liability system. Likewise, the risk assessment and mitigation measures (Chapter 3.3) would also be likely to violate these provisions, as the new rules would force providers to have knowledge of the

20 CJEU, La Quadrature du Net and others, joined cases C-511/18, C-512/18 and C-520/18.

21 **CJEU, La Quadrature du Net, paras. 152-156.**

22 **CJEU, Schrems I, C-362/14, para. 94.**

23 Similarly, the only reference in the Impact Assessment to the 2014 Digital Rights Ireland judgment on the Data Retention Directive (joined cases C-293/12 and C-594/12) is para. 42, where the CJEU recalls that the fight against serious crime is an objective of public interest.

24 CJEU, Poland v Parliament and Council, 2022, C-401/19, para. 90.

25 https://www.eff.org/deeplinks/2022/08/general-monitoring-not-answer-problem-online-harms

content of people's private messages in order to conduct risk assessments. This undermines the rationale of the EU's conditional horizontal liability regime, and will lead to intrusive monitoring of user content.

Digital services and platforms should not have knowledge of, or control over, people's private online communications. The confidentiality of communications is a vital tenet of democratic societies, the ePrivacy Directive and the EU's fundamental rights regime. Bringing third parties – especially commercial entities – into people's private conversations and exchanges will be a severe and disproportionate infringement of this right.

As we have argued in "10 principles to defend children in the digital age", such a limitation on the right to privacy would be justifiable only in the event of reasonable, warranted suspicion, and at that point, only with the specific involvement of a judicial or law enforcement authority.[26] This is not the case in the CSAR.

What's more, the necessity of the CSAR is called into question by the substance of the DSA. The latter already contains a variety of tools which, following the DSA's implementation, will have a significant positive impact in the fight against CSAM.

In general terms, the DSA covers any illegal content (Article 3.h) and includes special provisions on criminal activities and law enforcement. As far as hosting services, including online platforms, are concerned, the DSA mandates a mechanism for an entity or person to report illegal content, with safeguards. Providers must give a Statement of Reason for why information is being deleted. Not only must individuals be able to report posts containing allegedly illegal content, but also other entities, including so-called "Trusted Flaggers". These trusted flaggers must be pre-approved by the DSA's independent national regulator, the Digital Service Coordinator.

Furthermore, the DSA foresees a set of provisions on risk assessment and mitigation for very large online platforms, which, furthermore, will be subject to independent audits. Such risks cover a variety of aspects, including risks related to the dissemination of CSAM.

As such, the difference in approach to risk mitigation between the CSAR and the DSA creates a lack of legal clarity and coherence between the relevant provisions – something that has not been sufficiently explained or justified by the CSAR proposal.

Considering the comprehensive scanning of online content that the CSAR's rules would entail, as well as the Regulatory Scrutiny Board's view that the proposal may amount to general monitoring obligations, a more thorough analysis should have been performed by the Commission before putting forward the CSAR proposal.

Such an analysis would have revealed that the CSAR is highly likely to violate the prohibition of general monitoring according to the CJEU and the DSA/eCommerce Directive, and probably also the limited liability system.

2.4 The ePrivacy Directive

The e-Privacy Directive (ePD) covers specific privacy and data protection safeguards in the electronic communications sector. It was adopted in 1997, and revised in 2002 and 2009.

The ePD was created to ensure privacy and to protect personal data in the electronic communications sector by "complementing and particularising" matters covered in a general way by the main legal instrument, the Directive on Data Protection, now the General Data Protection Regulation (GDPR).

For example, the confidentiality of the metadata of communications and information stored or accessed on an individual's device is specifically protected under the ePD.

Article 6 of the GDPR on the lawfulness of processing does not apply in the ePD context. Instead, the legal basis for processing communications data must be expressly provided for by law in either the ePD itself (Articles 5, 6 and 9) or by national or EU law restricting the right to the confidentiality of communications in accordance with ePD Article 15(1).

This is generally more restrictive for data controllers than the GDPR, for example considering the absence of legitimate interest as a legal basis for processing.[27]

Moreover, under ePD, any processing of communication data beyond the transmission of the communication itself formally constitutes an exception to the confidentiality of communications laid down in Article 5(1), and under CJEU case law, such exceptions must be interpreted strictly.[28]

To execute detection orders under Articles 7-11 of the CSA Regulation, the rights and obligations provided for in Articles 5-6 of the ePD (confidentiality of communications) are restricted in accordance with and by analogy of ePD Article 15(1).

That provision allows Member States to adopt legislative measures to restrict the confidentiality of communications only "when such restriction constitutes a necessary, appropriate and proportionate measure within a democratic society" including "to prevent, investigate, detect and prosecute criminal offences" and, crucially, in compliance with the Charter.

2.5 The Temporary Regulation

The CSA Regulation follows the Temporary Regulation 2021/1232/EU, which aims to legalise the voluntary scanning of interpersonal communications by service providers.

The aim of this Regulation is to enable certain online communications services to continue using technologies to detect and report child sexual abuse online, and remove child sexual abuse material on their services, effectively by moving the applicable legal framework for such practices from the ePD to the GDPR.[29]

Its basis for doing so was widely criticised, with several MEPs noting that if the law were challenged at the CJEU, it would likely be invalidated.[30] EDRi has raised concerns about the lack of judicial oversight and transparency of the regulation, the risk of general surveillance, and the possible violation of several fundamental rights.

The Temporary Regulation has a limited duration and narrow scope, restricted to voluntary activities of certain online services during an interim period of 3 years, which is set to expire in August 2024, unless it is extended by the co-legislators.

Despite several requests from EDRi, the European Commission has never published statistics on the effectiveness of the scanning under the Temporary Regulation, for example what percentage of reports made on the basis of voluntary scanning was actually confirmed to be CSAM; or what percentage of reports led or even contributed to arrests, prosecutions and subsequently to convictions.

For the last two years, the European Data Protection Board (EDPB) has been trying to produce a list of rights-respecting scanning tools to support the implementation of the Temporary Regulation, which allows for "voluntary" scanning of private messages by providers. The EDPB is yet to produce any recommendations. This would suggest that there are no quick solutions that the CSAR could use either – at least not ones that respect fundamental rights.

2.6 The Child Sexual Abuse Directive (2011)

The 2011 Child Sexual Abuse Directive is an EU law focusing on preventing and prosecuting child sexual abuse in Europe. However, its rules are fragmented and have not been consistently applied by EU Member States. This suggests that new legislation may be premature at best in an environment where Member States are not yet doing everything they can to protect children.[31]

The most recent review of the law by the European Parliament in 2017 showed many areas where Member States had failed to implement the necessary changes to sufficiently protect children. In 2022, continued non-compliance led the European Commission to launch new infringement proceedings against Ireland, Portugal, Spain and Italy.[32]

As such, the benefits of this important piece of legislation to protect children are still to be fully realised. Yet the European Commission is currently pursuing an update (recast) of the Directive, with consequences for the CSAR (which relies on the Directive for provisions including definitions).[33]

26 https://edri.org/our-work/chat-control-10-principles-to-defend-children-in-the-digital-age/

27 **When the European Electronic Communications Code (EECC) Directive extended the scope of the ePD to cover number-independent interpersonal communications services in December 2020, providers of such services could no longer rely on the GDPR for their generalised scanning practices. The ePD contains no provision permitting such scanning, rendering it unlawful under the ePD framework.**

28 CJEU, C-119/12 Probst, para. 23 as well as the extension data retention case law interpreting ePD Article 15(1).

29 **Recital 15 of Regulation (EU) 2021/1232.**

30 https://www.politico.eu/article/european-parliament-platforms-child-sexual-abuse-reporting-law

31 https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2017)598614

32 https://www.europarl.europa.eu/doceo/document/A-8-2017-0368_EN.html

2.7 The Terrorist Content Online Regulation

Over the past years, the EU has been attempting to counter purported "terrorist content" online by pressuring public hosting service providers such as social media to censor such content more quickly and more systematically.

The Terrorist Content Online Regulation (TCO) 2021/784 forces hosting providers to respond to content removal orders issued by the Competent Authorities of Member States (mostly law enforcement) within one hour, and requires hosting providers to prevent the dissemination of "terrorist content" by adopting certain proactive measures.

A major concern for the functioning and freedom of the internet is the extension of the upload filter regime the EU introduced for copyright to "terrorist content".

Requiring internet companies to monitor everything we say on the web not only has grave implications for freedom of speech, but also follows a dangerous path of outsourcing and privatising law enforcement.

EDRi has long advocated for the full respect for fundamental rights in the TCO Regulation. Unfortunately, the Regulation, which still includes very dangerous measures, was adopted by the EU's co-legislators in 2021. There are key learnings for the CSAR.

2.7.1 "Voluntary" upload filters

The TCO Regulation strongly encourages platforms to make every practical effort to remove terrorist content, including by relying on their terms of service. Given current content moderation practices, this frequently involves the use of ill-suited automated content filtering technologies.[34]

The CSAR equally incentivises voluntary measures via its risk assessment and mitigation measures. Measures that involve generalised scanning of user content, effectively constituting upload filters, are indirectly mentioned as possible mitigation measures in the CSAR under the heading of industry best practice (e.g. Recital 18). At the same time, the proposal states – almost in passing – that providers' measures must be in accordance with Union law.

This creates a remarkable lack of legal clarity, since industry "best-practice" measures, especially those involving generalised scanning, may very well conflict with EU law on data protection and confidentiality of communications.

The purported need for the Temporary Regulation, after the scope of the ePrivacy Directive was extended in December 2020, is prima facie evidence of that inherent conflict between industry measures and EU law. The CSAR proposal is vague on these conflicts, and this lack of clarity is likely to have a detrimental effect on users' fundamental rights when service providers implement mitigation measures to comply with the CSAR. The consequences of this are explored further in Chapter 3.3.

2.7.2 "Voluntary" referrals

The CSAR allows Coordinating Authorities (Article 32) and the EU Centre (Article 49) to send "referrals" to service providers for their voluntary consideration of whether the notified content constitutes illegal child sexual abuse material.

EDRi has criticised the possibility of referrals in the TCO Regulation proposal, since the adjudication of illegal content should not be left to the voluntary consideration of service providers, especially for serious offences.

As with the TCO Regulation, we are particularly concerned that the CSAR's referrals will be used instead of removal orders for content where the potential illegality is not obvious, since service providers are likely to remove notified content voluntarily, especially for content related to serious criminal offences such as terrorism and child sexual abuse. However, for precisely this type of content, independent judicial review is of the utmost importance to protect freedom of expression and access to information.

33 https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13073-Combating-child-sexual-abuse-review-of-EU-rules

34 https://www.euractiv.com/section/digital/opinion/misguided-solution-to-terrorist-content-will-have-bad-consequences-for-our-rights/

2.8 The General Data Protection Regulation (GDPR)

Adopted in 2016 and entering into application in 2018, the EU's General Data Protection Regulation (GDPR) – and its policing counterpart, the Data Protection Law Enforcement Directive (LED) – established world-leading data protection rules. This framework recognised that the protection of personal data is not a singular issue, but that a large number of other fundamental rights are also put at risk when the right to data protection is infringed.

In particular, the GDPR establishes key rules including ones on transparency of data processing; the need for a specific legal basis to process personal data; enhanced protections for special (protected) categories of data; and rights to redress.

Firstly, the GDPR establishes that the processing of biometric data (for example fingerprints or facial templates) is prohibited (Article 9.1) except under specific conditions. The CSAR, however, proposes that platforms can use "age verification" measures to reduce the risk of CSAM dissemination on their platform (Article 3.2.b). Whilst no specific method for age verification is put forward in the CSAR, the "online age assurance" industry already offers many purported "solutions" which process biometric data.

We have concerns about how such methods misappropriate the legal basis of consent and create serious risks to people's fundamental rights.[35] This issue of age verification will be explored in more depth in Chapter 3.3, but already raises concerns about how provisions in the CSAR might undermine GDPR rules.

Secondly, the GDPR requires the processing of personal data to be justified by one of its legal bases, for example "legitimate interest" (Article 6.1.f). Legitimate interest is a specific and limited basis, and the burden rests on the data controller to demonstrate that their legitimate interest complies with the conditions of the GDPR.

Following the publication of the draft CSAR, Commission representatives suggested that they intend the generalised scanning

of hosting services (e.g. social media posts) to be permissible under the GDPR basis of legitimate interests.[36] The Commission noted that Article 6.1.c (legal obligation) could not apply to such practices because whilst this generalised scanning is strongly encouraged, it is not legally required under the CSA Regulation.

This is problematic for several reasons. The controller that would purportedly have a legitimate interest in such scanning practices would be the platform or service provider. Legitimate interest does not override the ban on processing special categories of data set out in GDPR Article 9, and the exceptions of said article typically require a specific basis in Member State or Union law.

The few exceptions enumerated in Article 9 that are compatible with legitimate interest are specifically meant to reconcile the GDPR with freedom of expression, religious freedom and the freedom of association, not to enable generalised surveillance and upload filters.

As a consequence, legitimate interest would not be a suitable basis for scanning for CSAM by a service provider as they would fail to meet the criteria of Article 9 of the GDPR. We also consider it unlikely that meeting the CSAR's undefined, and potentially coercive (see Chapter 3.3) risk mitigation obligations could meet the necessary threshold to be considered a legitimate interest for the controller, in light of the balancing act required by the GDPR.

The CSAR's approach to risk mitigation is to incentivise certain outcomes without defining how to reach them, which the Commission calls "technologically neutral". However, this could compel providers to take measures which may in fact contradict the GDPR (such as scanning social media posts on the basis of legitimate interest or using biometric data for age verification), whilst hiding the specific details out of the purview – and therefore democratic scrutiny – of the legislative process.

At the same time, the risk mitigation obligations of the CSAR do not meet the standard set by Recital 41 of the GDPR for a legal obligation to be sufficiently specific and foreseeable to the data subjects that may be subject to the processing of personal data. This by itself causes an incompatibility with the GDPR, but also with the European Convention on Human Rights.

As a result, such measures for hosting, ancillary chat and social media services do not have a clear basis in the GDPR. It is notable that the penalty for non-compliance with the CSAR is 6% of global turnover, compared to the GDPR's 4%. Providers may therefore be incentivised to process personal data in ways that could contradict the GDPR over the CSAR.

35 https://edri.org/wp-content/uploads/2021/11/EDRI_RISE_REPORT.pdf

36 https://www.3-is.eu/

2.9 Children's digital rights

2.9.1 The "Better Internet for Kids" (BIK+strategy)

When the European Commission published the proposal for a CSA Regulation, this was accompanied by an update to the 2012 "Better Internet for Children" strategy, coined "Better Internet for Kids", or BIK+.[37]

Read alongside the CSA Regulation, the BIK+ strategy provides additional information about the Commission's intentions and visions for the CSAR, as well as its intersection with other current legislation, like the DSA and the electronic IDentification, Authentication and trust Services (eIDAS) Regulation.

BIK+ indicates that one of the intentions of the CSA Regulation is to impose age verification measures on virtually all online communications services. It suggests that digital identity documents will be given to under-18s, which should ring alarm bells in terms of the threat to children's rights to privacy and data protection.

2.9.2 International recommendations on protecting children's privacy and data

The Council of Europe Lanzarote Convention requires that state parties – including EU Member States – take action to tackle child sexual abuse.

The EU Charter, the Convention on the Rights of the Child, any many other European and national instruments assert the responsibility of states to protect children. These requirements must be read in conjunction with an appreciation for children's right to privacy of communications.

The UN Committee on the Rights of the Child's General Comment 25 on children's rights in relation to the digital environment calls on governments to ensure that "children's participation does not result in undue monitoring or data collection that violates their right to privacy, freedom of thought and opinion".[38] It continues by stating that:

"Any restrictions on children's right to freedom of expression in the digital environment, such as filters, including safety measures, should be lawful, necessary and proportionate" and that digital surveillance "should respect the child's right to privacy and should not be conducted routinely, indiscriminately [...] nor should it take place without the right to object to such surveillance".

UNICEF has also raised the need to protect both children's security and privacy online. Their toolkit on Children's Online Privacy and Freedom of Expression said that improving privacy and data protection for children is essential for their development and their future as adults.[39] The toolkit highlights that any monitoring tools should "bear in mind children's growing autonomy to exercise their expression and information rights".

Furthermore, child rights groups like CRIN warn that it is harmful for children to be subjected to generalised digital surveillance and denied safe, private online spaces.[40] Intrusive internet monitoring regulations also deprive survivors of safe spaces and may even disincentivise them from seeking help.[41]

This is especially pronounced for children who rely on digital communications for a range of important reasons, such as to escape abusive situations; to develop their sexual or gender identity (e.g. LGBTQI+ young people), especially where their family situation is not supportive or even puts them at risk; to seek support and solidarity as a survivor; or for undocumented children. As we will establish in this paper, the measures proposed by the CSAR may constitute a serious interference with the privacy rights of all children that use the internet. For example, if the CSAR's rules are implemented:

- A 15-year-old in a country where she is above the age of consent, who lawfully sends a topless selfie to her partner, could have her message routinely scanned (if the service is subject to a detection order), flagged as CSAM and then reviewed by moderators.

  They would be obligated to send it to the staff at the EU Centre, who would then be obligated to send it to Europol and national police. The image would be analysed by the police, which could lead to an investigation, including the notification of parents, which may be especially harmful if the young person is legitimately exploring an LGBTQ+ sexual identity. As shown by our case study (see Chapter 3.5), even after the image is confirmed as innocent, the person's data might still be retained by police.

- A provider of an app for encrypted messaging could be forced via a detection order to implement technology (such as "client-side scanning") that scans the content of their users' messages before they send them. This technology would insert a vulnerability into all the users' devices.

  Children who use the messaging app to communicate with friends and to let their

parents know that they are safe when going to and from school would find their phones more vulnerable to hacking by criminals. This could give the latter access to children's personal information, location data, daily behaviour patterns and other sensitive information.

An undocumented young person who has fled persecution in a third country and is now in the EU, but does not have status, would be unable to get a digital identity document.

Without this digital identity document, they would not be able to verify their age when trying to use email, messaging or social media apps that use age verification to meet the CSAR's risk mitigation requirements. This would prevent them from being able to contact friends or family, or perhaps to seek health, welfare or legal advice.

37 https://digital-strategy.ec.europa.eu/en/policies/strategy-better-internet-kids

38 https://undocs.org/CRC/C/GC/25, paragraphs 18 and 59.

39 https://sites.unicef.org/csr/files/UNICEF_Childrens_Online_Privacy_and_Freedom_of_Expression(1).pdf

40 https://home.crin.org/issues/digital-rights/childrens-right-digital-age?rq=digital%20age

41 https://www.linkedin.com/pulse/why-i-dont-support-privacy-invasive-measures-tackle-child-hanff

3. Analysis of key articles in the CSA Regulation

Chapter Summary

3.1 Providers The wide scope of the providers and services subject to the CSAR will unnecessarily and disproportionately infringe on the rights of all internet users of all ages around the world, rather than being targeted against CSA perpetrators;

3.2 Content Whilst there are already concerns about the accuracy and reliability of "known" CSAM detection and concomitant threats to fundamental rights, the search for "new" CSAM and the prediction of grooming even further exacerbates these risks;

3.3 Risk assessment and mitigation The proposed risk model will incentivise service providers to take the most intrusive measures possible in order to avoid facing legal consequences, including age verification and potentially some generalised surveillance;

3.4 Detection Orders It will likely be impossible for detection orders to be served in a way that is targeted. As such, they will usually constitute unlawful general monitoring, and cannot be improved with safeguards. The threat and ensuing risks are even greater for encrypted communications, with the Commission intending that providers implement client-side scanning (CSS) methods despite serious security and human rights risks;

3.5 Case study – Ireland The new case study on CSAM reports to Irish police demonstrates the very real risk and potential consequences of false alarms and potential data retention, which will be exacerbated under the new rules proposed in the CSAR;

3.6 Reporting Obligations The reporting obligations in the CSAR are impractical, may overlap with the DSA, and should better take into account survivors' needs;

3.7 Removal Orders Removal orders can in theory be sufficiently targeted but should be limited to courts. They are also likely to be hampered by blocking orders (see 3.8);

3.8 Blocking Orders Blocking orders technically cannot work at the URL level and will disproportionately impact legal content at the domain level;

3.9 The EU Centre The Centre fails to meet its commitment to independence from Europol, and risks creating a complex bureaucracy which is likely to make it harder for CSAM to be removed from the internet, as well as for perpetrators to be investigated;

3.10 National authorities Creating new national authorities will likely worsen capacity and other issues already preventing the effective removal of CSAM by law enforcement.

> **"Every device subject to CSS will therefore be made technically much more vulnerable to attacks and hacking."**

3.1 Providers in scope

3.1.1 Which providers are in scope of the CSAR?

The CSAR will apply to virtually all online platforms and services on the EU market (even if based outside the EU), and therefore practically all the digital activities and forms of communication of children and adults alike (Articles 1.1, 1.2 and 2.f.iii), meaning:

Interpersonal communications services

- Online and app-based chat services, even those that are currently end-to-end encrypted: Facebook Messenger, WhatsApp, Signal, Telegram, etc., and the direct message components of platforms like Instagram, Twitter, LinkedIn, Reddit, etc.;

- Email: Gmail, Exchange Online, Proton Mail, etc.;
- Dating apps (Tinder, Grindr, etc.), chat rooms, instant messengers, Slack, and other chat-based services;
- Telephone calls and SMS messages;
- Services where the interpersonal communication is "ancillary", such as gaming;

Application stores

- The Apple App Store, the Google Play Store and alternative Android stores;

- Software repositories, like those enabling people to download Linux packages;

Hosting platforms and services

- Any place where users can store information, such as iCloud, Google Cloud, Microsoft Azure, Nextcloud or other cloud infrastructure service;

- File sharing/exchange services like DropBox, WeTransfer or Wikisend – even if only used for private storage without public access through shared links;

- Blog and Podcasting services like Wordpress, Squarespace or buzzsprout;

- Online services, as defined in the DSA, including the content of social media services and community platforms, like Twitter, Facebook, Reddit, YouTube, TikTok and LinkedIn, therefore meaning that posts on these platforms are in scope;

- By extension, anyone who uses a commercial hosting service, even for private purposes – such as to host your family or work email server, or to run a work or family cloud – would have their content subject to the risk mitigation measures and detection orders which the service provider must follow;

Internet access services, sometimes known as internet access providers.

There are no reduced obligations for micro or small and medium enterprises (SMEs). However, there are some specific circumstances where micro or SMEs will get free-of-charge support, for example when asking the EU Centre to perform an analysis of representative data to inform their risk assessment (Article 3.3).

3.1.2 Analysis of the providers in scope

The overwhelming majority of "information society services" users are regular, innocent people, using the services for legitimate reasons: to communicate with friends and family; to store or share photos of cherished moments; to work (notably lawyers, psychologists and other professions that rely on confidentiality); to play games; to learn; to build communities; to access healthcare; and to live their lives in an increasingly digitalised world.

The CSAR proposal even acknowledges that the issue is the misuse of platforms and services (Article 1.1). Yet the inclusion of these providers means that the CSAR's rules will impact *all* their users.

Whilst some CSA offenders exchange CSAM via these services or platforms, others have the capacity to develop their own services, for example on the so-called "dark net", which will be able to entirely circumvent the CSA Regulation.[42]

To echo the EDPS and EDPB, the proposed CSAR is therefore likely to do great harm to regular people, with a very limited impact on stopping CSA criminals. Whilst it is reasonable to conclude that at least some of these services should be subject to additional rules or practices (but not the ones put forward in the CSAR, as we argue throughout this paper), it is disproportionate to include such a wide range of services in the CSAR's scope.

For example, in the case of phone calls, Detection Orders would amount to wiretapping. In the case of text messages, Detection Orders would amount to interception. This contradicts rule of law requirements to pursue investigations only in genuinely individual, targeted, warranted cases – in online spaces, just as in offline ones.[43]

The fact that CSAM is shared at scale in the EU does not entail that all services and platforms are responsible for this, nor that all their users should be subjected to surveillance – especially given that the reported scale is not prima facie representative of the scale of CSA (see Chapter 2.2).

Whilst the European Commission has explained that the various types of orders that can be issued under the CSAR are designed to respond to this necessity and proportionality challenge, we will argue that the proposed orders fail to meet these criteria.

3.1.3 Specific issues for software application stores

The definition of software application stores in the CSA Regulation is taken from the Digital Markets Act (DMA, 2022/1925), where it means a type of online intermediation service, which is focused on software applications as the intermediated product or service.

This definition is quite broad, as the wording potentially covers any online service focused on the distribution of software. Besides the well-known large app stores for smartphones operated by Google and Apple, which are probably the intended scope of Article 6 in the CSAR, there are independent app stores for smartphones, software repositories for Linux distributions (Ubuntu, Debian, and many others), as well as websites focused on hosting software applications.

The obligations in the DMA only apply to a limited number of gatekeepers (probably only Google Play Store and Apple App Store), yet the obligations in Article 6 of the CSAR apply to all software application stores, independent of their size and economic resources.

This means a potentially very large number of app stores will be required to follow the CSAR's risk assessment rules.[44]

Most applications in these app stores will not be associated with providers of hosting services or interpersonal communications services, which are required to make their own risk assessments under Article 3. Therefore, independent software application stores cannot expect to rely on risk assessments conducted by other service providers, as foreseen by Article 6.2.

Moreover, unlike the major app stores which are closely integrated into smartphone operating systems, independent software application stores generally allow anonymous download of software without any registration (user account) or login to the service (access control).

This design makes it impossible for them to take measures to prevent child users from downloading software, where a significant risk has been identified, and implement age verification measures as required by Article 6.1.c.

43 https://edri.org/our-work/chat-control-10-principles-to-defend-children-in-the-digital-age/

44 **For example, the Linux distribution Debian offers 51,000 software packages as of 2022 (the number of applications may be lower since applications are often formed from multiple packages)** https://wiki.debian.org/DebianIntroduction

3.2 Content in scope (known and new material; grooming)

3.2.1 What content is in scope of the CSAR?

The types of content that service providers will become liable for under the CSA take three main forms. "Known" CSAM (Article 2.m) means "potential child sexual abuse material detected using the indicators contained in the database of indicators referred to in Article 44.1, point (a)".

The indicators refer to "material previously detected and identified" as CSAM by either a judicial or administrative authority, or by a national Competent Authority (Article 36.1). According to Recital 62 and Article 44.1, the EU Centre will build and manage the three databases of indicators. Moreover, the Impact Assessment suggests that while the indicators for known CSAM will be a form of "hash value", for unknown material and grooming, this would be artificial intelligence classifiers.

"New" CSAM (Article 2.n) is defined as known material, except for the indicators, which refer to "material previously not detected and identified" as CSAM (Article 44.1.b) "in accordance with Article 36(1)". This means that indicators of new CSAM will be submitted to the EU Centre by national Coordinating Authorities.

"Solicitation" (or "grooming") means the "solicitation of children for sexual purposes" (Article 2.o). According to Article 44.1.c, there will also be a database of indicators of solicitation, such as "language indicators" (Article 44.2.c). As the CSAR notes in its Explanatory Memorandum, the search for unknown content or grooming is even more intrusive than for known images.

3.2.2 Analysis of content in scope

Known CSAM

The search for "known" CSAM in the CSAR – which is already scanned for by many digital service providers using tools like PhotoDNA – is deeply problematic. In the Impact Assessment (IA) accompanying the proposal, the Commission explains that its assessment of the accuracy of scanning/detection tools is based entirely on self-reported statistics from the developers. It states, for example, that "Thorn's CSAM Classifier can be set at a 99.9% precision rate", citing simply "Data from bench tests" without providing any information on the test sets.[45]

Aside from being very vague, this statement is potentially misleading, as precision is not the same as accuracy.

The precision rate describes the precision for the detection of a specific category (e.g. a lot of skin is shown); while this can be high, the accuracy can still be relatively low, as accuracy describes the overall proficiency of the model. The Impact Assessment also fails to provide the rate of false negatives (how much CSAM is not detected), and the rate of false positives (how much material is incorrectly flagged as CSAM) of Thorn's tool.

The IA continues that "Microsoft has reported that, in its own deployment of this tool in its services, its accuracy is 88%", again with no additional verification or transparency. Similarly, there are no independent statistics on the accuracy of PhotoDNA.

The IA simply states that "PhotoDNA has a high level of accuracy" and the "rate of false positives is estimated at no more than 1 in 50 billion, based on testing (Testimony of Hany Farid, PhotoDNA developer)". It is also worth noting that PhotoDNA is tunable, so its accuracy is highly dependent on its configuration.

The European Commission's "technical expert group" has suggested that PhotoDNA is in need of an update "to keep up with the latest developments (and make it less vulnerable to manipulation)" (page 310).

This suggests a much less positive picture than what has been put forward by Farid about his tool. His claims about PhotoDNA – and those of other suppliers of scanning technology – reflect a serious lack of due diligence by the European Commission to verify what they have been told by private entities.

It is also becoming apparent that perceptual hashing methods such as PhotoDNA are easily reversible; they amount in effect to a black-and-white thumbnail image of the allegedly illegal image.[46]

For this reason, the use of PhotoDNA would mean distributing CSAM images in which sexual abuse and individuals can be recognised. Apple's proposed use of neural hashing in 2021 similarly turned out to be easily manipulated.[47]

A Freedom of Access to Information (FOIA) request made by EDRi member Gesellschaft für Freiheitsrechte seeking more information about these claims confirmed that the Commission uses industry claims, specifically from Thorn and Meta (Facebook), without any independent verification.[48]

However, a report from LinkedIn revealed that only 41% of the content identified by PhotoDNA as known CSAM on their platform in 2021 actually constituted CSAM.[49] Whilst this figure is not directly comparable with accuracy, it reveals that the success rates of such technologies in practice are significantly lower than what is claimed by the Commission and the companies developing the scanning technology.[50]

3.2.2.1 New CSAM

When it comes to "new" (aka unknown) CSAM, this AI-based technology has an even higher rate of false alarms than for known material. As became apparent in the Copyright Debates of the 2010s, artificial intelligence (AI) filters do not work well.[51]

In the intervening years, the technology has substantially improved, but still not to the extent that it can be relied on to identify possible crimes with an acceptably low error rate (which, as explained in Chapter 2.3, is a necessary precursor for such technologies to be considered lawful by the CJEU.)

From a societal perspective, there are good reasons why the search for new CSAM is so difficult. Social workers and law enforcement agents spend decades building up the knowledge and experience to be able to differentiate between acceptable and unlawful conduct, and still do not get it right all the time. In 2022, sophisticated image-recognition algorithms can still mistake a dog for a cat.[52]

So they will be hard put to tell the difference between a topless sunbather or a child's bath-time photo from an abuse scenario, or to infer whether a person is a teenager or just a young-looking adult. Context is vital in distinguishing between unlawful CSA and legitimate expression, and machine-learning technology cannot understand context, as it has no common sense.

The predictable outcome will be a flood of false alarms which will take up valuable time that could have been spent investigating actual cases of CSA.

This is not hypothetical; the Ireland case study in Chapter 3.5 reveals that hundreds of people are being falsely identified as disseminating CSAM for exactly these reasons. The 2021 report from Meta about the scanning of private messages similarly emphasises the inevitable existence of false alarms for AI-based tools.[53]

Meta reported that over less than two months, 207 Facebook or Instagram accounts had to be reinstated after detection reports falsely identified that they were disseminating CSAM, with thousands of other users still appealing the deletion of their accounts.

45 **Impact Assessment can be downloaded at** https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12726-Fighting-child-sexual-abuse-detection-removal-and-reporting-of-illegal-content-online_en

46 https://www.anishathalye.com/2021/12/20/inverting-photodna/

47 https://gangw.cs.illinois.edu/PHashing.pdf

48 https://www.asktheeu.org/en/request/technologies_for_the_detection_o#incoming-39916

49 https://edri.org/our-work/internal-documents-revealed-the-worst-for-private-communications-in-the-eu-how-will-the-commissioners-respond/

50 **On the basis of the Temporary Regulation, providers can choose to scan the private communications of their end users, but must meet several criteria, including reporting requirements.**

51 https://felixreda.eu/2017/09/when-filters-fail/

52 https://twitter.com/ellajakubowska1/status/1539543255309860864

53 https://transparency.fb.com/sr/eu-csam-derogation-report-2022/

Growing examples of grave false accusations of CSA, for example the newly-reported case of a parent being investigated by authorities for trying to seek medical advice for their child, demonstrate how severe a false alarm can be.[54]

3.2.2.2 "Live-generated" abuse

One of the main arguments in favour of including new CSAM and grooming detection in the scope of the CSA Regulation is the desire to tackle the problem of "live-generated" abuse.

The 2022 EPCAT International and WeProtect report specifically considers the rise in technology-facilitated CSAM such as live streaming.[55]

However, the main recommendation to tackle this problem given by the EPCAT report – based on the testimony of social workers and survivors – is to increase young people's awareness of hotlines where they can report what has happened to them, as well as to improve access to institutional reporting (police, social services and other authorities).

The use of scanning technologies is not mentioned by the report as a possible solution. Once again, we emphasise that the legitimate need to protect children does not entail that intrusive scanning technologies should be used for this purpose.

3.2.2.3 Grooming detection

The crime of grooming occurs when an adult interacts with a child with the intention to "engage in sexual activities" or "produce ... child pornography" (Articles 3.4 and 5.6 of the EU's 2011 Child Sexual Abuse Directive). As the Child Sexual Abuse Directive sets out, the crime applies only if the child is under the age of consent in their Member State.

The CSAR aims to use a combination of language analysis and probabilistic technology to identify potential grooming behaviours. By seeking out patterns or certain behaviours, the proposal endeavours to predict – and perhaps even stop – CSA before it happens.

This seemingly noble endeavour does not, however, withstand legal scrutiny. People cannot be detained for crimes that they have not committed or that they are not genuinely intending to commit. There must be a strong indication that a person is grooming a child for an investigation to be lawful, just as probable cause is required for law enforcement to acquire a warrant to search a suspect's home or devices. Yet probabilistic, AI-based predictive tools suffer from the same inaccuracies and false alarms as technology for identifying new CSAM, making them unreliable and unlikely to meet the CJEU's threshold.

The analysis of behavioural patterns also falls very short of probable cause. We agree with the specific conclusion of researchers at the UK Intelligence Agency that, when it comes to grooming detection, "[i]t is hard to envisage how such an algorithmic

probability of malicious activity with little supporting evidence could be used to convince a judge that the investigation was necessary".[56]

If this technology is used as the starting point for any investigation into a potential perpetrator of CSA – even if subsequent corroborating evidence is found – the defendant would be able to argue that the accusation against them has been made under unlawful procedural conditions.

There is therefore a real risk that suspicions based on the CSAR's grooming detection obligation will not support warrants, and may fail to lead to prosecutions. There is a further issue of the volume of false alarms that grooming technologies in particular will create.

The performance of natural language processing (NLP) models based on machine learning (ML) is such that error rates exist at several percent, even in state-of-the-art models. Such error rates cannot be deployable at the scale the CSAR proposes, and will instead overwhelm genuine cases with false alarms.

Aside from the technical and procedural concerns, we do not believe that technology is the right solution to online grooming. As explained by the Child Rights International Network (CRIN):

> **"The best defence against online grooming ... is informed and engaged parents who discuss the internet with their children from an early stage and can recognise the warning signs (such as emotional withdrawal), so that children feel able to report and discuss anything that has made them uncomfortable."[57]**

Whilst the European Commission claims that age verification is a way to ensure that the proposed grooming detection is targeted and legitimate, these purported safeguards are not adequate, and furthermore will open the door for other abuses and threats to online privacy and free expression. This will be explored in Chapter 3.3.

3.2.2.4 Children above the age of sexual consent

The content in scope of the CSAR uses the legal definitions from the 2011 CSA Directive. CSAM is defined by the Directive as material constituting "child pornography" (Article 2.c) or "pornographic performance" (Article 2.e). The Directive requires Member States to criminalise notably the production, acquisition, possession, deliberate access to and distribution of such material.

In Article 8 of the Directive, Member States are given the discretion not to criminalise material involving children who have reached the age of sexual consent, where that material is produced and possessed by those children for their own private use, and no abuse is involved.

This means that children above the age of sexual consent – an age limit which varies between Member States from 14 to 17 years – can distribute sexual images of themselves to close friends or their partner

(e.g. "sexting") without committing a crime. If the images are distributed to a wider circle of persons, for example by someone betraying the trust of the young person who produced the image, or by illegally hacking a child's device or account, that distribution will, of course, be criminalised.

This important distinction in the CSA Directive protects teenagers from sexual abuse, while allowing them to explore their sexuality, including in an online context.

Yet these exemptions for the private use of material by children above the age of sexual consent are not incorporated into the draft CSAR. This means that certain online activities by children, which are perfectly legal and furthermore important for sexual self-development and free expression (such as sexting using interpersonal communications services), will be in scope of the CSA Regulation because the material falls under the definition of CSAM in Article 2.l.

The material involved, which is legal as long as it is distributed privately between peers or stored on hosting services for the young person's personal use, will be subject to the provisions of the CSA.

Besides being reviewed by content moderators from the service providers and then reported to law enforcement, which in itself is highly intrusive and disruptive, the young people affected are likely to have their private communications blocked (in the case of detection orders) and may lose access to online communication services that they use. In the mandatory risk mitigation measures, providers are incentivised to prohibit any sexual images of children, even if they can be distributed legally for private use under the national law implementing the CSA Directive.

Solicitation of children is defined according to Article 6 of the CSA Directive, which requires the criminalisation of solicitation for sexual purposes of children below the age of sexual consent. The CSAR seems to expand the definition of solicitation so that it applies to any child, even if above the age of sexual consent.

For example, Article 7.7 of the CSAR states that detection orders for solicitation of children shall apply where one of the users is a child, meaning a person below 17 by the definition in Article 2.j. Private communications between, say, an 18-year-old and a 16-year-old will be subject to such detection orders for solicitation, even if the "grooming" activity is not criminalised in their Member State.

By tacitly ignoring the relevant exemptions for children above the age of sexual consent in the CSA Directive, the CSA Regulation effectively broadens its scope to cover material and activities which are legal under Member States' national laws.[58]

This may be an unavoidable consequence of the reliance on automated tools in the CSAR proposal, since automated tools, such as AI for detecting unknown CSAM, are generally unable to take the proper context into account (e.g. private distribution among child peers above the age of sexual consent; knowing which legal age(s) of consent apply, etc.).

As a result, the proposal will have wide-ranging consequences on young people's private communications, and unduly interfere with their normal discovery of sexuality in the course of human development, insofar as this involves online activities.

In the Impact Assessment (pp. 32-33), the Commission notes that differences between US and EU law on what constitutes CSAM, including the varying legal ages of consent across Member States, lead to many reports from NCMEC about material which is not illegal in the EU.

This is used to motivate the CSAR proposal and the creation of the EU Centre; but our analysis shows that precisely the same problems will apply to reports to the EU Centre.

54 https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html

55 https://ecpat.org/wp-content/uploads/2022/01/05-01-2022_Project-Report_EN_FINAL.pdf

56 https://arxiv.org/abs/2207.09506, p.20

57 https://home.crin.org/issues/digital-rights/childrens-right-digital-age?rq=digital%20age

58 **The material and activities in question will not be criminalised because of this, since criminalisation remains a matter for national law (implementing the CSA Directive) and not the CSAR. However, the material and activities will be in scope of the CSAR, which means they will be subject to requirements for risk assessment and mitigation, detection orders, reporting obligations, including ultimately to Europol and national law enforcement, all of which can be highly distressing and intrusive for the child, as well as detracting resources from investigating genuine CSA cases.**

3.3 Risk assessment and mitigation (Articles 3,4, 5 and 6)

3.3.1 What are the risk assessment and mitigation rules?

Hosting providers and interpersonal communications services (email, instant messengers, chat apps, etc.) shall identify the risk of use of their services for the purpose of online child sexual abuse (Article 3), take "reasonable mitigation measures" (Article 4) and report on both the analysis and measures taken to the Coordinating Authority (Article 5).

Providers of app stores shall analyse the risks for each app that they offer (Article 6 .1.a), based on those apps' risk assessment processes (Articles 3 and 4). They must also take "reasonable measures to prevent child users from accessing" services which have identified a high risk of solicitation (Article 6.1.b), and have age verification for their stores (Article 6.1.c).

Article 3.2.b lays out several criteria which will reduce the perceived risk: the existence of an explicit prohibition of online child sexual abuse (OCSA) in the platform or service's terms and conditions; measures to enforce that prohibition; age verification measures; and effective ways for end users to report suspected CSAM. Article 4.1 describes potential mitigation measures, including changes to the core technical and procedural elements of the platform or service, increasing internal supervision, or cooperating with other entities, including trusted flaggers as defined in the DSA. Safeguards include requiring measures to be "targeted and proportionate" to the risk and with "due regard" for fundamental rights (Article 4.2).

Where there is any risk of solicitation on the service identified in the risk assessment, the service "shall take the necessary age verification and age assessment measures to reliably identify child users on their services" (Article 4.3). Article 3 describes that the (common) ability to share pictures or videos by private message is one of the factors that create a risk of solicitation. Article 3.5 subsequently requires providers to consider whether there is "any remaining risk" that, even after they pursue mitigation measures as described in Article 4, OCSA could still happen on their platform.

3.3.2 Analysis of risk assessment and mitigation rules

Whilst there will likely be a benefit to providers considering and reasonably reducing the risk of OCSA on their platforms or services, the model as proposed in the CSA Regulation is dangerously broad, vague, and will likely incentivise generalised monitoring and the obligation to seek knowledge of the contents of communications, potentially in violation of the DSA, eCommerce Directive and CJEU case law (see Chapter 2.3).

Although Article 4.1 only requires providers to take "reasonable" measures, Article 3.5 requires that they must consider the existence of "any" remaining risk, which they must either address, or be liable to be served with a detection or removal order.

Since there is almost always a risk of a service being used for OCSA, all providers will be forced to take measures to mitigate risk, and may still face subsequent orders. The proposal states that the risk assessment must take into account "the manner in which the provider designed and operates the service" (Article 3.2.d).

This is very problematic for services that use end-to-end encryption to protect the information of their users. In these cases, the content is only accessible by the sender and the recipient of the transmission. Since the provider does not have access and is unable to monitor or interfere, this could be considered to be a high risk.

As a consequence, this could force providers to take measures to mitigate this risk by, for example, downgrading the security properties of their service. Whilst the risk assessment process does aim to provide anonymised statistics to support risk assessments in encrypted scenarios, there is no information about how information that would adequately portray the specific risk could be sufficiently anonymised.

For encrypted and unencrypted services alike, the proposed format of the risk assessment is furthermore problematic because it assumes that the service provider will generally monitor the behaviour of users in order to have sufficient knowledge to conduct an accurate risk assessment.

This may itself be in conflict with EU law, especially for interpersonal communications services subject to the ePD where any processing, other than the transmission of user communications, must be provided for by law (see Chapter 2.5). Whilst the GDPR might provide a basis for hosting services to undertake risk assessments, such a legal basis is not clear and could violate Recital 41 of the GDPR (see Chapter 2.8).

Providers are held responsible to decide which measures should be taken to address the risks identified. Given that they are legally liable, this should clearly be seen as a coercive incentive to resort to the heaviest and most intrusive measures, rather than incentivising the use of those which are respectful of rights to privacy

and data protection. What's more, there are very limited safeguards to protect against this: Article 4.2 sets out criteria that providers must meet. But there is no mechanism to check or verify whether the measures actually meet those criteria, nor what can be done if a measure is taken that is not proportionate or violates fundamental rights.

The criteria themselves are vague, limited and open to broad interpretation. What, for example, would count as sufficient "due regard" for fundamental rights? And how can risk measures be "targeted and proportionate" when providers are required to take actions across their entire platform or service? The proposal only stipulates an outcome that providers must achieve, with few checks on how they choose to do so. This may incentivise providers to take disproportionate actions, with no way for them to be held accountable for doing so.

3.3.3 Specific risks of age verification measures

The one place where there is specificity in the selection of risk mitigation measures is age verification. Article 4.3 of the proposal requires age verification to be performed for any hosting or interpersonal communications service where there is a risk of solicitation. In effect, this means every chat service, every instant messenger, and every e-mail service.

It becomes clear that the CSA Regulation could make age verification the reality for virtually every form of online communication when considering all of the following cumulatively: obligations for app stores to have age verification and to prevent under-18s from accessing apps with a purported high risk of solicitation (Article 6); the general incentivisation of age assessment across a range of services (Article 4); the requirement for services to know the demographic of their user base for risk assessment purposes; and the broad scope of the CSA Regulation. In the context of young people in situations of abuse, these measures could be seriously misguided and harmful. For survivors whose abuser is a family member, for example, removing their access to secure communication apps could increase isolation and deprive them of access to support.

As discussed in Chapter 2.2, the age verification industry already offers "solutions" which include the excessive use of people's biometric data. By definition, using biometric data for age verification will lead to the systematic processing of children's biometric data. This runs contrary to the work of child rights organisations like Defend Digital Me, which reminds us that children's biometric data are especially sensitive and must be treated with utmost care.[59]

Even when biometric data are not used, age verification still comes with risks. Firstly, age verification may require the user to verify themselves using an identity document. As is currently being explored in the negotiations for an EU law on digital identity wallets (eIDAS), making identity documents a precursor for internet access can have repercussions on people's privacy, data protection, non-discrimination and

other fundamental rights. Depending on the chosen architecture, the use of age verification can exacerbate manipulative surveillance advertising, create risks of security breaches, and lead to scope creep, whereby digital identity becomes a precursor to participation in social life and access to online services.[60]

When used to control access to digital communication, such age verification practices can effectively eliminate any potential to be anonymous, making the work of whistleblowers, journalists and human rights defenders harder, if not impossible. Such practices will also exclude anyone without the right identity documents.

This will be especially pronounced for those who face structural discrimination and exclusion, such as Roma and Sinti communities, homeless people, and undocumented people – including undocumented children – and anyone else that faces digital literacy or other barriers to accessing the latest technology. As undocumented people often have several minoritised identities, such age verification measures are also disproportionately likely to exclude people of colour, non-EU nationals, sex workers, and other minoritised groups.

3.3.4 Does the CSAR allow generalised scanning outside of detection orders?

The question of whether generalised detection (scanning) is possible under Article 4 of the CSA Regulation has already proven contentious. Some organisations, such as Thorn, have interpreted generalised scanning as being impossible under Article 4, but have called for the new rules to be amended to allow this, arguing that it would otherwise reduce the amount of scanning that is undertaken.[61]

This is consistent with public claims by Commissioner Johansson, including at the press conference to launch the CSAR, that no detection will happen outside of a detection order and that encrypted communications will not be undermined, except via an order.

In contrast, EDRi has outlined concerns that the proposed CSA Regulation is, at best, ambiguous on this question. First of all, the European Commission has confirmed that they envisage certain generalised scanning, for example of hosting services, to continue on the basis of the GDPR's legitimate interests provision. As already explained in Chapter 2.8, we have serious concerns about whether the basis of legitimate interests would truly be permissible.

What's more, many online services are designed specifically to avoid knowledge of the content shared by users on that service. As a result, being able to accurately perform the risk assessment required by them under Article 3 or 6 of the CSAR could amount to the unlawful general monitoring of interpersonal communications as well as of services within the scope of the GDPR, without a clear basis in the GDPR.

Thirdly, the Impact Assessment for the CSAR describes in more detail some of the Article 4 risk mitigation measures that the Commission foresees being used. In

particular, Article 4.1 refers to "technical measures" and "operation or functionalities of the service" that would influence the risk of OCSA. Based on our reading of Annex 9 of the impact assessment, it is hard to see how the "technical measures" and "operation of the service" could apply to anything other than a recommendation to use client-side scanning, which is a technical tool for scanning private encrypted communications.

As such, there seems to be a desire to use generalised scanning in encrypted environments even outside of detection orders. We see, therefore, a risk that the CSA Regulation may use Articles 3 and 4 to smuggle in forms of detection that *prima facie* do not seem to be within its scope.

Even if the text is clarified so that such general surveillance unequivocally cannot happen under Article 4, the issue of generalised scanning under the CSAR would not be resolved. This becomes apparent with the fundamental inability of detection orders to be targeted.

59 https://defenddigitalme.org/corporate-accountability/biometrics/

60 https://en.epicenter.works/document/3865

61 https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12726-Fighting-child-sexual-abuse-detection-removal-and-reporting-of-illegal-content-online/F3313611_en

3.4 Detection Orders (Articles 7-11)

3.4.1 What are detection orders and how will they be used?

The Coordinating Authority of each Member State will have the power to issue, subject to judicial or independent authorisation, a detection order requiring hosting providers and interpersonal communications services in their Member State to detect OCSA on a specific platform or service (Article 7). These orders can cover all three types of material: known, new and grooming (Article 10.1).

These orders can be served only when there is evidence of a "significant risk" of the service being abused, and the reason for such an order outweighs the negative consequences (Article 7.4). These orders are applicable for up to 2 years. Articles 7, 8 and 10 contain several provisions aimed at safeguarding and mitigating the risks posed by the Detection Orders, and Article 9 contains provisions for redress. Article 11 would allow the Commission to issue guidelines.

3.4.2 Analysis of detection orders

In practice, it will be possible for Member States to serve a detection order to any hosting provider or instant messaging platform in their country to monitor all information shared by their users.

As the logic that underpins the CSAR proposal is that not enough is being done to detect CSAM online, and detection is only supposed to happen under a detection order, this suggests that the European Commission would envision a high number of such orders being issued. Since it is impossible to differentiate between criminal content and legitimate content without analysing it, all content needs to be included in the assessment.

Such an order will thus constitute a "general monitoring obligation", which is unlawful (see Chapter 2.3).

We are concerned that the wording of "significant risk" (Article 7.4.a) to justify a detection order is misleading, as such a level of risk is deemed to exist when the mitigation measures in Articles 4 and

6 do not prevent the "likelihood" of the service being abused "beyond isolated and relatively rare instances" (Recital 21). In practice, this situation will become the rule rather than the exception for many services, since the risk of abuse is very hard to eliminate "beyond isolated and relatively rare instances". Most services are therefore likely to receive detection orders.

Detection orders may only be given when the reason for such an order "outweighs the negative consequences". Nothing in the proposal makes clear that undermining encryption would be considered such a negative consequence. The recitals, in particular 22, 23 and 26, are very unclear and grant no special protection to end-to-end encryption.

Whilst safeguards in Article 7 are supposed to provide some protections against this, they are close to meaningless: if the risk is limited to an identifiable part or component of the service, the order should only apply to that part or component (Art 7.8.a).

Given the risk identification structure, it is unlikely that the risk can be limited to an identifiable part or component of the service, and detection orders will therefore be general and indiscriminate (entire service) rather than targeted. The decisive factor for the orders is therefore not proportionality, but risk, which is broadly defined. According to the text, "effective and proportionate safeguards" need to apply only "where necessary" (Art 7.8.b) without any indication of what this means or who would decide what is necessary.

As such, the proposed detection orders should be considered specific only in terms of content and technologies, but not safeguards or scope, and will therefore frequently imply a form of generalised detection.

This contradicts the recommendation of the UN High Commissioner for Human Rights that any "interference with the private communications of individuals should only be carried out when authorised by an independent judiciary body and on a case-by-case basis".[62] By basing detection orders on risk, rather than on reasonable suspicion, they can never be genuinely case-by-case.

In the context of end-to-end encrypted services, the downgrading of the security properties of the service that would be required to fulfil a detection order is exceptionally dangerous and can never be targeted or proportionate. That is because end-to-end encryption relies on the technical integrity of the whole service.

If measures are built into a service to allow future access by providers or by law enforcement, there is no longer any technical guarantee of end-to-end confidentiality. Every person using that service will therefore be vulnerable to intrusion and potentially also hacking by state and non-state actors.

This would also violate Article 16 of the Charter: the freedom to conduct a business by effectively preventing digital services from providing secure, trusted communications services to users in the

EU.[63] It would also violate the consumer right to choose privacy-protective digital services.

Lastly, purported data protection safeguards for detection orders are clearly insufficient. Only in the case of an issuance of a detection order for grooming specifically, and if there is diverging opinion of the provider and/or the EU Centre, are a data protection impact assessment and a prior consultation procedure at the data protection authority required.

There are only minimal requirements for something to be considered "sufficiently reliable detection technologies". There is, for example, no requirement for a publicly-available and independent assessment of the reliability and effectiveness of the technologies applied.

3.4.3 Encryption and client-side scanning

One of the key debates surrounding the CSA Regulation so far has been whether it poses a threat to the integrity of end-to-end (E2E) encryption, whether it would force providers to use "client-side scanning" techniques, and whether those techniques are safe and respect fundamental rights.

The Commission has claimed that the CSAR is justified in including encrypted communications in its scope because of the threat that these services, which cannot be easily accessed by law enforcement, pose to investigations into CSA.

This framing of encryption overlooks the fact that it is a vital human rights tool, with organisations across the world emphasising that the security of people's private lives frequently relies on E2E encryption.[64]

The UN High Commissioner for Human Rights, for example, has emphasised the important role of E2E encrypted services for civilians trying to protect themselves and their families following the Russian invasion of Ukraine in 2022.[65]

Such services would likely be less safe and secure under the CSAR. Without E2E encryption, we lose confidence in our private communications, and without that, our ability to claim and enjoy practically all our fundamental rights becomes much more difficult and, in many cases, less safe.

Furthermore, the key claim of the CSAR proposal that there is "nothing police can do" to investigate evidence of serious crimes in E2E encrypted environments other than using detection technologies is not true.

Law enforcement authorities currently have more access to surveillance data than ever before, and the forthcoming EDRi paper "State access to encrypted data: A digital rights perspective" emphasises the importance of the protection of fundamental rights when undertaking state hacking.[66]

Read in conjunction with EDRi's "10 principles to defend children in the digital age", it is clear that law enforcement has many methods at their disposal to pursue criminals, even those that may misuse E2E encrypted services, without the CSAR.[67]

3.4.4 Technical assessment

Our analysis of the CSAR shows that it would dangerously undermine E2E encryption by compelling even providers who offer E2E encryption to lower the security of their service in order to be able to conduct any detection. This is a fundamental point of how E2E encryption works, and cannot be improved with developments in the accuracy or efficiency of technological tools like CSS.

The Impact Assessment shows that the Commission clearly intends CSS techniques to be employed by providers offering E2E encrypted services. Claims by Commissioner Johansson that if there is no available technology then providers would not be forced to use something that does not exist appear to be disingenuous; the Commissioner and staff in DG HOME have made it clear that they consider CSS to be a safe and viable technique for the EU Centre to make available for providers to comply with the CSAR's detection orders.[68]

This contradicts the advice of the Commission's own technical expert group in preparing the CSAR proposal, as well as cybersecurity experts around the world.[69] In the Impact Assessment to the CSA Regulation, the Commission's expert group makes high-level comments on the key considerations for several potential methods of scanning for known CSAM in E2E encrypted environments. After an initial assessment, the experts selected their top three most "promising" detection methods, all of which are forms of CSS:[70]

- **The experts suggest that "On-device full hashing with matching at server" (a type of client-side scanning, or CSS) is a viable option, despite assessing its protection of privacy as "medium-low" and its secureness (including its resilience to abuse by malicious actors) as "medium-low". Thus, despite assessing that this method of CSS is neither privacy-respecting nor secure, the expert group recommended the immediate roll-out of this technique (p.310);**

- **The group also looks favourably on "On-device partial hashing with remaining hashing and matching at server" (another CSS method), calling its feasibility "medium" despite only ever having been piloted as a proof-of-concept, and as such, there is no evidence that such a solution could be effectively scaled up. What's more, the protection of privacy and level of secureness for this method are both listed as "medium-low";**

- **The third purportedly viable option is "Secure enclaves in ESP server". Not only do the experts explain that the privacy and security of this technique are "medium-low", but also that feasibility is "medium-low".**

All of these CSS proposals involve the on-device analysis of data before being encrypted or after being decrypted. Despite showing that they all suffer from serious privacy and security risks, this assessment has underpinned the claim from the European Commission that detection in E2EE environments is safe, secure and respects fundamental rights. The UN High Commissioner for Human Rights, however, classifies CSS as a technology with the potential to endanger fundamental rights:

> **"Client-side scanning also opens up new security challenges, making security breaches more likely. The screening process can also be manipulated, making it possible to artificially create false positive or false negative profiles."[71]**

CSS breaks the whole purpose and function of end-to-end encrypted communication, which is the assurance of confidentiality against the service provider.

Looking at Apple's 2021 proposal for on-device scanning, for example, reveals that the technical implementation of CSS is done in a way that prevents the user from removing it. In effect, this means that surveillance software is hosted on mobile devices which are often vulnerable to "zero days exploits" (unmitigated software vulnerabilities) and can therefore be abused by malicious actors.

These actors can include not just cyber-criminals, who infect user devices to subvert them for criminal purposes, but also local opponents such as abusive partners, who increasingly use technological surveillance as tools of control and abuse.[72] Every device subject to CSS will therefore be made technically much more vulnerable to attacks and hacking.

62 https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report, para 57.a

63 https://www.patrick-breyer.de/wp-content/uploads/2021/03/Legal-Opinion-Screening-for-child-pornography-2021-03-04.pdf

64 **For example:** https://www.fightforthefuture.org/news/2022-10-13-make-dms-safe-orgs; https://www.hrw.org/tag/encryption; https://edri.org/take-action/our-campaigns/keep-it-secure/

65 https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report

66 **It will be available shortly after the publication of this paper at:** https://edri.org/our-work/?category=position papers

67 https://edri.org/our-work/chat-control-10-principles-to-defend-children-in-the-digital-age/

68 **For example,** https://ec.europa.eu/commission/commissioners/2019-2024/johansson/blog/children deserve-protection-and-privacy_en

69 **For example,** https://arxiv.org/abs/2110.07450; https://privacyinternational.org/sites/default/files/2022-09/SECURING%20PRIVACY%20-%20PI%20on%20End-to-End%20Encryption.pdf

70 **Impact Assessment can be downloaded at** https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12726-Fighting-child-sexual-abuse-detection-removal-and-reporting-of-illegal-content-online_en

71 https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report

72 https://tech.cornell.edu/news/how-domestic-abusers-use-smartphones-to-spy-on-their-partners/

3.5 Case study - Ireland

Newly-released information from EDRi-affiliate the Irish Council for Civil Liberties (ICCL) provides evidence to support our concerns about the proportionality and lawfulness of the CSA Regulation.

In a case study published for the first time alongside this position paper, we demonstrate the difference between claims of accuracy and effectiveness made by the European Commission compared to reality, as well as the tangible threat that the scanning of private communications poses. Specifically, this case study evidences low rates of CSAM being identified by current scanning tools, high levels of false positives, and the potentially illegal retention of innocent people's data by police.

3.5.1 Background

In early 2021, ICCL requested information from the Irish police force, An Garda Síochána, about the referrals (reports of suspected child sexual abuse material) it receives from NCMEC. In October 2021, ICCL received answers from the force's Online Child Exploitation Unit (OnCE). ICCL has since asked further questions to clarify the responses and still awaits a reply.

The Irish police explained that they have received referrals from NCMEC since 2015, and from UK authorities and the FBI between 2010 and 2015. They confirmed that the number of referrals received per year varies: 2848 referrals in 2017; 6812 in 2018; 3888 in 2019; 4192 in 2020 and, as of October 2021, approximately 3500 referrals had been received.

3.5.2 Low accuracy and high rates of false alarms

As discussed at length in this paper, laws that restrict people's fundamental rights need to demonstrate their legitimacy and lawfulness, including by proving that proposed measures are necessary and proportionate. One of the criteria that can be used to assess necessity and proportionality is whether the law can be effective in achieving its stated goal(s).

Of the 4192 reports that the Irish police received in 2020, they told ICCL that 852 reports (20.3%) actually turned out to depict CSAM. ICCL also asked them about the number of referrals that lead to criminal investigations per year. They replied: "Using 2020 as an example, a total of 4192 referrals were received

from NCMEC. 409 of these referrals were actionable [9.7%], and from those referrals 265 files were completed [6.3%]."[73]

The Impact Assessment to the CSA Regulation states that the accuracy of scanning technology for known CSAM is 99%, with Commissioner Johansson publicly stating that false positives occur at a rate of 1 in 50 billion for known CSAM technologies, and at less than 0.1% for new CSAM.[74]

In August 2022, the Commissioner increased her accuracy claim for known detection to 99.9%.[75] These technologies are currently in use in the EU under the temporary ePrivacy Derogation and contribute to a substantial number of NCMEC reports.[76]

Although we do not have sufficiently granular data to know exactly how many of the 2020 reports to the Irish police were from scanning technologies (rather than user reports, for example), the limited publicly-available information shows that across Big Tech services, CSAM is detected "proactively" (i.e. by scanning technologies), not through user reports, in the vast majority of cases actioned (Meta puts this figure at 99.1% and Microsoft at 99.4%).[77]

This does not translate directly to NCMEC reports, but it confirms a high prevalence of automated scanning, meaning that we can reasonably expect a large proportion of the reports that NCMEC sent to the Irish police to relate to content automatically flagged by scanning tools.

It is evident, therefore, that the accuracy claims made by Commissioner Johansson technically cannot mean that 99.9% of the content that PhotoDNA or other scanning technologies flag as being CSAM actually turn out to be CSAM. In Ireland, only 20.3% of 2020 reports contained CSAM, with less than a third of those genuine reports subsequently being "completed".

Of those that were "actionable" or "completed", the OnCE said that they do not retain information on how many investigations have led to prosecutions, nor the number of prosecutions that have led to convictions.[78]

These figures also highlight a high number of false positives: "OnCE doesn't use a specific categorisation of non-illegal. A total of 471 were marked as being not Child Abuse Material in 2020 from a total of 4192. [...] 940 referrals included IP addresses which could not be progressed further. [...] 606 were marked as below the threshold. 75 were self-generated. 333 were marked as viral. 51 were adult."

The consequences of this are significant: at least 471 reports (greater than 1 in 10), but probably far more, were false positives, and at least 940 reports (greater than 1 in 5) did not contain information that would allow police to pursue them further.[79] Based on available data, we can reliably assume that a significant number of these false alarms came from scanning technologies. It is not mathematically possible, therefore, for false alarms generated by scanning technologies in use in the EU to be limited to 1 in 50 billion.

When asked to detail the general nature of the content which triggers these false referrals, the police replied that "OnCE will not action a referral further for a number of reasons on the basis of its content, the following are examples: children playing on a beach, topless content, nudist, adult content, etc."

This proves that these legitimate pieces of content are already wrongfully reported as CSAM. Dealing with such false alarms can also take already scarce resources away from investigating genuine cases. As the Dutch police have already indicated, expected levels of both genuine and false reports under the CSA Regulation will be at a scale that is not possible for the police to handle.[80]

In sum, these figures contradict the European Commission's claim that scanning tools are so robust that there is no need to worry about accuracy, reliability or false positives. It also emphasises what Netzpolitik has called the "grossly distorted picture" created by the figures put forward by the European Commission when talking about the scale of CSAM.[81]

This same issue is clearly present in what the Commission refers to as the "accuracy" of scanning tools, which we have shown cannot be conflated with the (significantly lower) number of reports that actually contain CSAM.

At the core, this case study helps show that scanning technology is neither sufficiently accurate nor demonstrably effective at identifying CSAM, and a clear link between reports and convictions has yet to be demonstrated.

3.5.3 Violations of free expression, data protection and the presumption of innocence

When asked what the Irish police does with the IP addresses and identifying information pertaining to NCMEC referrals after the content is found to be innocent, they confirmed that the following data relating to all NCMEC referrals are retained: NCMEC CyberTip number, date received, suspect email address, suspect screen name, suspect IP address and reporting Electronic Service Provider.

However, the force admitted in its response to ICCL that, having spoken with the Garda Data Protection Unit:

> **"there may be no legal basis to retain data relating to (1) suspect email address, (2) suspect screen name, (3) suspect IP address in the first place in cases which are clearly not child abuse material – for example referrals involving images and videos of children playing on a beach as it may not be proportionate to do so."**

The police also confirmed that they would seek a legal opinion on the practice of retaining personal data pertaining to innocent people wrongly flagged as suspect sharers of CSAM online.

These facts show that innocent people's highly intimate data are retained in police databases, despite the police knowing they are innocent.

The implications are manifold: it emphasises our concerns that legitimate free expression will be criminalised under the CSAR. It shows that not only do false alarms exist at a much higher rate than the European Commission claims, but further that these false alarms have already led to violations of people's rights to data protection and the presumption of innocence.

3.5.4 Consequences for the CSA Regulation

This case study emphasises that CSAM scanning technologies cannot be as accurate or reliable as the European Commission claims, and that false alarms are a present and serious issue, with major civil liberties consequences.

Furthermore, it emphasises the lack of evidence that scanning practices are effective. As such, this case study clearly questions the legitimacy, necessity, proportionality – and therefore lawfulness – of the CSAR's proposed measures.

73 **ICCL has asked An Garda Síochána to provide an explanation or definition for the terms "actionable" and "completed" but so far they have not provided this.**

74 https://multimedia.europarl.europa.eu/en/webstreaming/committee-on-civil-liberties-justice-and-home-affairs_20221010-1500-COMMITTEE-LIBE

75 https://ec.europa.eu/commission/commissioners/2019-2024/johansson/blog/children-deserve-protection-and-privacy_en

76 **87% of NCMEC reports in 2021 came from Instagram and Facebook:** https://netzpolitik.org/2022/ncmec-figures-explained-how-the-spectre-of-millionfold-abuse-haunts-european-policy-makers/, **both of which deploy this scanning technology on their messages:** https://transparency.fb.com/sr/eu-csam-derogation-report-2022

77 https://transparency.fb.com/data/community-standards-enforcement/child-nudity-and-sexual-exploitation/facebook/#proactive-rate; https://www.microsoft.com/en-us/corporate-responsibility/digital-safety-content-report?activetab=pivot_1%3aprimaryr3

78 **This is consistent with UK figures where in 2021, 102 000 of NCMEC's 29 million reports were passed on to UK law enforcement. That year, the number of UK arrests for alleged CSA was a far smaller 6500, with no evidence of how many of those arrests were a result of the NCMEC reports, nor how many led to convictions. Intelligence agencies hope to establish a causal link between reports and arrests but so far this has not been established. Source:** https://arxiv.org/abs/2207.09506. **Note that while we find many statistics in this paper to be useful, we strongly disagree with many of its assertions, particularly those relating to privacy and lawfulness, which the authors note is outside the scope of their paper (despite the fact that they allege to draw conclusions about privacy and lawfulness).**

79 **The true number of false positives is likely to be significantly higher than 471, as the police have not confirmed what categories such as "below the threshold" mean, nor whether the adult content was included in the figure of 471. We therefore estimated false positives at the most conservative rate based on the available data.**

80 https://debatgemist.tweedekamer.nl/node/29579

81 https://netzpolitik.org/2022/ncmec-figures-explained-how-the-spectre-of-millionfold-abuse-haunts-european-policy-makers/

3.6 Reporting Obligations (Articles 12 and 13)

3.6.1 What are the CSAR's reporting obligations?

The CSAR obliges hosting providers and interpersonal communications services to report OCSA of which the provider is made aware to the new EU Centre (Articles 12 and 13).

3.6.2 Analysis of reporting obligations

The CSAR proposal largely follows the current US framework for reporting obligations of service providers.

When providers of hosting services and interpersonal communications services become aware in any manner (other than through a removal order) of any information indicating online sexual abuse, they must promptly submit a report to the EU Centre (Article 12). The threshold for reporting in Recital 29 (reasonable grounds) seems rather low, requiring that doubt about the potential victim's age would specifically not prevent the provider from submitting reports.

Many of the reports will originate from automated detection tools deployed by providers, either through the execution of detection orders or voluntary scanning measures. Therefore, a very large number of reports to the EU Centre, including false alarms, must be expected.

Concerns have been raised that US-based companies will be legally unable to report CSAM to the EU Centre, due to US laws limiting the onward sharing of CSAM (even for reporting purposes). As such, the current requirement for service providers to report CSAM seems infeasible in many cases. What's more, the DSA already requires providers to act on the basis of any illegal content about which they become aware, again emphasising our necessity concerns.

A better approach would be to reconceptualise "reporting" in terms of what will actually help survivors. As Chapter 4 will explore, no evidence has been put forward to demonstrate that reporting suspected CSAM to the EU Centre will accelerate its removal or increase the likelihood of a prosecution.

On the contrary, the complex and bureaucratic provisions of the CSAR could make both of these factors harder.

By looking at what a meaningful "report" would be from the perspective of a survivor, child protection organisations point to several much more meaningful and effective options.

According to EPCAT and the WeProtect Global Alliance, this includes ensuring that young people are sufficiently educated about how to report abuse, that they have trusted adults to help them navigate this difficult process, that they can be guaranteed privacy and anonymity in the case of legal action, and that police do not treat them as if they themselves are criminals.[82]

3.7 Removal Orders (Articles 14 and 15)

3.7.1 What are the CSAR's requirements for removal of CSAM?

Under Article 14 of the CSAR, removal orders can be issued by a national judicial or administrative authority (as requested by the Coordinating Authority) but only against hosting service providers.

Removal orders are designed to remove specific pieces of content (identified with a URL and any additional information necessary).

Once the provider receives the order, the CSAR requires that it should act against it within 24 hours. Article 14.3 details that removal orders should include key information such as the identification of the Competent Authorities, name of the provider, specific uniform resource locator (URL) and information about the redress available.

There are redress measures in Article 15. This includes a right to information for users who provided the material, although this may be strictly limited (following Article 15.4).

82 https://ecpat.org/wp-content/uploads/2022/01/05-01-2022_Project-Report_EN_FINAL.pdf

3.7.2 Analysis of removal orders

Removal orders, which were created for the TCO Regulation (see Chapter 2.7), are designed to ensure that Competent Authorities can require the removal of specific pieces of content. Contrary to the mass scanning of content via upload filters in hosting services or client-side scanning for chat messages, removal orders can in theory be necessary and proportionate (in this case, targeted) actions if, and only if, the necessary safeguards are met.

In the CSAR, the Coordinating Authority in the Member State where the hosting service provider is established or represented can request competent judicial or administrative authorities to issue an order to the service provider to remove access (in all Member States) to one or more specific items identified as CSAM.

Since judicial authorities include prosecutors, which are not independent authorities in most Member States, we argue that such orders should only be issued by a court.

For the TCO Regulation, the International Committee of Jurists commented about the risks resulting from the lack of independence *"leading to excessive, arbitrary or discriminatory interference with the freedoms of expression, religion, assembly and association online as well as with rights to privacy and data protection."* [83]

The time-frame for removal (derived from the final TCO Regulation) may be sufficient for most big corporations. However, the CSAR notes that some flexibility, for example regarding the capacity of smaller providers or in case of force majeure, may be accepted in certain cases to ensure that they have the capacity to remove the content as soon as possible (Article 14.5).

The redress mechanisms in Article 15 include the right to challenge a removal order before the courts of the Member State of the competent authority. This is a necessary safeguard for cases where CSAM may have been wrongfully identified.

[83] https://www.icj.org/european-union-independent-judiciary-and-effective-remedies-must-be-at-the-core-of-the-eu-regulation-on-terrorist-content-online-warns-icj/

3.8 Blocking Orders (Articles 14 to 18)

3.8.1 What are the CSAR's requirements for blocking content?

Under Articles 16, 17 and 18 of the CSAR, internet service providers (ISPs) can be required to block access to websites containing CSAM.

These orders can be issued by a national judicial or administrative authority (requested by the Coordinating Authority) and are envisaged to be used to block EU access to content that is not subject to other rules in this proposal (e.g. because it is hosted in a third country with no EU presence).

The specific rules for such blocking are elaborated in Article 17, and redress measures in Article 18.

3.8.2 Analysis of blocking orders

Article 16 refers to CSAM hosted outside the EEA and that is hosted by service providers that refuse to take it down.

This is an uncommon situation; research indicates that CSAM typically is hosted within the EEA or the USA, and that service providers take most CSAM down within seven days of being notified.[84] This calls into question the contribution that blocking orders would provide to combating the dissemination of CSAM, even as a measure of last resort, given that the vast majority of CSAM is distributed through hosting services in countries with advanced infrastructures and a well-developed rule of law (namely, the United States and Netherlands).

Blocking orders are issued at a Universal Resource Locator (URL) level.[85] Blocking at the URL level has the notable advantage of ensuring that orders can be targeted to specific material which has been identified by Competent Authorities as illegal (CSAM). In theory, this can significantly reduce, and in principle eliminate, the risk of over-blocking (blocking legal content).

However, today almost all internet traffic is encrypted when transmitted between the end-user requesting it and the server delivering it (HTTPS for web traffic). For HTTPS and other encrypted internet traffic, it will not be technically possible for the IAS provider to execute blocking orders at the URL level. IAS providers cannot deploy detection technologies at the device level of end-users, since internet access does not take place through a specific app controlled by the IAS provider.[86]

In short, there is no possible way for the IAS provider to circumvent encryption because they do not control the encryption, but simply transmit internet packets (encrypted or not).

Given this pervasive use of transport-level encryption (e.g. HTTPS), blocking orders will have very limited value for Coordinating Authorities due to the general impossibility of ISPs to implement them.

With HTTPS, it would only be possible to block access to illegal content at the website (domain) level, which immediately raises the issue of over-blocking, since the order would effectively cover all URLs, present and future, pointing to that website. It is highly unclear whether the blocking measure in Articles 16-18 can be used at the domain/website level because this requires a proportionality assessment not foreseen by the proposal, where the blocking order is targeted to specific CSAM.

Blocking at the domain level would affect legal content at the hosted website, which in many cases will constitute a disproportionate interference with freedom of expression and access to information. Additionally, the blocking orders in Articles 16-18 are meant for situations where CSAM practically cannot be taken down, neither with a formal removal order nor with a request to the hosting service provider ("referral").

There is a risk, however, that Coordinating Authorities (or other Competent Authorities) may prematurely resort to blocking orders because they are generally easier to handle than removal procedures. Blocking orders only require contact with the domestic IAS providers (besides the independent judicial authorisation), whereas removing CSAM could require more cumbersome cooperation with authorities in other countries.

While IAS providers could technically implement blocking orders at the domain level, such blocking is not effective, as there are numerous methods of circumvention.

Removing CSAM from the servers where it is hosted is the only effective way of preventing access and further online distribution of the illegal content, yet blocking orders could make it less likely that Coordinating Authorities will pursue the more procedurally complex removal orders.

84 https://inhope.org/media/pages/articles/annual-reports/8fd77f3014-1652348841/inhope-annual-report-2021.pdf, p.15

85 **An example of a URL is a webpage address, but elements on a webpage, for example an image, can have their own URLs.**

86 **If the IAS provider deploys a "man-in-the-middle attack" on HTTPS traffic in order to inspect the URLs accessed and block certain URLs, the connection to the website will be rejected (with a certificate warning) by the user's browser, because the IAS provider cannot present a valid certificate for the domain name. Public campaigns to combat online fraud and promote good cybersecurity practices advise users never to ignore these certificate warnings and proceed to the insecure website. Web browsers also make it increasingly hard for users to proceed to a website with a certificate warning. For these reasons, we consider it technologically impossible for IAS providers to block URLs when HTTPS is used, because any attempt to do so would literally break (destroy) security on the entire internet.**

3.9 The EU Centre

3.9.1 What is the EU Centre?

One of the key features of the CSA Regulation is the creation of an independent EU hub, the "EU Centre", which would, in theory, partially replace EU law enforcement's reliance on the US-based NCMEC. It would also provide a triage role for reports of potential CSAM from providers (Article 48.1) and manage the list of available detection technologies (Article 50.1).

A large portion of the CSAR is dedicated to the procedural establishment of the EU Centre, including provisions on its Management and Executive Boards (Articles 56-63) as well as operational elements, such as the fact that it will share administrative functions like HR staff and IT infrastructure with Europol (Article 53).

3.9.2 Analysis of the EU Centre

We have serious doubts about whether the EU Centre model, at least in its current form, can be effective, and raise concerns that it is not sufficiently independent from Europol. What's more, given the technical issues with the detection of known CSAM – and especially so for new CSAM and grooming – and the fact that providers are not required to conduct a human review prior to submitting reports, the scale of reports to the Centre would likely be unmanageable.

The EU Centre, an administrative authority, is tasked with assessing reports from providers before forwarding them to Europol (which in itself is cause for concern, given that there is no explicit basis for Europol to receive these data) and the relevant national law enforcement agency for action.

However, the criterion in Article 48.3 for forwarding reports from the EU Centre to law enforcement is very broad since all reports that are not "manifestly unfounded" must be forwarded to Europol and law enforcement authorities of the relevant Member State. Recital 65 explains that this covers all reports where it is not immediately evident, without any substantial legal or factual analysis, that the reported activities do not constitute OCSA.

It is not clear why the EU Centre cannot perform a more detailed analysis before forwarding reports to law enforcement, especially as the very purpose of the EU Centre is to function as a civilian agency

between users and service providers on one hand, and law enforcement agencies on the other.

The need for such an independent intermediary entity becomes all the more important given that reports can be submitted from service providers based on detection technologies with high error rates, as described in Chapter 3.2. Furthermore, there are no sufficient mechanisms in the proposed CSAR to ensure a high quality of reports to the EU Centre, nor to ensure that the reports subsequently passed on to national law enforcement are of high quality.

On the contrary, Dutch police have confirmed that they do not currently believe they would be able to handle the volume of reports that they would receive under the CSAR.[87]

The EU Centre will not rely on its own, arguably superficial, assessment of the illegality of the reported material for the task of building databases of indicators for identifying future CSAM. These databases shall be updated solely on the basis of material which has been identified, after a "diligent assessment", as CSAM by Coordinating Authorities or other Competent Authorities (Article 44.3).

This creates a complex information flow: the EU Centre receives reports from providers, forwards them to law enforcement authorities unless manifestly unfounded, and then receives them again from national Coordinating Authorities if the material is confirmed to constitute CSA. It is, of course, positive that the databases of indicators are only updated based on properly validated information, but the same high standards should apply to reports from service providers before they are forwarded to law enforcement.

Instead, the proposal implicitly seeks to maximise the amount of information submitted to law enforcement (including Europol) from providers. The number of reports forwarded to national police forces is likely to exceed any reasonable law enforcement capacity for investigation and prosecution, but the information forwarded will likely be retained in Europol and/or national law enforcement databases (as demonstrated in the case study on Ireland in Chapter 3.5), and in some cases possibly subjected to intrusive data-mining analysis by predictive policing systems.

Another purported aim of the EU Centre is to develop an approach to tackling CSAM that reflects EU rights and values rather than being reliant on US entities and the interests of private companies. However, only Thorn and Microsoft currently have the relevant technology as defined in the Impact Assessment, and so we believe that the EU Centre will have no choice but to employ technology from one or both of them.

Thorn is a US-based organisation functioning both as a not-for-profit and as a commercial entity, both selling and providing their scanning software free of charge. In 2022, Thorn came under fire by Netzpolitik, which revealed that Thorn has lobbied extensively on the CSAR proposal

and stands to make significant financial gains if the use of scanning technology becomes mandatory in the EU.[88]

There is also a concern that the EU Centre could disempower or even displace the work of hotlines. Hotlines are the network of organisations across EU Member States (and the world) working on the front lines to remove CSAM from the internet, support survivors, and undertake other crucial tasks.

Many hotlines rely on the EU for vital funding, which has become increasingly precarious and unpredictable in recent years. Hotlines are very important entities because they have vast amounts of national context and expertise – something that the EU Centre will lack. Yet the EU's network of hotlines does not currently have a specific legal basis at the national or European level, despite the sensitivity of the content with which they work; this is something that the CSAR fails to address. It further risks disempowering hotlines through its centralised model:

> **Centralised models for the removal of content from online platforms and services – and indeed as foreseen for the EU system – can add 6 weeks to the time it takes to remove abuse content from the internet compared to decentralised approaches (like private-sector contractors or certain types of hotlines, which currently report swift takedown).[89]**

87 https://debatgemist.tweedekamer.nl/node/29579

88 https://netzpolitik.org/2022/dude-wheres-my-privacy-how-a-hollywood-star-lobbies-the-eu-for-more-surveillance/

89 https://www.lightbluetouchpaper.org/2022/05/11/european-commission-prefers-breaking-privacy-to-protecting-kids/

3.10 National Authorities

3.10.1 What role do national authorities play?

Along with the EU Centre, the CSA Regulation creates "Competent Authorities" (Article 25): judicial or independent administrative bodies designated by each Member State to carry out the application and enforcement of the CSAR.

One of the national Competent Authorities in each Member State will be nominated as the single point of contact, or "Coordinating Authority", for their country (Article 25.2), and the entity with ultimate enforcement responsibility. Competent Authorities can complement their activities with voluntary requests to certain service providers to remove CSAM (Article 32).

The Coordinating Authority must be an independent administrative authority (Article 26) and it has significant powers, such as ordering the cessation of infringements of the CSA Regulation and imposing administrative fines.

Under Article 31, Coordinating Authorities will be permitted to search through the content of hosting providers, presumably at a large scale. They have jurisdiction over all providers established in their territory (respecting the country of origin principle), and there is no ability under the proposal for cross-border enforcement.

3.10.2 Analysis of national authorities

Along with the EU Centre, the creation of National Authorities is designed to coordinate enforcement of the CSA Regulation. However, the complexity and size of this system raise similar questions about whether this will be an efficient and effective way to remove CSAM from the internet. The Dutch child protection hotline EOKM, for example, finds that the most effective way for them to deal with CSAM in the Netherlands is to remove it as soon as it has been reported to them.[90]

Furthermore, the country of origin principle could also make the CSAR significantly more difficult to enforce. Since the GDPR entered into force in 2018, the Irish Data Protection Commission (DPC) has received criticism for its sluggishness in dealing with data protection investigations.[91]

Ireland receives a very high number of data protection complaints because many of the large tech companies offering online services in the EU are registered in Ireland. This would suggest that a large number of investigations for the CSAR's Coordinating Authorities would fall on the Irish authority. The same goes for the Netherlands, where the vast majority of the EU's hosting providers are based. Ireland and the Netherlands would therefore bear the brunt of the CSAR's enforcement, which is likely to significantly slow down enforcement.

The CSAR proposal gives no reassurance for how Ireland and the Netherlands would be motivated to deal briskly with the enormous administrative burden that the CSAR would place on them on behalf of the entire Union, leading to concerns about the ability of the system to be effective in practice.

If we compare the multi-year backlog of data protection cases at the Irish DPC to the case of child sexual abuse material, we foresee a situation where the CSAR would leave abuse imagery online for years, ultimately failing survivors and allowing for their re-victimisation. Lastly, whilst the Coordinating Authorities are tasked with performing a balancing test when issuing orders as an attempted safeguard, their child protection mandate creates concerns about whether they will be able to perform an impartial balancing test.

Investigations from journalists show that today, the failure to protect children is comprised of several factors, such as a lack of capacity from law enforcement and judicial bodies to deal with the volume of abuse cases and imagery that are reported to them, and failures of public and private actors to properly respond to CSAM (e.g. to remove it) after they are notified about its existence.

In a recent case in Germany, for example, investigative reporters found 20 terabytes of CSAM that had remained online for years because, despite knowing about its existence and potential to be further disseminated, police stated that they did not have the "human resources" to remove the material.[92] The CSAR will likely make such situations even worse.

In the Netherlands, there are a few so-called "bullet-proof" hosters. These providers do not act on valid reports of material of sexual abuse of children on their servers, not even when made by the Dutch hotline. Law enforcement has the power, but seems to lack the capacity to take down this material.[93]

There are known cases in the Netherlands where the police took such a long time to pick up a case that it became nearly impossible to do a meaningful investigation.[94] As a result, these cases get shelved, with the offender going

unpunished. Again, ensuring that law enforcement has sufficient resources at their disposal is a far less intrusive and more effective measure than generalised surveillance.

Rather than tackling these serious issues and enabling more thorough investigatory work by law enforcement, the CSA Regulation will exacerbate existing problems by vastly increasing the number of reports to platform moderators, the EU Centre, and law enforcement. As we have already explained, many of these reports will be erroneous.

90 https://www.weprotect.org/wp-content/uploads/EOKM-Annual-report-2021.pdf

91 **Among many others:** https://noyb.eu/en/irish-dpc-burns-taxpayer-money-over-delay-cases

92 https://www.tagesschau.de/investigativ/panorama/kinderpornografie-loeschung-101.html

93 https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2020Z18360&did=2020D39646

94 https://www.nrc.nl/nieuws/2022/05/05/slachtoffers-zedendelict-vinden-weinig-steun-bij-justitie-en-politie-2-a4122859 and https://www.nrc.nl/nieuws/2022/05/05/na-jaren-deed-ze-aangifte-maar-haar-zaak-belandde-op-de-plank-a4122794

4. Conclusions and Recommendations

4.1 All pain for little gain? The social implications of the CSAR

4.1.1 Efficiency concerns

As discussed in Chapter 2, one of the key criteria for assessing the necessity, and therefore lawfulness, of the proposed CSAR, is whether its methods will be effective in achieving their aim.

As highlighted throughout this analysis of the CSAR's key articles, it is clear that not only are the proposed measures unlikely to be effective, they are in fact more likely to be counterproductive.

As already established, the Commission has not put forward evidence that "voluntary" scanning and increased detection under the Temporary Regulation has led to increased access to justice for survivors. Rather than tackling existing serious issues or enabling more thorough investigatory work by law enforcement, the CSAR will increase the volume of reports, including many false alarms, to the EU Centre and law enforcement authorities – without helping them to do their jobs more effectively. This may obfuscate genuine cases of CSA amongst the high number of false alarms, in particular for national law enforcement that is already systemically under-resourced and over-burdened.

And the complex system of orders and authorities at best lacks evidence of effectiveness; at worst, it is likely to hamper existing efforts to tackle CSAM that have shown to be effective, such as the removal of CSAM by national hotlines.

What's more, Europol and the Commission warn that some child abuse and exploitation networks are sophisticated cyber-criminals.

If this is the case, they would be well-equipped to move from conventional message services, such as WhatsApp or Signal, or conventional hosting services, to their own messaging services, Tor hidden services (the so-called dark net), or overseas services that will not comply with the CSAR. They would also be the most equipped to trick systems such as client-side scanning, which the next section shows can be manipulated to deliberately generate false negatives (to evade detection of CSAM) or false positives (to deliberately generate a malicious false alarm).

Our analysis of the CSAR's key articles thus strongly supports the European Data Protection Board and Supervisor's conclusions that the CSAR will have regrettably little effect on stopping child sex offenders, but a significantly negative effect on society at large.[95]

The UN High Commissioner for Human Rights adds that measures like the CSAR could "choke" the development of "vibrant, pluralistic democracies".[96]

4.1.2 Privacy and safety are mutually reinforcing rights

As emphasised throughout Chapter 3, the CSA Regulation takes an unrealistic, technosolutionist approach which threatens the most basic principles and practices of digital security.

It shows a fundamental misunderstanding of technology, for example the infeasibility of URL blocking in today's digital environment (see Chapter 3.8) or the issue of inaccuracy and false alarms of scanning technologies (Chapter 3.3).

But arguably the most profound technological error put forward by the Commissioner is the claim that it is possible to safely and securely scan for CSAM (or any other content) in end-to-end encrypted services, most likely using client-side scanning (CSS) techniques.

For these reasons, it has been argued that proposed detection orders may compromise people's personal digital devices to the extent that this would entail an impermissible violation of the very essence of the right to privacy as enshrined in Article 8 of the Charter. [97]

What's more, privacy and security are mutually-reinforcing rights, which can have an impact on a wide range of fundamental rights as discussed in Chapter 3.2.

[95] https://edps.europa.eu/press-publications/press-news/press-releases/2022/combat-child-sexual-abuse-online-presents-serious-risks-fundamental-rights_en

[96] https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report

[97] https://europeanlawblog.eu/2022/06/07/does-monitoring-your-phone-affect-the-essence-of-privacy/

The UN High Commissioner for Human Rights explicitly warns that regulations which undermine end-to-end encrypted communication pose a great threat, not only to the right to privacy and freedom of speech of the population at large, but in particular to children and victims of gender-based violence.[98] A campaign launched in October 2022 by over sixty organisations further warns about how actions that undermine encryption pose a particular threat to victims of gender-based violence and can suppress people exercising their human right to healthcare, in particular reproductive healthcare such as abortions.[99] There are many other examples of how the safety and security of children and adults alike rely on the integrity of their digital communications, which has been further highlighted throughout this paper (see especially Chapter 2.9).

4.1.3 The importance of freedom of expression online

The scope of the proposed CSAR includes hosting service providers that store user-generated content (e.g. social media platforms). This is likely to have a disproportionately detrimental impact on users' right to freedom of expression and opinion as defined in Article 11 of the Charter.

Risk assessment and mitigation measures (Articles 3 and 4) would strongly incentivise social media platforms, among others, to conduct the generalised scanning of public-facing content, such as posts and tweets, despite risks that this is incompatible with the GDPR, DSA and CJEU case law prohibiting general monitoring (see Chapter 3.3). Based on detection orders, these providers could be compelled to undertake such generalised scanning, even extending it to the private message functionalities of their services.

In order to comply with such requirements and orders, which include known and new CSAM as well as "grooming" detection in their broad scope, services that host significant amounts of user-generated content (in practice, any popular social media service or community discussion forum) will ultimately have to rely on automated decision-making to deal with this volume of content.

As we noted in the previous chapter, it will not be technically possible for providers to comply with this without creating a significant and likely unmanageable number of false alarms.

In particular, for text search (which is required for grooming detection) to do any useful work, orders would have to search for very specific text strings (relating for example to specific victims, suspects or offences). Otherwise, they would have to allow investigators to refine their searches progressively, in effect using the scanning system as a search engine that could look through the messages of hundreds of millions of EU residents and their correspondents.

These measures will risk not only catching large amounts of legitimate content in their broad net, which could constitute generalised upload filters that could amount to or at least enable censorship of legitimate content, but further may also have a chilling effect on future free expression.

This is because even just the knowledge that your posts and messages are or may be being scanned could discourage people from fully expressing themselves online. This would be especially pronounced for anyone who seeks to hold power to account – journalists, investigative reporters, human rights defenders, protesters and activists – including youth activists.

The current lack of transparency about scanning practices and technologies – which the CSAR does not tackle – increases these risks even more.

Lastly, the voluntary referrals discussed in Chapter 2.7 further emphasise how the very structure of the CSAR may incentivise over-removal and suppression of legitimate content, by encouraging providers to act "proactively" as much as possible.

98 https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report

99 https://www.fightforthefuture.org/news/2022-10-13-make-dms-safe-orgs

4.2 Withdraw the CSA Regulation

The draft CSA Regulation is an attempt from the European Commission to propose a set of measures that are illegal under EU law, irrespective of the importance of their goal.

In the past three years, DG HOME has wrestled with how to create a law that would allow people's private digital lives to be subject to disproportionately invasive scanning. We suspect that the challenge to find a credible legal basis for such practices is what led to the repeated delays in the law that has now been proposed.

Ultimately, the Commission's attempt to find this basis has not been successful. Given its incompatibility with the Charter, including its violation of the essence of several rights in contradiction with various other well-established principles of EU law, and its incompatibility with the DSA, we do not see how the EU could credibly approve the CSAR.

Whilst the rights of the child demand that the EU takes action to protect children, this does not mean that the EU can take any measure at any cost.

118 civil society organisations – including those representing adult and child survivors of online sexual violence, children's health and privacy organisations, press freedom groups, cybersecurity experts, women's rights groups, and many other social justice and fundamental rights groups – have called on the EU to withdraw the proposed regulation.[100]

Major concerns about the law have also been raised by the EDPS and EDPB,[101] the UN High Commissioner for Human Rights,[102] the German and Austrian governments,[103] the Czech Committee for EU Affairs,[104] and others.

Amendments cannot be sufficient to rectify the CSA Regulation's lack of a legitimate legal basis; the ideology of surveillance; disregard for privacy, security and effectiveness; naivety about technology; and failure to think through the consequences not just on society at large, but also on children themselves. We urge the co-legislators to withdraw the CSAR proposal.

100 https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report

101 https://www.fightforthefuture.org/news/2022-10-13-make-dms-safe-orgs

102 https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report

103 **For example:** https://netzpolitik.org/2022/chatkontrolle-interne-dokumente-zeigen-wie-gespalten-die-eu-staaten-sind/; https://netzpolitik.org/2022/wissenschaftliche-dienste-chatkontrolle-darf-so-nicht-in-kraft-treten/

104 **The Opinion of the Committee states that "The Committee on European Affairs considers it crucial to strike a consistent balance between the extremely important protected interest of preventing child sexual abuse and the protection of the right to privacy and excessive interference with that right, stressing that the resulting regulation must respect the right to protection of encrypted communications."** https://www.psp.cz/sqw/text/orig2.sqw?idd=216063&pdf=1

4.3 Identify elements for further exploration

If the Commission were to go back to the drawing board to work on a new proposal that does justice to the seriousness of the issue, there are a limited number of provisions in the current proposal that could provide a basis for further work.

The first is risk assessments. Providers could be reasonably required to assess the risk their platform or service poses to children and other vulnerable users, such as victims of intimate partner violence, as this is often associated with child abuse. Then, they could be required to determine how to limit the resulting danger as far as is reasonably practical and compatible with fundamental rights law. [105]

The second is user reporting. Given that most new CSAM and grooming is reported by users, while service firms often make this inconvenient to minimise costs, it is entirely appropriate for the law to require effective ways for end-users to report suspected CSAM, and to ensure that providers have sufficient human resources to deal with these reports.

All of these measures are currently only optional under Article 3.2.b. We further note that the DSA includes provisions on user reporting; we therefore suggest carefully studying the barriers to abuse reporting, not just by and on behalf of children, but by and on behalf of other vulnerable users including intimate partner abuse survivors, before pursuing new legislative measures.

There are also elements of removal practices which – with a significant overhaul of the process for how they are actioned by national and centralised authorities to increase efficiency – could in theory be conducted in a rights-respecting way.

Lastly, there are likely benefits to a coordinated approach and a genuinely independent EU Centre whose focus is on enabling national activities. This could support existing efforts and reduce EU dependence on US-based child protection services (e.g. NCMEC), which exist in a different legal framework.

4.4 Pursue alternative approaches

As we have argued, based on the existing law of the European Union, the CSA Regulation would not be lawful. Yet there is broad societal and political agreement that more needs to be done to keep children safe and to stop perpetrators.

Regrettably, some argue that – despite its evident flaws – the CSA Regulation is the best option on the table because of how important it is to do "something". Some argue that there is no other way to tackle this problem.[106] This is not true. CSA is a complex social problem, and the primary response must be from local law enforcement, with a supporting role played by social services, schools, family members, and other local guardians, who are often in the best position to notice and, therefore, report, suspected CSA.

The "Don't look away, Report it!" campaign from EPCAT and InHOPE, for example, teaches that there are warning signs that we can all look out for to help protect children from abuse and trafficking.[107]

Experts from child protection hotlines equally warn that effective solutions must tackle the role of adults who facilitate CSA or who stay silent despite knowing about abuse.[108]

We call on the EU to pursue a combination of short-term measures to tackle the harm and distress to which victims of CSA are subjected, as well as long-term measures to reduce the prevalence of CSA in society, empower survivors, and prevent reoffending.

We hope that this final section will contribute to a nuanced discussion which combines the need to better protect children with the fact that the CSAR's proposed solutions are dangerous, as well as technically infeasible.

105 https://www.who.int/publications/i/item/inspire-seven-strategies-for-ending-violence-against-children

106 **Claim from online campaign about the launch of the CSA Regulation, emphasised by Commissioner Johansson:** https://twitter.com/YlvaJohansson/status/1556252202481713153/photo/2

107 https://www.inhope.org/EN/dont-look-away

108 **For example, see Arda Gerkens, EOKM, speaking at:** https://www.paultang.nl/en/event-csam/

4.4.1 Long-term structural and societal measures

There are several key themes that child rights experts repeatedly propose as the most effective ways to prevent child sexual abuse, as well as ensuring that when it does happen, it is identified and stopped quickly, and that perpetrators are held to account. They include:

- **Requiring service providers to provide effective means for victims and other users to escape harassment or abusive encounters, and to report abuse to the platform moderators in order to block abusers, preserve evidence, and share evidence with local police, teachers, parents or others as appropriate;**
- **Increasing children's awareness of and access to hotlines, institutional reporting (police, social services and other authorities), and support mechanisms, as emphasised in the EPCAT report;[109] as well as investing more in these services and other survivor victim support services, in particular focusing on empowering survivors; [110]**
- **Digital literacy and education, as emphasised by child rights groups, including CRIN:** *"From an early age and throughout their development, children should be taught about their digital rights, the opportunities of the internet, as well as the risks it poses and how to confront them. This way children will be empowered with the knowledge to make informed choices about their activity online without the need for restrictive policies"*;[111]
- **Ensuring long-term health care by professionals for potential perpetrators and to rehabilitate offenders, recommended by many child protection groups;**
- **"Push[ing] for reforms that will open [...] closed institutions... to scrutiny, prevent cover-ups, and allow victims to access justice", tackling structural abuses of power, as well as requiring better and more consistent criminal record checks for people who work with children, as recommended by an investigation into CSA in the French Catholic Church; [112]**
- **Trauma-informed interviewing by police so that young people aren't made to feel like "criminals" and instead have trusted adults to help them navigate CSA reporting;[113]**
- **The World Health Organisation (WHO) recommends tackling the "social tolerance of both victimisation of girls and perpetration by boys and men", which drives low levels of reporting and contributes to victim blaming. The WHO specifically mentions the need to change "gender norms relating to male entitlement over girl's and women's bodies"; [114]**
- **Investing in "primary prevention" and developing "evidence-based strategies" to prevent CSA, which the US Center for Disease Control (CDC) notes are not currently common in the global fight against CSA.[115]**

4.4.2 Tackling short-term issues of CSAM dissemination online

Long-term approaches are likely to be the most sustainable and effective measures, not just for tackling the spread of CSAM online, but for stopping child sexual abuse and exploitation before a child is harmed in the first place. We also recognise that the spread of CSAM online is a form of re-victimisation.

It causes harm and trauma to survivors, can incentivise perpetrators, and is distressing to others that have to view it, such as police officers, child protection case workers, platform moderators and judges.

In addition to the areas for further consideration discussed above, there are several other ideas that the co-legislators should consider in order to have a short-term impact on the spread of CSAM.

4.4.3 Address current failings

The 2011 Child Sexual Abuse Directive contains many provisions requiring EU Member States to do more on a national level, yet has not been fully implemented. This means that there is already a clear, lawful blueprint that Member States could follow to improve the protection of children in their countries.

The CSA Directive is currently being "recast" (reviewed with a view to an updated version of the law), and we believe that it provides the opportunity to resolve many pressing barriers to justice, as well as to implement longer-term solutions.

In this sense, we welcome the actions taken by the European Commission to launch infringing procedures against Member States for failing to implement aspects of the Child Sexual Abuse Directive. [116]

The issue of existing law enforcement procedures and mechanisms is also crucial. Currently, the CSA Regulation can force providers of interpersonal communications services (Article 14) to remove content within 24 hours, and Competent Authorities can ask other types of providers to remove CSAM (Article 32). However, such tasks are the responsibility of police.

As discussed in Chapter 3.9, investigations reveal that law enforcement agencies are systematically failing survivors. Not only would the CSAR make their job even harder, but it will be far more effective to invest in these existing teams and process with expertise - but lacking resources to help children on the front line.

4.4.4 Enable national hotlines to increase their capacity

As discussed in Chapters 3.9 and 3.10, national hotlines play a vital role in protecting children from sexual abuse and exploitation. Globally, hotlines are already overburdened. For example:

"NCMEC does not open or view every image file submitted in a CyberTipline report. [...] Based on the volume of CyberTipline reports NCMEC receives, it is not possible to review all reports much less all image files."[117]

Increasing both EU and national funding to European hotlines, as well as committing funding earlier in the process, would boost the capacity and reduce the precariousness of these vital organisations. This would also be a proven method of removing CSAM from the internet, without the sizeable investment in a new bureaucratic infrastructure that the CSA Regulation proposes.

Since increasing investment in hotlines would not require new capabilities but would instead enable existing capacity to have more impact, such an approach would be likely to reduce the dissemination of CSAM online faster than the approach proposed by the CSAR.

It is important to note, however, that the EU's network of hotlines does not currently have a specific basis in law, despite the sensitivity of the work that they do.

To accompany increased resources, their national legal basis should be clarified urgently. This should be complemented with increased funding to broader victim support organisations, including legal advice services, counselling and mental health services, as well as those that inform survivors of their rights and support them to claim those rights.

4.4.5 Use the DSA implementation to better tackle CSAM

The DSA already contains rules such as notice and action (i.e. removal) of any illegal content of which the platform becomes aware, which includes CSAM. Article 4.1 of the CSA Regulation also explains that the trusted flaggers defined in the DSA can support the detection and removal of CSAM. The priority should be ensuring effective enforcement of such laws – including investing in mental health support for anyone whose job it is to review CSAM, as well as training in trauma-informed interviewing for people in direct contact with survivors.

4.4.6 Invest in lawful, targeted investigation techniques

To make better use of existing mechanisms and structures, we further recommend investing in lawful investigation techniques such as those outlined in EDRi's "10 principles to defend children in the digital age".[118]

Recent experiments in law enforcement innovation, such as the collaboration between child protection group *L'Enfant Bleu*, the French national police, and Europol on the Undercover Avatar project, show that it is possible to protect children online without resorting to measures that rely on surveillance of private communications.[119]

This is particularly pertinent when it comes to considering the effectiveness of measures to protect children online. In the course of around 5 weeks, we note that the Undercover Avatar project engaged with 1200 children who were suffering, or were at risk of suffering, domestic abuse, including sexual abuse. This enabled the French police to intervene to help 360 children who were in a "dire situation" of abuse.[120]

This demonstrates that more targeted, support-focused measures like Undercover Avatar can be extremely effective. However, despite its great success, Undercover Avatar was terminated because of a lack of funding.

4.4.7 Work together for a safer internet for all

The CSAR and other similar legislation are sometimes framed as a zero-sum game between children's rights on the one hand, and digital rights and data protection advocates on the other.

Such a false dichotomy is damaging, unrealistic and in contradiction with a holistic human rights-based approach.

We fully believe that bringing children's rights groups, women's rights groups, digital rights groups, educators, social workers, groups representing minoritised people who especially rely on private online communications (such as sex workers, undocumented people and queer communities), governments (including law enforcement), lawmakers and policymakers together in a constructive environment would help us to collaborate to find additional solutions that will help answer the real question: how can we keep children safe while fully upholding fundamental rights?

109 https://ecpat.org/wp-content/uploads/2022/01/05-01-2022_Project-Report_EN_FINAL.pdf

110 https://www.ciase.fr/rapport-final/

111 https://home.crin.org/issues/digital-rights; **a similar recommendation about education and empowerment is made by EPCAT, Ciase and the other organisations mentioned in this section.**

112 **Ibid.**

113 https://ecpat.org/wp-content/uploads/2022/01/05-01-2022_Project-Report_EN_FINAL.pdf

114 https://www.who.int/publications/i/item/inspire-seven-strategies-for-ending-violence-against-children

115 https://www.cdc.gov/violenceprevention/childsexualabuse/fastfact.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fviolenceprevention%2Fchildabuseandneglect%2Fchildsexualabuse.html

116 https://www.europarl.europa.eu/doceo/document/A-8-2017-0368_EN.html

117 https://epic.org/wp-content/uploads/amicus/algorithmic-transparency/miller/US-Exhibits-NCMEC-Declaration.pdf

118 https://edri.org/our-work/chat-control-10-principles-to-defend-children-in-the-digital-age/

119 https://www.fabriceplazolles.com/enfant-bleu-undercover-avatar

120 https://www.europol.europa.eu/media-press/newsroom/news/europol-excellence-award-in-innovation

Ireland Case Study

Annex

Questions (ICCL) and answers (An Garda Síochána):

1. When did An Garda Síochána start to receive referrals from NCMEC?
"This office started directly receiving NCMEC referrals in 2015. Before this referrals from NCMEC were received via UK, FBI etc., as far back as 2010."

2. How many referrals has An Garda Síochána received per year?
"The number of referrals received is different year on year. In 2017, 2,848 referrals were received; in 2018, we received 6,812. In 2019, 3,888 referrals were received. In 2020, we received 4,192 and so far in 2021, we have received approximately 3,500."

3. How many suspect IP addresses have the Gardaí received per year?
"The number of IP addresses received each year is not recorded at this office. Each referral is unique and a referral received may have numerous different IP addresses contained within it."

4. How many referrals contained the same offending content per year?
"Duplicate content can be a feature of NCMEC referrals. While we know that this does happen, regularly content can be deemed to have gone 'viral', the number of recurring duplicate content referrals are not recorded."

5. How many referrals have led to the launch of a Garda investigation per year?
"Using 2020 as an example, a total of 4,192 referrals were received from NCMEC. 409 of these referrals were actionable, and from those referrals 265 files were completed."

6. How many investigations have led to prosecutions per year?
"This information is not retained at OnCE."

7. How many prosecutions have led to convictions per year?
"As above."

8. How many referrals contained non-illegal content per year?

"OnCE doesn't use a specific categorisation of non-illegal. A total of 471 were marked as being not Child Abuse Material in 2020 from a total of 4,192. This is the focus of the OnCE unit. 506 referrals were marked as being age undetermined. 940 referrals included IP addresses which could not be progressed further. 852 referrals were marked as Child Abuse Material. 606 were marked as below the threshold. 75 were self-generated. 333 were marked as viral. 51 were adult."

9. What percentage of referrals contained non-illegal content per year?

"Please see above."

10. What is the general nature of the non-illegal content which has triggered false referral to An Garda Síochána?

"OnCE will not action a referral further for a number of reasons on the basis of its content, the following are examples: Children playing on a beach, topless content, nudist, adult content, etc."

11. What does An Garda Síochána do with the IP addresses and identifying information pertaining to NCMEC referrals after a) an investigation is complete and b) after the content is found to be non-illegal?

"The following data relating to all NCMEC referrals received is retained at OnCE: NCMEC Cybertip No., Date received, suspect email address, suspect screen name, suspect IP address and reporting ESP.

Actioned NCMEC referrals are retained in full at OnCE. With specific reference to Question 11 above, the processing of personal data for the purposes of law enforcement falls under Part 5 of the Data Protection Act 2018.

While Section 94(3)(a) of the Data Protection Act 2018 states that a data controller can restrict access to data held for the purposes of the prevention, detection or investigation of offences, the apprehension or prosecution of offenders or the effectiveness of lawful methods, systems, plans or procedures employed for the purposes of the matters aforesaid, I am to report that I have spoken to [redacted], Garda Data Protection Unit who has advised that there may be no legal basis to retain data relating to (1) suspect email address, (2) suspect screen name, (3) suspect IP address in the first place in cases which are clearly not child abuse material – for example referrals involving images and videos of children playing on a beach as it may not be proportionate to do so.

Clearly we are covered retaining referrals and related data which involve Child Abuse Material even when the investigation is complete. It is my recommendation that a definitive opinion on the lawfulness from a Data Protection viewpoint of our practice in OnCE in retaining certain data from all NCMEC referrals be obtained from the Garda Data Protection Officer. I will draft correspondence seeking such an opinion under separate cover and forward same via your office."

Mass surveillance. Random Censorship. Content Restrictions.

Companies and governments
increasingly restrict our freedoms.

—

DONATE NOW:

https://edri.org/take-action/donate

Press enquiries
press@edri.org

Brussels office
brussels@edri.org

Phone number
+32 2 274 25 70

Visit us
Rue Belliard 12
1040 Brussels
Belgium

Follow us
Twitter
Facebook
LinkedIn
Youtube

European Digital Rights (EDRi) is the biggest European network defending rights and freedoms online. We promote, protect and uphold human rights and the rule of law in the digital environment, including the right to privacy, data protection, freedom of expression and information.

www.edri.org

Designed by Gemma Portella - www.gemmaportella.com - gemma@gemmaportella.com