TIMELEX

White paper on CSAM detection and prevention mechanisms under current and proposed European data protection legislation

A critical assessment and reflections on policy options for the future

Date	14/03/2024
Author(s)	Hans Graux and Jolien Clemens, Attorneys-at-law at Timelex

While the authors of this White Paper were granted full intellectual freedom and editorial control, their research was funded by Microsoft.

TABLE OF CONTENTS

Executive summary

This White Paper was drafted in the context of current discussions in the European Union on a new legal framework for the detection and prevention of child sexual abuse materials (CSAM), in particular with respect to number-independent interpersonal communications services (ICS). Since 2021, ICS providers have relied on a provisional regime, known as the Interim Derogation, which provides an exception to the general confidentiality rules of the ePrivacy Directive, and allows ICS to process personal data for the sole purpose of detecting and removing CSAM, subject to specific safeguards. Most stakeholders today agree that in the backdrop of a growing prevalence of this crime, there is a necessity for a long-term legal framework that properly addresses this crime, while still satisfying the requirements of the EU's privacy legislation.

It is in this context that the European Commission proposed in May 2022 a Regulation laying down rules to prevent and combat child sexual abuse. At the time of writing, the complexity of finding the right middle-ground between child safety online and privacy has meant that discussions on this proposal are still ongoing. With the end of the Derogation looming, co-legislators have had to extend the lifetime of the Interim Derogation to 3 April 2026.

The CSAM Regulation Proposal, however, also fundamentally changes the approach to detection and prevention, by focusing on mandatory detection orders issued by competent authorities, without creating an explicit mandate for voluntary screening currently seen in the Interim Derogation. A recurring justification for this approach is that detection and prevention of CSAM in the context of ICS are such a powerful intrusion into the fundamental rights to privacy and data protection of the users, that they can only be justified by an order from a competent body, mandated under specific legislation. A conditional mandate for ICS providers to act voluntarily under specific safeguards would not be conceptually permissible as a matter of EU law.

This paper seeks to take a narrow look at this specific facet of the long-term CSAM proposal, namely how to reconcile ICS' providers ability to detect and prevent CSAM with the fundamental right to privacy. The paper analyses the legal framework and prior case law from the Court of Justice of the European Union ("CJEU"), and concludes that it would be possible and desirable to create a permanent mandate for voluntary CSAM detection and prevention within the CSAM Regulation, under strict safeguards that build upon and enhance the approach of the Interim Derogation.

The authors find that the legal framework and the existing case law is significantly more nuanced than is often assumed in discussions. The CJEU consistently emphasizes the importance of context when assessing the lawfulness, legitimacy and proportionality of intrusions into fundamental rights. It has, in prior instances, accepted even large scale automated assessments of personal data, provided that these were coupled with clearly defined and effective measures that appropriately mitigate potential negative impacts on the persons concerned.

In the context of CSAM detection and prevention, such measures can be based on an appropriate and targeted assessment of the ICS, and of the risks and effectiveness of detection tools (since e.g. hash based detection of known CSAM has virtually no false positives). Other measures include human intervention prior to taking further action, transparency towards data subjects, stratified and proportionate response mechanisms that consider the distinctions between e.g. adult offenders

mutually exchanging CSAM and teenagers exchanging images of themselves, as well as prior authorisations by competent authorities, and post hoc verifications in case of incidents or changes in the ICS, or in the detection technologies used. There is no basis to decide generically that no such measures are conceivable as a matter of principle, given the fact that the CJEU has accepted less stringent measures in contexts where the public interest was not as high.

Moreover, any assessment of the viability of voluntary detection by ICS providers cannot end purely with the consideration of the privacy and data protection rights of persons communicating via ICS. Due consideration must also be given to the rights of CSAM victims. The exchange of CSAM is a clear and catastrophic infringement of their rights to privacy, data protection, human dignity, physical and mental integrity, and fundamentally the children's right to protection and care in a democratic society. When considering the feasibility or appropriateness of detection and prevention, their rights are a crucial part of the consideration of the processing context, that should not be swept aside on the basis of the principles of the ePrivacy Directive.

The White Paper does not argue against the introduction of detection orders under the CSAM Regulation. Such orders can indeed play a decisive role in compelling negligent, passive or unaware ICS providers to act appropriately. Nor does this White Paper argue in favour of an unbounded and open mandate for voluntary detection. To the contrary, it notes that the safeguards that currently exist in the Interim Derogation could be strengthened and improved.

Centrally however, it concludes that the continued existence of a mandate for voluntary detection and prevention of CSAM by ICS providers is legally feasible in accordance with EU law by ensuring that these measures are targeted as a result of an ICS-specific risk assessment, and by building in a range of additional safeguards.

Moreover, the existence of such a mandate is a necessary complement to a regime based on detection orders. If the Interim Derogation is allowed to expire, and the CSAM Regulation's proposed text is unmodified, the legal mandate that currently unambiguously permits ICSs to engage in voluntary CSAM screening will disappear. Voluntary screening would be reduced; and in instances where it does continue, this will be done with reduced transparency, and fewer guarantees of harmonised and effective safeguards. The dissemination of CSAM may increase altogether, and at any rate CSAM will not be as effectively combated. This is clearly not desirable from a public policy perspective.

This White Paper argues that voluntary CSAM detection and prevention should be given a clear and unambiguous fiat under the CSAM Regulation, as a measure that is permissible for the affected ICS, provided that a prior risk assessment justifies it, and that certain safeguards are met. It is possible and desirable to establish a co-regulatory model, where the ICSs are allowed to assume responsibility, while building a governance framework that ensures lawfulness, legitimacy and proportionality. In this manner, the Derogation can more effectively contribute to the fight against CSAM, in a manner that aligns with European requirements in relation to the fundamental rights to privacy and data protection.

A. Introduction – CSAM detection and prevention by information society service providers in the EU

In today's society, certain categories of information society services (ISS) providers play a critical role in enabling a broad range of electronic communications between private citizens. Many of these ISS providers conceivably have the ability to support the detection and prevention of child sexual abuse materials (CSAM) exchanged via their networks, and are thus able to support the fight against such materials online. There is however a strong debate in the EU why and to what extent ISSs should play a role in policing communications, or how ISSs must design their services to ensure maximum safety for their users.

This white paper examines only a single complex facet of that broader question: **How (and, if at all) can CSAM detection and prevention by ISSs be reconciled with the fundamental right to data protection in the European Union?** This paper focuses on one particular category of ISS, notably interpersonal communications services (ICS), which is a specific set of electronic communication services covered by the EECC. From a data protection perspective, these services are subject not only to GDPR but also to the ePrivacy Directive, which is also reflected in the European Commission's proposed Regulation on preventing and combatting child sexual abuse (2022) (hereinafter the CSAM Regulation Proposal).

The legal backdrop of this question is complex for several reasons. Firstly, and perhaps most importantly, any action to be undertaken in relation to CSAM – including detection, blocking, or removal - requires a balancing act between multiple fundamental rights. The EU traditionally applies a strong legal framework protecting the fundamental right to privacy and data protection in general, and to telecommunications confidentiality in particular. Like any other fundamental rights, these cannot be infringed upon except where this is lawful, proportionate, necessary, and genuinely meet objectives of general interest recognised by the European Union, or the need to protect the rights and freedoms of others. Systematic monitoring and analysis of private communications within an ICS to detect any presence of CSAM is clearly at tension with that fundamental right.

But similarly, the existence and dissemination of CSAM via an ICS (or generally via an ISS) infringes upon several other fundamental rights, such as the right to human dignity, the right to physical and mental integrity, and every child's right to such protection and care as is necessary for their well-being. ISSs inaction on this would thus be both legally and morally difficult to defend.

Secondly, even if one accepts that ISSs are within their rights – or perhaps that they are obliged – to act against CSAM, there can be questions around the scope of what they exactly are allowed or required to do. In the EU – which is the main focus of this white paper – ISSs have to abide by multiple sets of data protection rules, particularly the GDPR but also the ePrivacy Directive for ICS. As will be further explored below, in 2021 a targeted Regulation was adopted, known as the Interim Derogation,[1]

[1] Regulation (EU) 2021/1232 of the European Parliament and of the Council of 14 July 2021 on a temporary derogation from certain provisions of Directive 2002/58/EC as regards the use of technologies by providers of number-independent interpersonal communications services for the processing of personal and other data for the purpose of combating online child sexual abuse, see https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32021R1232.

which established a provisional framework to allow voluntary CSAM detection and prevention by ICSs in deviation from the ePrivacy Directive, under certain circumstances. The Interim Derogation was intended to be a temporary solution, as it contained a sunset clause under which it shall apply only until 3 August 2024.

To prevent any regulatory and policy gap, a Proposal for a Regulation laying down the rules to prevent and combat child sexual abuse[2] was published in 2022. This new CSAM Regulation would replace the Interim Derogation, and put an entirely new framework in place. The proposal however takes a very different perspective towards the possibilities and obligations of ISSs in relation to CSAM, notably by imposing mandatory risk assessments for hosting providers and ICS providers, as well as introducing the possibility for authorities to issue detection orders. Detection orders could be issued for known and unknown CSAM for hosting providers, but could extend to grooming for ICS. Part of the drive behind this change in approach is the consideration that a mechanism which focuses on detection orders, rather than on voluntary actions, can be more easily reconciled with European data protection law.

In this white paper, we will examine current data protection frameworks and their limitations for effectively combatting CSAM detection and prevention by ISSs, both under current legislation and under the CSAM Regulation Proposal. While we will touch upon the position of ISSs under EU data protection law in general, we will particularly focus on the situation for ICS created by the Interim Derogation. We will also evaluate the approach of the CSAM Regulation from a data protection perspective, and assess its strengths and weaknesses. Finally, we will examine potential solutions that could strengthen the EU legal and policy framework in the future.

B. ISS providers and telecommunications privacy under EU data protection law

1. Introduction of the data protection legal framework in the European Union

The data protection legal framework applicable to information society service providers is to a large extent harmonized at EU level. ISSs are subject to a number of EU legislations which intend to protect the right to privacy of the user and their right to the protection of their personal data.

As aforementioned , these two fundamental rights are enshrined in the Charter of Fundamental Rights of the European Union (hereinafter "the Charter")[3], in which the right to privacy is based on the

[2] Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse. COM/2022/209 final; see https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A209%3AFIN

[3] Charter of Fundamental Rights of the European Union, 2000/C 364/01, see https://www.europarl.europa.eu/charter/pdf/text_en.pdf.

fundamental right to private life as protected by Article 7 of the Charter, and the right of protection of personal data is included in Article 8.

The **General Data Protection Regulation**[4] (hereinafter "**GDPR**") is arguably the main data protection framework applicable to ISSs who are established in the EU or who direct their services towards EU citizens.[5] The GDPR applies when processing personal data of natural persons in general. Therefore, it will be applicable to ISSs that are processing content data (communications, images, videos, etc.) and metadata (i.e. device and connection information, location data, etc.) of users on their services.

While ISSs mainly process personal data to provide the core functionalities of these services (i.e. communications services, hosting services, etc.), the data is often also processed for ancillary reasons in order to ensure that the services are safe and secure, i.e. to combat fraud, spam, and illegal content in general. When processing this kind of data, ISS will have to respect the key principles of the GDPR[6].

A certain segment of these ISSs, namely those that provide electronic communication services, will also be subject to the **ePrivacy Directive**.[7] As stated in Article 1 (2), the ePrivacy Directive *particularises* and *complements* the European legal framework on the processing and free movement of personal data in the Union[8]. In other words: the general principles of the GDPR remain fully applicable to those ISSs that are (also) providing electronic communication services. The ePrivacy Directive provides a set of specific rules on data protection in the area of electronic communications, such as on the confidentiality of electronic communications, the treatment of traffic data (including data retention), spam and cookies, etc.

While the ePrivacy Directive is arguably no longer optimally suited to the fast-changing nature of the electronic communications sector, attempts to review and replace it have thus far failed.[9]

In the next Chapter of this White Paper, we will examine how the ePrivacy Directive impacts a particular set of electronic communication services, notably interpersonal communication services or ICSs, that are currently already voluntarily scanning and detecting CSAM on their services, including under the Interim Derogation. Before doing so however, it is important to examine how earlier interferences in the fundamental rights to privacy and data protection have been addressed under EU law. This is useful, since it provides insights into the limitations that ICSs have to respect when engaging in voluntary detection and prevention and the limitations that the EU fundamental rights framework

[4] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regards to the processing of personal data and the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), see https://eur-lex.europa.eu/eli/reg/2016/679/oj.
[5] Article 3 (1) and (2) of the GDPR.
[6] Article 5 (1) of the GDPR.
[7] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communication sector (Directive on privacy and electronic communications).
[8] At the time of adoption of the ePrivacy Directive, this was a reference to the Data Protection Directive 95/46 EC. Presently, this should be understood as a reference to the GDPR.
[9] A proposal for an ePrivacy Regulation was published on 10 January 2017.[9] The discussions on this proposal have however been stalled at the Council for almost 6 years, and it is uncertain whether the proposal will be adopted in the foreseeable future. The ePrivacy Directive therefore is and remains presently the law of the land, as a complement to the GDPR.

imposes on the EU legislator – i.e. the question whether the Interim Derogation and the CSAM Regulation Proposal strike the right balance on the basis of prior case law.

2. The case law of the Court of Justice on an interference of fundamental rights

2.1. The Court of Justice of the European Union's three-factor test

Any legislation enacted by the European Union which has an impact on fundamental rights, must meet the criteria that can be found in Article 52 (1) of the Charter, which are further developed in case law of the Court of Justice of the European Union (hereinafter the "CJEU"). Article 52 (1) of the Charter states that any limitation on the exercise of the rights and freedoms recognized by the Charter must be provided for by law and respect the essence of those rights and freedoms. Taking into account the principle of proportionality, this means that limitations are only possible if they are necessary, and if they meet objectives of general interest recognized by the European Union or the need to protect the rights and freedoms of others.

This is an important constraint that must be respected by the European legislator as well, as the CJEU has the right to annul legislation enacted by the European Parliament and the Council if it is not in line with the fundamental rights in the European Charter; and the CJEU has indeed done so in the past. Any legislation that allows ICSs to scan their services on the occurrence of CSAM and to report it to law enforcement authorities clearly has a significant impact on the right to private life (including private communications) and the right to protection of personal data. For that reason, this section will provide a short overview of selected relevant case law of the CJEU, as the criteria developed therein will also be applied to the proposed CSAM Regulation.

The CJEU (as well as the European Court of Human Rights of the Council of Europe) applies the following legal criteria to establish whether a limitation on the exercise of a fundamental right can be justified:

1. **Lawfulness**: The interference must be prescribed by law
2. **Legitimacy**: The interference must pursue a legitimate aim
3. **Proportionality**: The interference must be proportionate

Concerning the criteria of **lawfulness**, the CJEU has stated that any EU legislation under consideration must lay down clear and precise rules which determine the scope and the application of the intended measure and must impose minimum safeguards.[10]

The principle of **legitimacy** is interpreted as pursuing an objective of general interest recognized by the Union or the need to protect the rights and freedoms of others.

Lastly, the principle of **proportionality** requires that the limitations imposed by the legislation under consideration must be appropriate and necessary to meet the legitimate interests pursued or the need

[10] This principle is derived from case law of the European Court of Human Rights in ECHtR, H.R., Liberty and Others v. the United Kingdom, 1 July 2008, no. 58243/00, § 62 and 63; Rotaru v. Romania, § 57 to 59, and S. and Marper v. the United Kingdom, § 99

to protect the rights and freedoms of others; and that the objective of general interest could not reasonably be achieved through less intrusive means.

In 2014, in the context of electronic communications in particular, the CJEU has applied the above-mentioned criteria to the Directive 2006/24/EC[11] ('the Data Retention Directive'), and found that this Directive constituted an unjustified interference with Articles 7 and 8 of the Charter, in its Digital Rights Ireland Case. This Directive imposed an obligation on providers of publicly available electronic communications services or of public communications networks to retain certain data (mainly traffic and location data, excluding content data) which are generated or processed by them, for the purpose of investigation, detection and prosecution of serious crime.[12]

The Court considered the interference to be "wide-ranging and particularly serious", because it was applicable to all means of electronic communication and covered all subscribers and registered users.[13] In reaching this conclusion, the CJEU emphasized the fact that the rules of the Directive affected everyone, including individuals for whom there was no evidence linking their conduct, even indirectly, to serious crime.[14] Lastly, the CJEU also considered that the Data Retention Directive lacked substantive and procedural rules relating to the access by competent authorities to the data and to their subsequent use. As a result, the CJEU annulled the Data Retention Directive. Comparable decisions were later issued in two similar joined cases, examining whether two national implementation legislations of the Data Retention Directive in Sweden and the United Kingdom which required telecommunication service providers to indiscriminately, systematically and continuously retain certain data were compatible with the Charter of Fundamental Rights.[15]

In 2022, the CJEU also applied the interference criteria in a judgement on the Passenger Name Records (PNR) Directive (specifically the Belgian transposition law of the Directive).[16] This judgment is particularly interesting as the CJEU provides comprehensive guidelines on how large-scale predictive policing (surveillance) can take place. The PNR Directive sets out obligations for air carriers to collect and transfer information on passengers to competent authorities, so-called Passenger Information Units (PIUs), for the purpose of improving border control and combatting illegal immigration. These PIUs will - upon receiving the data - automatically process the PNR data by comparison, against both pre-existing databases and against "pre-determined criteria".[17] The CJEU found that the PNR Directive entailed a serious interference with the rights guaranteed in Article 7 and 8 of the Charter as it introduces a surveillance regime that is continuous, untargeted and systematic, including an automated assessment of the personal data of everyone that is using air transport services, without consideration of prior indications of unlawful activity. However, the CJEU concluded in this instance that a fair balance was struck – despite the indiscriminate nature of the process - because the Directive

[11] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communication services or of public communications networks and amending Directive 2002/58/EC (repealed), see https://eur-lex.europa.eu/eli/dir/2006/24/oj.
[12] Article 1 of the Repealed Data Retention Directive.
[13] Joined Cases C-293/12 and C-594/12, *Digital Rights Ireland Ltd*, § 37 and § 56.
[14] Ibid., §58.
[15] Joined Cases C-203/15 *Tele2 Sverige AB v Post-och Telestyrelsen* and C-698/15 *Secretary of State for the Home Department v Tom Watson and Others*.
[16] Case C-817/19 *Ligue des droits humains ASBL v Conseil des ministres*.
[17] These criteria are intended to identify persons involved in criminal or terrorist activities who are, as of yet, not known to law enforcement authorities.

(and the national transposition law) ensures the internal security of the EU (safety of EU citizens)[18], which are general objectives of general interest.

Regarding the proportionality assessment, the CJEU noted firstly that the interferences are strictly necessary, on the basis that the Directive requires the systematic transfer and automatic advance assessment[19] of the PNR data of all passengers on extra-EU flights (i.e. the scope of the Directive was limited to passengers flying between the EU and third countries; this was seen as a suitable constraint). The CJEU found that the objective of the PNR Directive would not be attained if the transfer and the assessment of the PNR data of air passengers would be restricted to a particular group of air passengers (i.e. those passengers for whom there are indications of terrorist involvement).

Interestingly, the CJEU explicitly acknowledged the fairly substantial number of false positives with the automatic assessment system used, but stressed that "*the appropriateness of the system […] essentially depends on the proper functioning of subsequent verification of the results […] by non-automated means*" (i.e. the human-in-the-loop).[20]

Finally, the CJEU highlighted the right to an effective judicial remedy, as enshrined in Article 47 of the Charter of Fundamental Rights of the European Union, by stating that the passengers concerned need to be informed how the pre-determined criteria and the systems used work, so that they can decide with full knowledge of the relevant facts whether or not to exercise their right to judicial redress.[21] Thus, in the PNR context, the Court considered the PNR Directive compatible with the human rights framework, despite the large and indiscriminate nature of the screening and the risk of false positives, taking into account the mitigating measures including the human verification and the availability of judicial remedies. Of course, another important element was the importance of the public interest being defended by the PNR Directive. This element will be briefly discussed below.

2.2. Public security vs. national security objectives

In the aforementioned cases, the CJEU developed a novel approach by tying the level of intrusiveness that is allowed to the objectives pursued by the intrusive measure. In *La Quadrature du Net and Others*[22], the CJEU distinguished three different types of public interest objectives, which may justify different types of interferences with fundamental rights. The hierarchy of public interest objectives can be construed as follows:

- Safeguarding **national security** when there is a 'serious threat', which is 'genuine and present or foreseeable';
- Combating **serious crime**, preventing 'serious threats' or 'serious attacks' on **public security**;
- Combatting **crime** and safeguarding public security

[18] The PNR Directive states that the PNR data that is collected in accordance with the Directive may only be further processed for the purposes of preventing, detecting, investigating and prosecuting terrorist offences and serious crime.

[19] The advance assessment is intended to identify the persons who were not suspected of involvement in terrorist offences or serious crime prior to that assessment and who should be subject to further examination by the competent authorities.

[20] Case C-817/19 *Ligue des droits humains ASBL v Conseil des ministres,* para. 124.

[21] Ibid., para. 210.

[22] Joined Cases C-511/18, C-512/18 and C-520/18, *La Quadrature du Net and Others v. Premier Ministre and Others*, ECLI:EU:C:2020:791.

The first objective "*safeguarding national security*" is defined by the CJEU as '*encompassing the prevention and punishment of activities capable of seriously destabilising the fundamental constitutional, political, economic and social structures of a country, in particular, of directly threatening society, the population or the State itself, such as terrorist activities*'.[23] It must be clear that this is a particularly high threshold to attain, and that the crime of child sexual abuse, while being a 'serious threat' to individuals, would be unlikely to be considered as a threat to national security.

The CJEU furthermore stated that for this first objective ("safeguarding national security") a serious interference may be justified, such as general and indiscriminate preventative retention of traffic and location data. However, mass retention must be for a limited period, i.e. for as long as there is a serious threat to national security, which must be genuine and foreseeable.[24] Furthermore, the CJEU highlighted that data retention cannot be systematic in nature and must be subject to certain limitations and safeguards (such as possibility of judicial review).[25]

The second objective "*combatting serious crime, preventing serious threats or serious attacks on public security*", the CJEU considered in its data retention case law that a targeted surveillance is allowed in this instance, which essentially meant that the individuals that might pose a threat to public security must be identified beforehand. However, the CJEU stated that this can be done using a geographical criteria by, for example, identifying places or infrastructures with a high incidence of serious crime.[26]

The third objective includes the prevention, investigation, detection and prosecution of criminal offences (irrespective of their seriousness). In at least one case, the CJEU has ruled that general and indiscriminate retention of subscriber data (i.e. data about civil identity) is permitted.[27]

3. Conclusion

While none of the above cases above are directly related to CSAM detection and monitoring in ICSs, they are relevant in CSAM discussions, as they point to the continuous and justified reliance of the Court on the three-factor test (lawfulness, legitimacy and proportionality) as well as the Court's consistent consideration for the processing *context* (notably the public policy objectives of any legislation that compromises fundamental rights), and of *appropriate safeguards*. Legislation that does not appropriately take this context into account has been struck down repeatedly, whereas legislation that contained appropriate checks and balances, that are commensurate to the policy objectives and the potential harm, has been upheld, even where the scope of the legislation is intrusive and affects a large group of persons.

In the next Chapter, we will assess whether the current Interim Derogation fulfils these requirements, and where potential challenges or problems can be identified.

[23] La *Quadrature du Net and Others*, para. 135.
[24] Ibid., para. 137.
[25] Ibid., para. 138.
[26] Ibid., para. 150.
[27] Case C-207/16, *Ministerio Fiscal*, ECLI:EU:C:2018:788.

C. The status quo for CSAM detection and filtering: the ePrivacy Directive and the Interim Derogation

1. The impact of the ePrivacy Directive on the voluntary detection efforts by ICSs

The rise of the internet and new types of communications has increased and diversified the risk factors for children online. Many ISSs and ICSs have reacted diligently and conscientiously, by commencing voluntary detection, reporting and removal of CSAM from their services.

From a legal perspective, such efforts became harder to defend for ICS with the entry into force of the European Electronic Communications Code in December 2020.[28] This Directive revised the definition of "electronic communication services" to explicitly include the category of ICS. The notion of ICS also includes network-independent services, particularly number-independent services (often also referred to as OTT services). The expansion of the definition of electronic communication services meant that network-independent services also became subject to the provisions of the ePrivacy Directive.

In the context of this White Paper, this was a significant shift. Prior to the EECC, the ePrivacy Directive only applied to 'publicly available electronic communication services' and services which were functionally equivalent. Until then, hosting service providers and providers of information society services that do not provide traditional electronic communication services did not fall under the scope of the ePrivacy Directive. They would fall under the GDPR of course, but not under the material scope of the ePrivacy Directive .

However, due to the enlargement of the definition of the electronic communications service providers, network-independent ICS providers that were previously not subject to the ePrivacy Directive now also had to comply with the relatively strict confidentiality obligations and the requirements regarding processing of traffic and communications data of that Directive. These requirements include the obligation to keep communications between two or more participants private and confidential, and this obligation could even cover automated monitoring and evaluation (meaning without human consultation) of the contents of communications could be covered. In view of these obligations, it became questionable whether voluntary detection by ICS on these services was still defendable.

To resolve this issue, the European legislator intervened through the adoption of the Interim Derogation[29], in which the legislator chose to approve of voluntary screening, subject to conditions. However, the Interim Derogation was only intended to provide temporary relief, as we will describe below.

[28] Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code (Recast), PE/52/2018/REV/1, see https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32018L1972.

[29] Regulation (EU) 2021/1232 of the European Parliament and of the Council of 14 July 2021 on a temporary derogation from certain provisions of Directive 2002/58/EC as regards the use of technologies by providers of number-independent interpersonal communications services for the processing of personal and other data for the purpose of combating online child sexual abuse, see https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32021R1232.

2. A temporary solution: the Interim Derogation

2.1. Introduction to the Interim Derogation

The Interim Derogation, which entered into force in August 2021, provides for a temporary derogation for certain service providers (including ICSs) from the confidentiality obligations in the ePrivacy Directive for the specific purpose of continuing their voluntary CSAM detection efforts. It must be highlighted that the Interim Derogation solely provides for a possibility – or perhaps more accurately a *permission* – for ICSs to use voluntary detection tools which fulfil the requirements set out in the Derogation to specifically detect and remove CSAM from their services. It does not and was never intended to entail an obligation for ICSs to implement a scanning technology on their services. This, among other elements, sets it apart from the examples of case law described above, all of which related to *obligations* to process certain personal data.

The Interim Derogation was designed to be technologically-neutral, meaning that it does not oblige ICSs to implement a specific detection technology if they want to rely on the derogation. ICSs can choose which technologies (if any) they deem appropriate; but the Derogation does put forward certain safeguards that need to be taken into account by these ICSs in their selection and use of any given CSAM detection tool[30]. Unsurprisingly, these align well with the general logic and principles of the GDPR:

- **Necessity and proportionality**: the processing of personal data by the CSAM detection tool must be proportionate, and limited to technologies that are used by providers for the sole purpose of detecting and removing CSAM on their services. The Derogation is however not limited to CSAM, but also covers solicitation of children (grooming);
- **Purpose limitation:** the processing should be limited to content data and to related traffic data, and the tool should only be used to detect known and/or unknown CSAM on their services (and not other types of unlawful content);
- **Technological safeguards:** the technology used must be in accordance with the state of the art in the industry and must be the least privacy-intrusive;
- **Data Protection Impact Assessment (DPIA):** service providers that are using a CSAM detection tool must perform a DPIA on the technology used, in order to identify and mitigate risks for the data subjects;
- **General duty of minimization of error rates of the technology used,** which includes keeping the number of false positives to a minimum.

Furthermore, the Interim Derogation imposes certain obligations on electronic communications service providers that need to be taken into account when implementing a detection technology[31]:

- **Internal procedures** need to be established to prevent abuse of, unauthorized access to, and unauthorized transfers of, personal and other data;
- **Human oversight** of the technology used and **human intervention** in the processing of personal data needs to be ensured;

[30] See Article 3 (1) a- c of the Interim Derogation.
[31] The specific obligations can be found in Article 3 (1) g) of the Interim Derogation.

- If unknown CSAM gets flagged by the technology used then there should be a **prior human consultation** before reporting the CSAM to law enforcement authorities;
- **Appropriate procedures and redress mechanisms** need to be established so that users can lodge complaints within a reasonable timeframe for the purpose of presenting their views;
- **Users need to be informed in clear, prominent and comprehensible way** of certain elements that are specified in the Interim Derogation (including the fact that the Derogation is invoked by the service providers, and any instances where their content is removed, their account is blocked or their service is suspended).

The Interim Derogation was designed with a limited applicability in mind: its closing Article currently[32] still notes that it shall expire on the 3rd of August of 2024.[33] This meant that the European Union was expected to seek a more permanent and robust solution to CSAM online. This more permanent solution is expected to be provided by the recent Proposal for a Regulation laying down the rules to prevent and combat child sexual abuse.

2.2. Assessment of the intervention by the European legislator: exceeding the principle of subsidiarity?

As stated, the ePrivacy Directive imposes a strict prohibition on voluntary detection unless a legal authorization in Member State law is provided (unlike the GDPR, which was previously applicable to many of the currently impacted ICS). This is reflected in the wording of Article 15.1 of the ePrivacy Directive, which states that Member States may adopt legislative measures to restrict the scope of certain rights and obligations in the ePrivacy Directive.

In light of the **principle of subsidiarity** which governs the EU's legislative competences, one might wonder whether the European legislator could have intervened in the first place, as Article 15.1 of the ePrivacy Directive clearly confers the legislative power to act to the Member States. The principle of subsidiarity seeks to safeguard the ability of the Member States to take legislative decisions and actions, and authorizes intervention by the Union in the areas where it has non-exclusive powers when the objectives of an action cannot be sufficiently achieved by the Member States (i.e. there is significant added value of EU intervention).[34] By conferring only legislative power to act on the Member States, the Directive suggests that, at the time of its adoption, the Member States were best placed to act on this matter.

None the less, the Derogation was adopted at the EU level. The European Commission thus took the position that, while Article 15.1 of the ePrivacy Directive conferred sole legislative power to the EU Member States, this wouldn't allow for effective action, and therefore that the principle of subsidiarity was not an encumbrance to the Derogation. The position is perhaps logical, but implies *de facto* that the Commission treated the conferment as unwritten, since the assessment would have been exactly the same if the ePrivacy Directive did not allocate legislative powers to the Member States.

[32] As commented above, on 15 February 2024 a political compromise was announced on the extension of the Interim Derogation until 3 April 2026; this comprise is still to be formally adopted at the time of drafting of this White Paper.
[33] See Article 10 of the Interim Derogation.
[34] See Protocol (No 2) on the application of the principles of subsidiarity and proportionality, 2008, OJ L 115/206.

A potential justification can be found in case law of the Court of Justice, namely the *Ex Parte British American Tobacco Case*.[35] An important consideration in this case was the fact that the CJEU stated that the EU legislature should have the possibility to amend or adopt certain EU legislation in order to properly carry out its task of safeguarding the general interests, particularly taking into account while doing so *any change in perceptions or circumstances*.[36] In this case, the CJEU took into account the progress in scientific knowledge regarding the dangers of smoking and the increased importance given to the social and political aspects of the anti-smoking campaign.

In much the same way one could argue that at the time of the adoption of the ePrivacy Directive, the EU legislature was not yet aware of how fast the technological changes would occur and the increasingly growing trend of CSAM dissemination on electronic communications services. The argument can therefore be made that perhaps back in 2002, the EU may have found that the Member States were best suited to adopt derogations to the ePrivacy Directive's confidentiality obligations, but that this has now proven to be insufficient due to the technological, societal and legal changes.[37]

2.3. The overall legitimacy and proportionality of the Interim Derogation

Since its inception, the Interim Derogation has been criticized by both the EDPB and the EDPS and by privacy organizations in general on the basis that it does not properly balance the right to protect children online on the one hand against the right to private life and communications (Article 7 of the Charter) and the right to protection of personal data (Article 8 of the Charter) on the other hand.

The EDPS provided a negative Opinion on the 10th of November of 2020 on the proposal for an Interim Derogation and stated that concerns with regards to legitimacy and the proportionality of the Proposal.[38] The main concern voiced by the EDPS is the fact that the Interim Derogation allows for a general, indiscriminate and automated analysis of all text-based communications with a view of identifying new potential infringements.[39] Moreover, the EDPS stated that the Interim Derogation does not provide for specific and effective safeguards against general and indiscriminate monitoring.

As noted in the analysis of the CJEU's case law however, general, indiscriminate and automated analysis of text-based communications are not sufficient to conclude that the fundamental rights of the data subject are disregarded. The Data Retention Directive was struck down on the basis of the fact that it was disproportionate, vague and did not contain appropriate safeguards. Retention was

[35] Case C-491/01, *The Queen v Secretary of State for Health, ex parte British American Tobacco (Investments) Ltd and Imperial Tobacco Ltd*.

[36] This case concerned a Council Directive on the labelling of tobacco products and on the prohibition of the marketing of certain types of tobacco and a Directive on the maximum tar yield of cigarettes. The Directives in question were highly criticized because there existed already harmonized rules on EU level for the purpose of eliminating barriers to trade in tobacco products (these were however only minimum harmonized rules, allowing the Member States to introduce stricter rules), which led critics to the conclusion that other harmonizing action by the EU was not necessary.

[37] This change of mindset of the EU legislature is even directly reflected in the text of the draft ePrivacy Regulation Proposal, which states in Article 11 that "*Union or Member State law may restrict by way of a legislative measure the scope of the obligations and rights provided for in Articles 5 to 8*".

[38] Opinion 7/2020 on the Proposal for temporary derogations from Directive 2002/58/EC for the purpose of combatting child sexual abuse online, see https://www.edps.europa.eu/sites/default/files/publication/20-11-10_opinion_combatting_child_abuse_en.pdf.

[39] Ibid., para. 26.

mandatory, and thus did not allow any consideration of context. These same cannot be said of the Interim Derogation:

- The Interim Derogation applies to a narrow use case – i.e. the targeted scanning of communications for the purpose of the prevention and the detection of CSAM – whereas the Data Retention Directive had a far broader use case including the prevention of all types of serious crime which needed to be further defined in national implementation laws. Vagueness was thus much less problematic in the case of the Interim Derogation.
- There is no obligation for ICSs – as was the case under the Data Retention Directive – to retain the data for the purpose of prevention and detection of serious crime. The Interim Derogation clearly mentions the fact that the processing of the data for the sole purpose of detecting and removing CSAM and reporting it to LEA; any extended data retention is not allowed.
- The Interim Derogation included specific safeguards that are intended to limit the interference with fundamental rights, specifically with the right to private telecommunications, i.e. a mandatory data protection impact assessment ("DPIA") must be performed, there must be a compulsory human review before submitting reports of unknown CSAM to other organisations, user redress needs to be ensured, etc. This is an important consideration to make, as the CJEU highlighted in the case law regarding the Data Retention Directive that the fact that these procedural safeguards were not sufficiently – or not even at all – included in the EU legislation, and were left to the scrutiny of the Member States. In contrast, in the PNR decision such safeguards were present, and deemed appropriate.
- Most importantly, –the data retention legislation included an obligation for all affected service providers to process personal data, without allowing them any consideration of the nature, context or scope of their processing activities, and whether this justified retention of any kind. This is fundamentally different from the Derogation, which (1) allows (but does not require) CSAM screening; (2) requires ICSs themselves to make the assessment whether this is necessary and justified in their situation; and (3) holds them accountable to a significant extent, not only through the DPIA obligation, but also through obligations to publish transparency reports on the processing of personal data to the competent supervisory authority and to the Commission. The approach of the Derogation supports accountability and proportionality, in a manner that is clearly better in line with prior CJEU jurisprudence.

Collectively, this renders it likely that an instrument such as the Derogation would survive the three-factor test (lawfulness, legitimacy and proportionality) in case of a challenge before the Court of Justice.

2.4. Remarks in relation to the Interim Derogation

Despite that general assessment, it is important to analyse some of the weaknesses of the Derogation. The main purpose of this assessment is not to criticise a regime that appears to be achieving many of its goals, but rather to seek out ways to strengthen and improve the approach from a data protection perspective, including in the finalisation of the proposed CSAM Regulation, while limiting the interference with the fundamental rights included in Article 7 and Article 8 of the EU Charter.

A first remark concerns **the transparency obligations**. The Derogation requires ICSs to inform users of the fact that they have invoked the derogation to the ePrivacy Directive and on "*the logic*" of the measures they have taken and "*the impact*" on the confidentiality of the user's communications. While there is thus some degree of transparency, these obligations are rather generic and abstract. Specifically, the obligation falls short of mandating ICSs to disclose specific details on the methodologies of detection technology employed, the type of personal data subjected to processing, the duration for which such data is retained, and the identity of parties with access to the data in instances of CSAM detection. While the ICSs also have to publish annual reports, it is not clear to what extent these are useful to the data subjects.

Simply by way of comparison, the recent Regulation (EU) 2021/784 of 29 April 2021 on addressing the dissemination of terrorist content online (hereafter the "Terrorist Content Online Regulation")[40] contains a more detailed provision on transparency, as a part of a much broader section in the Regulation on safeguards and accountability. This Regulation recognises that combating unlawful content should be based on a combination of mandatory orders (removal orders in that Regulation), and on voluntarily chosen measures by hosting providers. In contrast to the Interim Derogation, the Terrorist Content Online Regulation requires a clear disclosure of the policy of the ISS (rather than disclosing the invocation of a Derogation), an explanation of the functioning of the measures, and of the use of any automated tools.

The second remark is more of a **methodological** nature. Whereas, indeed the rate of accuracy of certain detection technologies used, such as hash matching technologies for known CSAM, is high, the same cannot be said for certain other technologies, such as the automated recognition of new CSAM and the use of AI text classifier technologies that aim to predict the occurrence of grooming in online communications. The Interim Derogation is not particularly precise on what type of technologies are allowed to be used by the ICSs, and on what extra risk mitigation measures should be implemented if certain technologies are used that might imply a higher risk to data subjects. This approach was intended to enable flexibility, since the Interim Derogation on this point builds on the accountability principle of the GDPR: the ICSs are likely best placed to determined what is feasible and appropriate for their own services, but should also be accountable in relation to their assessment. In practice, however, the Interim Derogation offers little guidance on how that assessment should take place.

Moreover, it should also be recognised that the largest players are likely to have sufficient resources to invest in state-of-the-art detection tools, trained human reviewers/analysts, extra audits on the hash lists used, etc. Smaller ICSs and startups do not have similar resources, which might lead to them using less reliable detection technologies, or without appropriate reflection on best practices in risk mitigation.

The last remark is more **procedural** in nature. The Interim Derogation introduces some procedural safeguards with regards to the detection and reporting of CSAM, by clarifying what ICSs are allowed to do (i.e. what they *might* do), and how they should inform their users of their actions. These general safeguards, however, do not outweigh the overall lack of clarity on the actual procedure that will be used by ICSs when they are reporting and removing CSAM. For a user of an ICS service, it may not be possible to determine the actions that will be taken, and what the impacts might be. Given the risk of

[40] Regulation (EU) 2021/784 of 29 April 2021 on addressing the dissemination of terrorist content online, see https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021R0784.

false positives – especially in relation to new CSAM or grooming, where detection possibilities are inherently less reliable – this is a problem that needs to be better addressed in the CSAM Regulation.

Most ICSs that are using voluntary detection tools are currently reporting to the National Centre for Missing and Exploited Children (NCMEC), either because they are an ICS established in the U.S., and are therefore required under U.S. law to report CSAM to NCMEC,[41] or they are an ICS established in the EU that decides to report to NCMEC on their own volition, since an equivalent EU alternative is not currently in place.[42] In this regard, the Interim Derogation merely stated that providers can process and store the data for the purpose of "reporting it to the competent law enforcement and judicial authorities or organisations acting in the general interest against child sexual abuse". There is no detail on how they should exactly report, i.e., details about timing, information to be included in the report, to whom they exactly need to report, etc. This relative vagueness implies that highly sensitive personal data may be processed in ways that are not easily foreseeable or appreciable for the data subject.

D. The proposed approach and anticipated impact of the CSAM Regulation

1. An overview of the new proposed approach of the CSAM Regulation Proposal

The CSAM Regulation Proposal is intended, as stated, to provide a permanent legislative framework to replace the Interim Derogation. The proposed rules will oblige certain ISSs (namely ICSs and hosting service providers)[43] to firstly assess and minimize the risk of misuse of their services (risk management and risk mitigation). Next, and depending on the outcomes of this assessment, it can include targeted obligations for certain providers[44] to detect such abuse, and to report it to an EU Centre that will be established under the Regulation; and subsequently, to remove or disable access to, or to block online child sexual abuse when ordered to do so by a competent judicial or administrative authority. This is done via so-called detection orders, removal orders and blocking orders.

To support this approach, the CSAM Regulation Proposal also introduces a new institutional framework, i.e. Member States will have to appoint Coordinating Authorities, which will be the primary national authority overseeing the consistent application of the Regulation. These coordinating

[41] The reporting obligation in US law is under the PROTECT Our Children Act (18 USC 2258A) and applies to ESPs with "actual knowledge" of facts and circumstances of child exploitation on their services. It is therefore similar to the reporting obligation in the EU proposal as this will apply to ESPs "becoming aware" of any information indicating potential online child sexual abuse on its services.

[42] An overview of the ISSs that reported to NCMEC in 2021 can be found here: https://www.missingkids.org/content/dam/missingkids/pdfs/2021-reports-by-esp.pdf.

[43] The obligations of the CSAM Regulation Proposal apply to all providers of hosting or interpersonal communication services offering services in the EU's digital single market. However, as the current White Paper is mainly addressing ICSs and to a lesser extent hosting service providers, we will use the term ICSs in the rest of this section on the proposed CSAM Regulation.

[44] See Article 1 (1) a) c) and d) of the CSAM Regulation Proposal: the detection, removal and blocking access orders apply to providers of hosting services and providers of interpersonal communication services.

authorities will be attributed specific investigatory and enforcement powers in relation to in-scope ISSs that fall under their jurisdiction.

The EU Centre will have a centralized and coordinating role towards online service providers in the risk assessment, detection, reporting and removal of CSAM. From an operational standpoint, the EU Centre will maintain an information-sharing system and a database of indicators and CSAM reports, and support procedural effectiveness by e.g. maintaining authoritative hash lists of known CSAM, and by assisting victims with removal requests. The EU Centre is also required to work with national law enforcement bodies and Europol, by reviewing the reports from the providers to ensure that they are not submitted in error, and will act as a knowledge hub for best practices towards the Member States themselves.

2. Possibility to voluntarily use detection technologies under the CSAM Regulation Proposal

The use of voluntary detection tools by ICSs is currently not directly addressed in the text of the CSAM Regulation Proposal. The only explicit recognition of the voluntary efforts that are currently undertaken by the ICSs can be found in one of the Recitals of the CSAM Regulation Proposal, which states that: "*providers are free to design and implement, in accordance with Union law, measures based on their existing practices to detect online child sexual abuse in their services and indicate as part of the risk reporting their willingness and preparedness to eventually being issued a detection order under this Regulation, if deemed necessary by the competent national authority*".[45] However, the recital clearly calls out the need to comply with Union law (including the ePrivacy Directive), and the right to continue existing practices is presented in the context of designing a response to detection orders.

The Proposal does require ICSs, along with other in-scope services, to conduct a risk assessment in order to evaluate the risk that their respective services are or will be used for the purpose of online child sexual abuse. Should the ICS be deemed to pose high-risk for CSAM the ICSs are required to implement certain 'mitigating measures'. The Proposal leaves the exact choice of the mitigating measures to the ICSs, and merely provides a list of safeguards that need to be taken into account by the ICS when implementing such measures.

Article 4 of the CSAM Regulation Proposal does mention some possible risk mitigation measures that can be implemented by the providers. One of these, namely "*the internal supervision of the functioning of services*" conceptually leaves room for a broad interpretation, and one might argue that voluntary detection technologies could fall under this phrasing. This would essentially build on the notion of further processing from the GDPR, namely that CSAM detection and prevention is not an independent processing activity that requires a separate legal basis, but rather that it should be considered as compatible with to the principal data processing activity of providing safe and secure ISS (much like the automated detection of viruses and automated blocking of spam and phishing). The fact that this possibility is suggested in the CSAM Regulation, in tandem with appropriate transparency safeguards, would support a finding that such screening is in line with the reasonable expectations of the data subjects. This would bring CSAM screening closer to other forms of general, indiscriminate and

[45] Recital 17 of the CSAM Regulation Proposal.

automated analysis of ICS, such as spam and virus detection, which are considered forms of compatible and lawful further processing of the personal data.

However, taking into account the case law of the CJEU on interference with fundamental rights, and the fact that ICS must adhere to the ePrivacy Directive in addition to the GDPR, opponents of this perspective would point out that voluntary detection and prevention of unlawful content are not an inherent part of ICS, and that they should be prescribed by law which lays down clear and precise rules that determine the scope and the application of the intended measure, and that impose minimum safeguards. The current general wording used in the Proposal in relation to risk mitigation measures would likely not satisfy these criteria.

The gist of the text of the Proposal appears to be that the only legally supported possibility for ICSs to continue to use detection technologies is provided by the legislature in the CSAM Regulation Proposal is under the mechanism of detection orders which will be discussed briefly below. Voluntary detection is likely to be eroded in practice, irrespective of the intentions indicated in the recitals.

3. Detection orders under the CSAM Regulation Proposal

The mechanism of detection orders is completely new, and intends to specifically target known and unknown CSAM online, providing at the same time greater legal certainty.[46] The proposed mechanism of detection orders is intended to work as follows: A competent judicial authority (or independent administrative authority) can issue a detection order requested by the Coordinating Authority to a service provider when there is evidence of a significant risk of the service being used for the purposes of online child sexual abuse and additionally it must "outweigh the negative consequences for the rights and interests of all parties affected".

The detection technologies and safeguards that can be imposed by detection orders are specified in Article 10 of the CSAM Regulation Proposal. Again, the Proposal remains in this regard technological-neutral, stating in Recital 26 that the choice of the technologies to be operated to comply with a detection order will be left to the provider. The safeguards that the detection technology must fulfil are similar as the ones that were already included in the Interim Derogation, namely effectiveness, proportionality, in accordance with the state of art in the industry and sufficiently reliable.[47]

The same can be said for the obligations that the ICSs will need to respect when implementing a detection technology to comply with a detection order. These obligations are the following[48]:

- take measures to ensure that the technologies and indicators, are used for the sole purpose of detecting the dissemination of known or new CSAM or the solicitation of children and *that the processing is strictly necessary to execute the detection order*;
- establish effective internal procedures to prevent and, where necessary, detect and remedy any misuse of the technologies, indicators and personal data, including unauthorized access to, and unauthorized transfer of, such personal data;
- ensure regular human oversight and, where necessary, human intervention;

[46] See Section 2 Detection obligations of the CSAM Regulation Proposal.
[47] Article 10 (3) of the CSAM Regulation Proposal.
[48] Article 10 (4) of the CSAM Regulation Proposal.

- establish and operate an accessible, age-appropriate and user-friendly mechanism that allows users to submit to it, within a reasonable timeframe, complaints about alleged infringements of the ICS's obligations, including about the use of detection technologies, removal or disabling of access to material, blocking users' accounts or suspending and termination the provision of the service to users. The complaints must be processed in an objective, effective and timely manner;
- inform the Coordinating Authority, at the latest one month before the start date specified in the detection order, on the implementation of the envisaged measures;
- regularly review the functioning of the measures mentioned above;
- inform the users of the services about the detection technologies used, including the way in which they operate and the impact on the confidentiality of user's communications, and about the reporting obligations to the to be established EU Centre.

4. Concluding remarks on the CSAM Regulation Proposal

The current text of the CSAM Regulation Proposal thus does not support voluntary detection efforts of ISSs. Article 1 (4) and one of the Recitals of the Proposal even expressly states that "*this Regulation limits the exercise of the rights and obligations provided for in Article 5 (1) and (3) and Article 6 (1) of the ePrivacy Directive insofar as necessary for the execution of detection orders*".[49] By explicitly limiting the scope of the derogation to the detection orders, the EU is removing the explicit legal basis for ICSs to continue to use voluntary detection technologies on their services, as in this case they would fall under the scope of the strict prohibition of the ePrivacy Directive. While there is some language suggesting otherwise in the proposal and in the Commission's Memorandum, the legislative support for voluntary detection is clearly eroded by the CSAM Regulation.

The continuation of voluntary detection efforts by ICSs is of paramount importance to ensure the proactively detection of CSAM online, even if they have not received a detection order which mandates them to do so. The ability to react quickly is essential in the fight against CSAM online. The fact that ICSs would have to wait on a detection order that is issued by a competent judicial authority necessarily means lost valuable time, and less consideration of the ability and willingness of at least some ISSs to act voluntarily in protecting users against CSAM as well as the fundamental rights of the CSAM victims. This does not seem in line with the EU's policy objectives, and creates a clear risk in the fight against CSAM.

In our concluding chapter, we will examine alternatives that might be considered: how could the EU retain (and improve) the possibility of voluntary detection, possibly via the CSAM Regulation itself, in a manner that would satisfy the interference test established in the case law of the CJEU? If a suitable answer can be found to this question, this would represent a significant boon to the fight against CSAM, in line with public policy interests, the priorities of diligent ICSs, and the preferences of children's rights organisations that have worked with ICSs to ensure the effectiveness of their efforts.

[49] Article 1 (4) CSAM and Recital 9 Regulation Proposal.

As a final remark, it is worth noting that the interinstitutional discussions have already proven to be very difficult, and it seems unlikely that a political agreement on a final CSAM Regulation will be reached before reaching the original deadline of the Interim Derogation. Given these challenges, the Commission had already proposed an extension of the Interim Derogation for a limited period of time in order to be able to continue political discussions. In January 2024, the EDPS provided a negative opinion[50] on the Proposal for an extension of the Interim Derogation, bringing forward similar arguments as the one that were mentioned in her previous Opinion on the Interim Derogation itself.

Nevertheless, in order to avoid the negative impacts of the potential impasse, on 15 February 2024, the LIBE Committee announced[51] that a political compromise was reached on the extension of the Interim Derogation by 18 months, i.e. until 3 April 2026. This however also implies that it is less likely that a political agreement on the CSAM Regulation Proposal would be reached before the June 2024 deadline.

E. Potential solutions: combining legal certainty, effectiveness, and fundamental rights

Firstly, the CSAM Regulation as proposed by the EU Commission, in combination with the triggering of the sunset clause of the Interim Derogation, would not unambiguously make the voluntary detection of CSAM unlawful for all ICSs. The text of the CSAM Regulation Proposal leaves some margin for interpretation: Article 4 of the Regulation does authorize – and in fact require – ICSs (and hosting service providers) to "take reasonable mitigation measures" to combat CSAM, which must be tailored to the risk that they have identified under the Regulation, to minimise that risk. The measures may include:

"*(a) adapting, through appropriate technical and operational measures and staffing, the provider's content moderation or recommender systems, its decision-making processes, the operation or functionalities of the service, or the content or enforcement of its terms and conditions;*

(b) reinforcing the provider's internal processes or the internal supervision of the functioning of the service;

(c) initiating or adjusting cooperation, in accordance with competition law, with other providers of hosting services or providers of interpersonal communication services, public authorities, civil society organisations or, where applicable, entities awarded the status of trusted flaggers in accordance with the Digital Services Act".

The limitations of the mitigation measures are described in paragraph 2 of the Article, which includes requirements in terms of effectiveness, proportionality, diligence and non-discrimination, as well as various other points. Collectively, these could be used as arguments to justify the further voluntary use of CSAM detention and prevention mechanisms, and as justifications as to why voluntary CSAM

[50] Opinion 8/2024 of the EDPS on the Proposal for a Regulation amending Regulation (EU) 2021/1232 on a temporary derogation from certain ePrivacy provisions for combating CSAM; see https://www.edps.europa.eu/system/files/2024-01/2023-1261_d0219_opinion_en.pdf

[51] See https://www.europarl.europa.eu/news/en/press-room/20240212IPR17636/child-sexual-abuse-online-agreement-on-extending-current-rules-until-april-2026

schemes could indeed satisfy the interference test under EU law: the entirety of the regulatory framework must be considered, and the fundamental rights challenges on this point are not at all comparable to those under the data retention cases.

By way of comparison, the PNR Directive was upheld by the CJEU, despite the fact that it was considered to entail a serious interference with the rights guaranteed in Article 7 and 8 of the Charter (as CSAM detection and prevention would undoubtedly also be), and despite the universality of the PNR transfers (in the sense that the data processing was not limited to persons with a particular profile, much as CSAM detection also would not be limited to suspicious persons). These elements of serious interference and indiscriminate application were not sufficient for the PNR Directive to fail the interference test, as the Court also considered that the Directive served a general objective of general interest (as voluntary CSAM screening unquestionably also does), and that the Directive involved sufficient safeguards – such as the application of a post-hoc individual human review, and the use of pre-determined assessment criteria. These measures could reasonably be implemented for voluntary CSAM detection and prevention also, so that the interference test could be satisfied.

However, the lack of a clear decision on this point within the CSAM Regulation - which states in its recitals that voluntary detection should remain possible, while not creating unambiguous rules on this point - is detrimental to the fight against CSAM.

With that in mind, and assuming that one agrees that voluntary CSAM detection and prevention should be sustained by legislation rather than being a possibility that could occur in a legal grey area, this final Chapter of this White Paper will explore some policy options for supporting voluntary CSAM schemes in a manner that is more likely to survive an interference challenge.

1. Option 1 – the status quo and a limited extension of the Derogation

A first and rather theoretical option is that the CSAM Regulation is abandoned, and the Interim Derogation is merely extended for a certain period of time.

This has not been seen as a viable option by the EU Commission, and the authors of this White Paper agree, on the basis that it would be clearly beneficial to institute a clear and harmonized EU level regime for mandatory CSAM detection and prevention as the CSAM Regulation envisages; and on the basis that the Interim Derogation has certain flaws as highlighted in the sections above. These issues cannot be resolved by merely extending the Derogation's duration.

It should be noted however that an urgent temporary extension of the Interim Derogation on the other hand (as the Commission proposed in November 2023, and as has been agreed upon at the political level on 15 February 2024) would be very beneficial, in the view of the authors of this Paper. While it is imperfect, the expiration of the Interim Derogation without any clear alternative would be a grave blow to the fight against CSAM, since it would create a new legal avenue for perpetrators to argue that evidence against them was collected in violation of their fundamental rights. This is a risk that the EU should not accept. With that in mind, an extension of the Derogation to ensure that no short-term gap exists in EU law, should indeed be considered.

2. Option 2 – adopting the CSAM Regulation Proposal and extending the Interim Derogation indefinitely

A second possibility would be to retain (and of course adopt) the CSAM Regulation Proposal as a legal framework for mandatory CSAM detection and prevention, while also removing the sunset clause from the Interim Derogation. This would essentially extend the Derogation's duration indefinitely, and thus treat it as a complementary and permanent framework to more emphatically support voluntary screening.

Since this approach arguably removes an inherent safeguard of the Interim Derogation – namely the fact that it only allowed temporary processing activities – such an extension might trigger a legal challenge before the CJEU, arguing that the indefinite nature of the Interim Derogation would violate the interference test. The outcome of such a challenge is unpredictable, but in principle the additional risk created by an indefinite extension (compared to the status quo where it is retained as a time-limited Derogation) should be limited. It seems unlikely that the CJEU would sustain the Interim Derogation as currently written, but reject an indefinitely extended "Permanent Derogation", since it would essentially have to rule that violations of the principles of legitimacy, lawfulness and proportionality are acceptable if they only last for a defined period of time.

Nonetheless, while a Permanent Derogation would provide enhanced legal certainty to ICSs performing voluntary detection– in the sense that their voluntary screening activities remain clearly qualified as permissible, including any activities initiated prior to the initial adoption of the Interim Derogation – the problems of the Derogation that have been briefly summarized above would remain unresolved. This is not only less than ideal from a public policy perspective, but also creates uncertainty for the ICSs themselves. The Interim Derogation was designed as a temporary stopgap measure, rather than as a permanent framework. To further support legal stability and ensure a better integration of the expectations towards voluntary and mandatory detection and prevention, a better and more stable alignment of both legal frameworks (the Interim Derogation and the CSAM Regulation) would be needed.

3. Option 3 – adopting a modified CSAM Regulation that sustains and improves voluntary CSAM detection and prevention

3.1. Suggestion for safeguards

A third option, which the authors of this paper consider to be preferable, is to modify the proposal for a CSAM Regulation by integrating a more explicit legal mandate for voluntary CSAM detection and prevention by ICSs, complementing the already foreseen approach that provides a legal framework for mandatory screening under specific orders.

This modification can build upon the approach of the Interim Derogation, by regulating the conditions under which voluntary screening is permissible, but should also integrate revisions to (i) better align the requirements for voluntary screening to those of mandatory screening; and (ii) to rectify some of the weaknesses of the Interim Derogation.

In this Paper, we have argued that, to ensure that voluntary screening satisfies the three-factor interference test with respect to fundamental rights, the Interim Derogation's weaknesses in relation to transparency, screening methods and reaction procedures should be addressed. There exist several options through which this could be done, any one of which (or a combination thereof) could be integrated into a revised CSAM Regulation that explicitly supports voluntary CSAM screening in ICS with safeguards that would in our opinion, satisfy the three-factor test:

- Firstly, comparable to the Terrorist Content Online Regulation, the permissibility to conduct voluntary CSAM screening can be made subject to a specific definition of safeguards and accountability requirements, which could include improved **disclosure duties** towards ICS users, in combination with an annual reporting obligation to the Coordinating Authorities. This would be one way to reduce the transparency gap that currently exists in the Interim Derogation, and would facilitate independent oversight and enforcement over voluntary detection and prevention practices.
- Secondly, **enhanced oversight mechanisms** could be introduced that build on the current requirement of the Interim Derogation to conduct a data protection impact assessment for any given technology and to subject these to a prior consultation. The approach of the Interim Derogation is arguably flawed, since the impact assessment (1) focuses only on data protection, and not on fundamental rights in general (other relevant rights include the right to non-discrimination, the right to a fair trial, and the presumption of innocence); (2) as a result, is submitted for prior consultation to data protection authorities under the GDPR, and not to Coordinating Authorities under the CSAM Regulation (which was of course inevitable, since there was no CSAM Regulation when the Derogation was adopted; but this is none the less a problem that should be fixed moving forward); and (3) the impact assessment is conducted at the level of "any specific technology used" – i.e. it targets a technology, rather than its specific use by a ICS, thus not recognising the importance of specific deployment choices that could be made by an ICS. These too are issues that could be rectified by an amended CSAM Regulation.
- A stronger variant of the DPIA approach would be for ICS providers to **submit their assessments to the Coordinating Authorities to obtain a prior authorisation** (either via explicit approval decisions, or by allowing the Coordinating Authorities to challenge the assessments where they deem necessary). While prior authorisation schemes are inherently less flexible (notably because of the inevitable doubts on whether re-authorisation is required when some elements of the solutions used evolve), this would constitute a strong safeguard.
- A further safeguard could consist of a more granular approach to CSAM that recognises that fundamental rights risks are different depending on the unlawful content and the detection technology applied. Concretely:
    - Known CSAM – hash-based matching technologies paired with known CSAM hash values have a very limited false positive rate. Voluntary detection and prevention that focuses exclusively on known CSAM is therefore a measure that is inherently more likely to pass the three-factor test, since the risk of false positives was a key element in e.g. the CJEU's PNR Decision.
    - Unknown CSAM and grooming inherently and inevitably are less accurate. They pose significantly higher risks of false positives, which can have detrimental impacts on the rights and interests of the persons concerned. Voluntary detection and prevention that includes unknown CSAM and grooming should therefore be held to a higher standard, such as e.g. the mandatory human intervention ("human in the loop") that is already foreseen in the current Interim Derogation. For such content, automated detections should be individually screened and evaluated by a skilled individual,

before any further action is taken (including before sharing the content with any third party); and no data should be retained where detected content was labelled as a false positive (except in a purely aggregate statistical form that could not be linked further to any individual person).

- Additionally, the CSAM Regulation could – like the GDPR and the Digital Services Act – encourage the **drafting of Codes of Conduct** for ICSs at the EU level, to be approved by the EU Centre in coordination with the European Data Protection Board. It is worth noting that this approach is not new, neither in general nor in relation to fighting unlawful online content: the Commission has supported the adoption of an EU Code of Conduct on countering illegal hate speech online, which was adopted in May 2016, and has been supported by Facebook, Microsoft, Twitter and YouTube. A comparable – but more tightly regulated – process could be implemented, resulting in a coregulatory model where the industry is allowed and encouraged to engage in CSAM detection and prevention, subject to substantive requirements set by an amended CSAM Regulation, and subject to approval of the Code at the EU level.

3.2. Building adequate safeguards in light of the case law of the CJEU

The authors have discussed in section B.2 of this White Paper the interference criteria developed by the CJEU on an interference of fundamental rights: lawfulness, legitimacy, and proportionality. In this section, the authors will discuss the way in which the interference criteria can be satisfied by implementing all or some of the above-mentioned safeguards.

Concerning the **first criterion** (lawfulness) it must be clear that the possibility to implement voluntary detection technologies must be more explicitly written into the CSAM Regulation. Moreover, in order to fulfil the criteria it is necessary that the law is sufficiently clear, precise and complete.

In that regard, the detection orders regime in the CSAM Regulation has been criticized by the LIBE committee in the European Parliament as not being sufficiently clear and precise with regards to the detection technologies that can be used.[52]

However, the authors are of the view that the technology-neutral approach of the CSAM Regulation should be applauded as it is a crucial aspect in the fight against CSAM online. Indeed, the types of abuse are rapidly changing as a result of technological advances and new ways of sharing CSAM online, which requires the continuous development of new technologies to tackle this. Compliance with the lawfulness criterion in this context does not require defining the specific detection technologies to be used. However, it should be noted that the Proposal should strongly recommend ICSs to use the detection technologies that will be developed (or provided) by the EU Centre, as it prevents that some smaller ICSs and/or start-ups will opt for non-transparent, commercial software.

Regarding the **second criterion** (legitimacy), it is clear that the CSAM Regulation fulfils an objective of general interest recognized in the European Union. The CJEU has recognized in its judgments on the Data Retention Directive that "*the fight against serious crime in order to ensure public safety constitutes a general objective recognized in the EU*"; and in relation to the PNR Directive that the need

[52] Committee on Civil Liberties, Justice and Home Affairs, Report on the proposal for a regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse, https://www.europarl.europa.eu/doceo/document/A-9-2023-0364_EN.html.

to ensure the safety of EU citizens was in accordance with the legitimacy requirement. The protection of children in the context of CSAM detection should satisfy this requirement as well.

The objective pursued by the CSAM Regulation Proposal is to harmonize rules that apply to prevent and combat child sexual abuse, which is a particularly serious crime.[53] Therefore, in the hierarchy of objectives of public interests as set out in section 2.2 of this report, the CSAM Regulation Proposal falls under "combatting serious crime", which according to *La Quadrature du Net*, only allows for limited and more targeted interference.

The detection tools used by ICSs have been criticized as constituting a general monitoring of interpersonal electronic communications which would be considered an excessive interference in light of the case law of the CJEU.

However, the authors consider that this assessment is incomplete as it focuses on only one aspect of "targeted interference", namely the volume of users affected. In reality, however, the assessment of whether or not the interference is targeted is more nuanced, and should take into account the *entire context* of the detection and prevention activities.

The CSAM Regulation Proposal requires ICSs to analyse and assess the risk of use of the service for the purpose of online child sexual abuse, which is a first constraint in terms of targeting. Next, the Proposal requires measures to be implemented which are explicitly linked to the outcome of this assessment, thus creating a second layer of targeting constraints.

When applying these principles to a concrete situation, we see for example that the use of hash based detection of known CSAM images inherently targets only communications with media content attached, and only triggers a response when there is a match to the CSAM database. In that sense, we believe that it meets with the targeted approach prescribed by the CSAM Regulation Proposal - i.e. only when a ICSs has identified a risk of misuse for a specific services, can it implement certain risk mitigation measures, including voluntary detection measures which are limited to that service, and which are targeted towards, and limited to, the risks identified.

The **last criterion**, namely the proportionality test, will inevitably be the most challenging to fulfil, due to the level of intrusiveness of some of the technologies used, i.e. the application of measures to ICS messages of all users, in order to detect the possible occurrence of CSAM. However, there are reasonable grounds to believe that the implementation of the above-mentioned safeguards will ensure a favourable balancing test in light of the case law of the CJEU mentioned in section B.2:

- As with the PNR Directive, it is strictly necessary for ICSs to apply the measures to the communications of all their users if a high risk has been identified, because it will be impossible to achieve the same objectives if the scanning would be limited to a sub-set of users (i.e. those of whom there is an indication that they are disseminating CSAM). Identifying a proclivity for CSAM would mean that the ICSs would have to start undertaking other types of far-reaching and privacy-intrusive profiling and data collection measures to assess the risk of their users, which would likely not meet the interference threshold. Moreover, in principle ICSs will not have access to the specific information that is necessary to classify its users in narrower way (i.e. location, age, gender, etc.): they do not need this information to provide the services, and it would therefore be a violation of the users' rights to privacy and data protection to make it

[53] See impact assessment of the CSAM Regulation Proposal.

available to the ICS. Next to this, users can also easily circumvent or falsify the classifiers, i.e. by the use of a VPN. In conclusion: the indiscriminate nature of the CSAM detection activity is inevitable, once the legitimacy and lawfulness test are satisfied.

- One of the main reasons which led the CJEU to the conclusion that the PNR Directive meets the interference threshold, in spite of the indiscriminate scale, was the fact that although false-positives were common with the automatic assessment of the passenger data, there existed ex post-verification by humans. The same can be said for the voluntary detection by ICSs, since most ICSs have in-house professional analysts that will verify the results of the detection technology before forwarding it to the EU Centre (who will also subsequently assess the content). In practice, this will mean that when content is flagged as potential CSAM by an automated detection tool, it will be subject to a dual verification process. The Proposal can however be improved in this aspect by including a strict obligation for ICSs that are voluntarily using detection tools to implement an ex-post in-house human assessment (i.e. clear rules need to be introduced on the requirements for analysts, the timing of performing ex-post assessments, etc.).
- Additionally, it is worth noting that the proposed list of measures above significantly exceeds the measures that were available in e.g. the PNR case, since they include enhanced transparency (towards data subjects and towards authorities), risk assessment, and oversight mechanisms that rely on prior consultation, prior authorization, and post-hoc controls by designated authorities (i.e. the obligation to report on the mitigation measures implemented by the ICS). It could also reasonably be considered to limit the mandate for voluntary CSAM screening to known CSAM content, since this significantly diminishes the risk of false positives, and therefore also the risk of unjustified harm.
- Lastly, an important aspect that is continuously highlighted by the CJEU is the need to ensure "an effective remedy" in accordance with Article 46 of the EU Charter. In case law of both the CJEU and the ECtHR this has commonly been interpreted as including an effective *a priori* authorisations and an *ex post* oversight body (which is considered independent and has power to issue binding decisions). We have proposed to introduce a range of enhanced oversight mechanisms in section 3.1 above, including an obligation to conduct an impact assessment prior to deploying a detection or prevention measure, and to subject this to a prior consultation with the Competent Authorities in the Member States. The ex post oversight by independent bodies is already explicitly provided in the Proposed CSAM Regulation for the detection order as redress is ensured, including through requests for re-assessment by Coordinating Authorities and the right to submit complaints to the Coordinating Authority. When explicitly introducing the possibility of voluntary scanning into the text of the Proposal, the European legislature should also mandate the ex post oversight by the Coordinating Authorities.

While it cannot be reasonably denied that there is an interference in the fundamental rights to privacy and data protection when deploying CSAM detection, the assurances and safeguards provided would thus exceed those that have been accepted as appropriate in other contexts.

4. Overall conclusion and recommendation

As this White Paper shows, the existing Derogation, while an important first step in the fight against CSAM, should not be considered a perfect tool. It has clear weaknesses as a framework for voluntary CSAM detection and prevention that should be addressed to ensure that European fundamental rights are appropriately protected. The proposed CSAM Regulation is a necessary complement to the Derogation, since it introduces new protections and creates a clear regime for mandatory screening.

However, if the CSAM Regulation is adopted as proposed by the European Commission, and the Derogation is allowed to expire (either in 2024, or in 2026 as is currently considered), the legal mandate that unambiguously permits ICSs to engage in voluntary CSAM screening as it stands will disappear. Voluntary CSAM screening would undoubtedly be reduced, and where it continues, would be done in a grey and much less regulated space where the measures of the Interim Derogation will not always be observed. The dissemination of CSAM may increase altogether, and at any rate CSAM will not be as effectively combated. This is clearly not desirable from a public policy perspective – and indeed it is ambiguous whether it is the intention or desire of the EU legislator to halt voluntary screening entirely.

To avoid a situation where voluntary screening is subject to interpretation and risk acceptance, where the impacts on fundamental rights of EU citizens will differ significantly on the basis of largely invisible criteria designed by ICSs, an approach should be found where voluntary CSAM detection and prevention is given a clear fiat under the CSAM Regulation, as a measure that is permissible for the affected ICS provided that certain safeguards are met that resolve the problems of the Derogation's regime. There are ample examples of such measures in prior legislative initiatives, not only in the CSAM Regulation and in the Interim Derogation itself, but also in the GDPR, the DSA, and in the Terrorist Content Online Regulation.

It is possible, on the basis of these frameworks, to design a voluntary screening approach that builds on a co-regulatory model, where the ICSs are allowed or required to assume responsibility, while building a governance framework on top of those measures that can better support lawfulness, legitimacy and proportionality. In this manner, the Derogation could be recast into an instrument that continues to support the fight against CSAM, in a manner that more clearly aligns with EU expectations in relation to the fundamental rights to privacy and data protection.