Chat Control and the Repercussions of Mandatory Communication Detection

Reconciling the EU Child Sexual Abuse Regulation Proposal (2022/0155) with Fundamental Rights, Privacy and Cybersecurity Standards

MARIAVITTORIA SCARSINI*

Abstract: This article examines the European Union's proposed Child Sexual Abuse Regulation (commonly referred to as "Chat Control") and the legal and technical challenges arising from its transition from voluntary to mandatory detection of online child sexual abuse material (CSAM). The study situates the proposal within the broader EU regulatory framework governing electronic communications, including the shift introduced by Directive (EU) 2018/1972 and the interim derogation established by Regulation (EU) 2021/1232, later extended by Regulation (EU) 2024/1307. While the proposal seeks to harmonise Member States' approaches and strengthen the fight against online child abuse, it presupposes the availability of reliable detection technologies and introduces extensive monitoring obligations for providers of hosting and interpersonal communication services. The article analyses the technical mechanisms likely to be used to comply with detection orders, including perceptual hashing, artificial-intelligence classifiers, and client-side scanning, highlighting their operational limitations, susceptibility to evasion or manipulation, and significant rates of false positives when applied at large scale. Particular attention is given to the interaction between these detection measures and widely deployed security technologies such as end-to-end encryption, as well as to the cybersecurity risks created by large-scale processing and storage of sensitive data. The paper argues that the proposed

framework raises substantial concerns regarding data security, transparency, procedural safeguards, and the proportionality of large-scale monitoring of private communications. Although combating online child sexual abuse represents a compelling public interest objective, the current proposal risks introducing systemic vulnerabilities and intrusive surveillance infrastructures without sufficiently addressing the technological and cybersecurity challenges involved. The article concludes that any durable EU regulatory framework should integrate rigorous safeguards, transparency, technological viability, and strong protection of secure communications.

Keywords: EU Proposal 2022/0155; Child Sexual Abuse Material; Scanning Technologies; Cybersecurity; End-to-end Cryptography.

Table of contents: 1. Introduction. – 2. CSAM Detection. – 2.1. Voluntary Detection – Legal Framework Preceding 2020. – 2.2. Legalized Control – Interim Regulation (EU) 2021/1232. – 2.3. Mandatory Detection – Proposal Regulation No. 2022/0155. – 3. Analysis of the Chat Control Regulation Proposal. – 3.1. End-to-End Cryptography, Scanning Technologies and False Positives. – 3.2. Safekeeping of Data and Cybersecurity Issues. – 4. Conclusion.

1. *Introduction*

The EU defines online child abuse and exploitation as "all acts of a sexually exploitative nature carried out against a child[1] that have, at some stage, a connection to the online environment"[2]. Child sexual abuse materials, commonly referred to as CSAMs,[3] encompass any materials depicting such abuse.

---

* Mariavittoria Scarsini is a final-year law student at the University of Trento, dedicated to the study of informational self-determination and the safeguarding of fundamental rights within the digital ecosystem. Following an academic exchange at NOVA School of Law in Lisbon, she refined her focus on the nuances of informed consent and the critical role of user perception in data transparency. Her current research explores the intersection of the GDPR, DSA, and DMA frameworks, with a particular emphasis on the legal challenges posed by dark patterns and the protection of minors online. By examining the gap between formal legal requirements and the actual awareness of data subjects, she aspires to advance effective models of digital autonomy for vulnerable users.

[1] "A child is defined in this context as any natural person younger than 18 years old." Office of the United Nations High Commissioner for Human Rights (OHCHR), Convention on the Rights of the Child art. 1 (November 20, 1989), available at https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child (last visited March 18, 2026).

[2] General Secretariat of the Council, *Operational Action Plan 2022: Child Sexual Exploitation*, Doc. No. 13589/21, pt. 2.1 (November 13, 2021).

[3] European Commission,Press Release, *Fighting Child Sexual Abuse: Commission Proposes New Rules to Protect Children* (May 11, 2022), available at https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2976 (last visited December 15, 2025).

According to the press conference held by the European Commission in May 2022, the COVID-19 pandemic has exacerbated the issue of the presence of pictures and videos depicting child pornography[4]. The EU has identified this problem as severe[5], mostly because despite its previous dedication to combat Child Sexual Abuse, the European Union territories remained the primary global fulcrum for hosting this content, as sixty-two percent of all known CSAM can be traced back to an EU country[6].

Until 2021, online platforms were implementing a self-regulatory policy voluntarily enacted by them, but it quickly became clear that this method repeatedly resulted in inefficiency, also because it had to work in synergy with the national legal framework of reference emanated by single EU Member States.

In light of CJEU's recognition that combating serious crime may justify limited interferences with privacy and data-protection rights[7], the European Union adopted Regulation 2021/1232 as an interim measure, which entered into force in 2021.

---

[4] European Commission, Press Release, *Fighting Child Sexual Abuse: Commission Proposes New Rules to Protect Children* (May 11, 2022), available at https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2976 (last visited December 15, 2025); Europol, *Exploiting Isolation: Offenders and Victims of Online Child Sexual Abuse During the COVID-19 Pandemic* 5, 17 (2020), available at https://www.europol.europa.eu/publications-events/publications/exploiting-isolation-offenders-and-victims-of-online-child-sexual-abuse-during-covid-19-pandemic (last visited December 15, 2025).

[5] Consolidated Version of the Treaty on the Functioning of the European Union art. 83, June 7, 2016, 2016 O.J. (C 202) 47 ("The sexual exploitation of children is considered a serious crime").

[6] Internet Watch Foundation, *Europe Remains "Global Hub" for Hosting of Online Child Sexual Abuse Material*, available at https://www.iwf.org.uk/news-media/news/europe-remains-global-hub-for-hosting-of-online-child-sexual-abuse-material/ (last visited March 14, 2026).

[7] Joined Cases C-293/12 and C-594/12, *Digital Rights Ireland Ltd v Minister for Communications, Marine & Natural Resources*, ECLI:EU:C:2014:238.

Given the temporary nature of the measure, in 2022 the new Child Abuse Protection Regulation[8] (also known as "Chat Control") was presented. A vigorous debate sparked among the Member States regarding the threats that this Regulation poses to the GDPR regulation[9] and to the cybersecurity of the Union, delaying the implementation of the new regulation and leaving no other choice but to extend the effects of the Interim Regulations until April 2028.

The Child Abuse Protection Regulation's aim is to render aspects of the previous regulation permanent, while enforcing mandatory screening measures for service providers, potentially taking down protections that were put in place to safeguard the privacy and security of users online.

This article analyses this Regulation proposal and the challenges that it poses. In particular, the author wishes to highlight how this legislation, however noble in its purpose, assumes as given aspects such as the existence of safe scanning technologies, while disregarding major cybersecurity compliance risks.

2. *CSAM Detection*

2.1. *Voluntary Detection – Legal Framework Preceding 2020*

In these past years, providers of electronic communication services have been able to voluntarily report CSAM to authorities in order to prevent or counter child sexual abuse online. This established practice is based on scanning online content, such as images, text and traffic data of

[8] *European Commission, 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 (2022).

[9] Regulation (EU) 2016/679 (General Data Protection Regulation), 2016 O.J. (L 119) 1.

communications, sometimes using historical data, and with the use of technologies such as hashing technology[10] and artificial intelligence[11].

These voluntary detection practices essentially consisted of detecting, removing or reporting CSAM from the aforementioned platforms.

Unfortunately, this policy has proven not to be adequate in order to stop the dissemination of such materials. This is due in part, because the level of involvement of providers varies greatly due to the voluntary nature of their practices to combat child sexual abuse: some internet service providers take no action at all, while others engage in low-quality reporting of CSAM or content with a low level of relevance to child sexual abuse[12].

This is troublesome since sexual predators frequently interact on open platforms, not just the dark web[13].

The overall situation was also aggravated by the lack of harmonization among Member States, which were individually trying to regulate the subject. This increased fragmentation of the legal background also influenced online service providers who, given the intrinsic cross-border

[10] The purpose of hashing is to identify different copies of the same digital file. First, an algorithm is used to assign a unique string of characters, called a "hash," to the digital file, and that hash is stored in a hash database. A second digital file is then put through the same hashing process, and the resulting hash is compared with the one already stored in the hash database to determine whether the files match exactly. European Union Intellectual Property Office (EUIPO), "Hashing," *Anti-Counterfeiting and Anti-Piracy Technology Guide*, available at https://euipo.europa.eu/anti-counterfeiting-and-anti-piracy-technology-guide/technologies-digital-media/hashing (last visited March 14, 2026).

[11] Regulation (EU) 2021/1232 of the European Parliament and of the Council of July 14, 2021, 2021 O.J. (L 274) 41, recital 7 (addressing the use of technologies by providers of number-independent interpersonal communications services for the purpose of combating online child sexual abuse).

[12] Leonore ten Hulsen, *Digital Fixes and Techno-Solutionism: The EU's Tech-Based Battle Against Child Sexual Abuse*, 16 *NEW J. EUR. CRIM. L.* 154 (2025).

[13] INHOPE, *Webinar: How Predators Online Hide in Plain Sight* (2024), available at https://inhope.org/articles/webinar-recap-how-predators-online-hide-in-plain-sight (last visited March 18, 2026).

nature of the provision of online services, struggled with compliance and unequal conditions, as well as took advantage of possible loopholes[14].

Consequently, the European Union opted to regulate the matter at the Union level. As a result, Directive (EU) 2018/1972[15] was adopted and entered into force at the end of 2020, modifying the definition of "publicly available electronic communications services" contained in Directive 2002/58/EC[16]. By expanding this definition to include certain online communication services, the reform brought internet service providers within the scope of Articles 5(1) and 6(1) of the ePrivacy Directive[17], thereby rendering their voluntary CSAM reporting practices unlawful[18]. This led to concerns among internet service providers about repercussions, with Facebook threatening to stop its voluntary proactive scanning if the Regulation entered into force[19], and a general paroxysm of preoccupation among other countries.

Due to these preoccupations, along with the ongoing intention of contrasting child abuse, the EU enacted Regulation 2021/1232 as an interim regulation.

2.2. *Legalized Control - Interim Regulation (EU) 2021/1232*

Interim Regulation 2021/1232 provided for a temporary derogation from certain provisions of Directive 2002/58/EC to allow for the processing of personal data by internet service providers for detecting and

[14] *European Commission, Proposal for a Regulation Laying Down Rules to Prevent and Combat Child Sexual Abuse,* COM (2022) 209 final, Introductory Memorandum § 2 (2022) ("Proportionality").

[15] Directive (EU) 2018/1972 (European Electronic Communications Code), 2018 O.J. (L 321) 36.

[16] Directive 2002/58/EC (ePrivacy Directive), 2002 O.J. (L 201) 37.

[17] *Id.*

[18] Recital 9 of Regulation (EU) 2021/1232; Directive (EU) 2018/1972; Directive 2002/58/EC.

[19] Gabriel JX Dance and Adam Satariano, *E.U. Privacy Rule Would Rein in the Hunt for Online Child Sexual Abuse* (N.Y. Times, December 4, 2020).

removing CSAM from their services, pending the enactment of a long-term EU legal framework against child sexual abuse[20].

In fact, the interim Regulation allows internet service providers to use technologies to detect, report, and remove CSAM without breaching the confidentiality obligations of the ePrivacy Directive.

Given the temporary nature of this Regulation, Member States started discussing a less intrusive and more privacy and cybersecurity-oriented permanent legislation; accordingly, the European Union chose to implement the Regulation that went into effect in 2021 until 2024.

Thus, in 2022 the Child Abuse Protection Regulation[21] (commonly known as "Chat Control") was presented. A dispute arose among Member States and legal associations[22] about the threats that this law poses to the GDPR and the Union's cybersecurity, consequently postponing the implementation of this new regulation.

Extensive criticism from *inter alia* the European Parliament Research Service, NGOs, the European Data Protection Board, the European Data Protection Supervisor, the European Commission Regulatory Scrutiny Board, and the Legal Service of the Council of the European Union was expressed. Above all, even the European Data Protection Supervisor has expressed its concerns about this extension[23]. It argues that the interim

[20] *Regulation (EU) 2021/1232*, recital 10.

[21] P*roposal Regulation 2022/0155.*

[22] *European Digital Rights, Fight Chat Control: Reports on the Standpoint of EU Member States on the Chat Control Proposal (2023), https://edri.org/our-work/chat-control-what-is-actually-going-on/ (last visited April 12, 2026).*

[23] European Parliamentary Research Service (EPRS), *Complementary Impact Assessment of the Proposal for a Regulation Laying Down Rules to Prevent and Combat Child Sexual Abuse,*DOI: 10.2861/016876. (last visited December 18, 2025); Ella Jakubowska, "Leaked Opinion of the Commission Sets off Alarm Bells for Mass Surveillance of Private Communications" (European Digital Rights, EDRi), available at: https://edri.org/our-work/leaked-opinion-of-the-commission-sets-off-alarm-bells-for-mass-surveillance-of-private-communications/ (last visited December 18, 2025); Daniel Boffey, *EU Lawyers Say Plan to Scan Private Messages for Child Abuse May Be Unlawful* (The Guardian, May 8, 2023), available at:

Regulation does not properly protect individuals' right to privacy and data protection, and the confidentiality of their communications.

For these reasons, the EU ultimately adopted Regulation (EU) 2024/1307 on 29 April 2024, amending the interim Regulation (EU) 2021/1232. It initially extended the Regulation's application until April 2026[24], and subsequently until April 2028[25], while pursuing a more appropriate long-term solution in the interim.

https://www.theguardian.com/world/2023/may/08/eu-lawyers-plan-to-scan-private-messages-child-abuse-may-be-unlawful-chat-controls-regulation (last visited December 18, 2025); EDPB-EDPS, *Joint Opinion 04/2022 on the Proposal for a Regulation Laying Down Rules to Prevent and Combat Child Sexual Abuse* (2022, available at: https://www.edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-042022-proposal_en (last visited December 18, 2025); European Commission Regulatory Scrutiny Board, *Opinion—Impact Assessment/Regulation on Detection, Removal, and Reporting of Child Sexual Abuse Online, and Establishing the EU Centre to Prevent and Counter Sexual Abuse* (2022) , available at: https://cdn.netzpolitik.org/wp-upload/2022/03/2022_03_Impact_Assessment_LEAK.pdf (last visited December 18, 2025); Council Legal Service, *Opinion of the Legal Service—Proposal for a Regulation Laying Down Rules to Prevent and Combat Child Sexual Abuse—Detection Orders in Interpersonal Communications* (Council of the European Union 2023) 8787/23, available at: https://www.bitsoffreedom.nl/wp-content/uploads/2023/05/20230426-opinion-legal-services-on-csar-proposal.pdf (last access December 18, 2025).

[24] *European Digital Rights, Fight Chat Control: Reports on the Standpoint of EU Member States on the Chat Control Proposal (2023), https://edri.org/our-work/chat-control-what-is-actually-going-on/ (last visited April 12, 2026).*

[25] *European Commission, Proposal for a Regulation Amending Regulation (EU) 2021/1232 as Regards the Extension of Its Period of Application*, COM(2025) 797 final (2025), https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2025)797&lang=en (last visited March 19, 2026).

2.3. *Mandatory Detection - Proposal for a Regulation to Prevent and Combat Child Sexual Abuse, Number 2022/0155*

Proposal number 2022/0155[26] in its intentions aims at giving targeted measures proportional to the risk of a misuse of a given service for online child sexual abuse and grooming[27].

In theory, the proposed Regulation seeks to comply with the underlying requirement of fairly balancing the various conflicting fundamental rights at stake that underlies that restriction, taking into account the specific context of combating online child sexual abuse and the importance of the public interest at stake while justifying its measures as removing obstacles from the internal market under article 114 of the TFEU[28].

In practicality, the proposal consists of introducing an obligation on providers of hosting services and providers of interpersonal communication services to detect and report online child sexual abuse[29], while imposing also other obligations, such as creating risk assessments to identify, analyze and assess the risk of use of their services for the purpose of online child material detection[30].

Although the proposed Regulation purports to resolve the compliance issues arising from the 2020 framework, it largely reiterates the same mechanisms that were temporarily authorised under the Interim Regulation.

3. *Analysis of the Chat Control Regulation Proposal*

The first approach the proposed Regulation takes to prevent and reduce child sexual abuse is to impose a monitoring obligation on internet

[26] *Child Abuse Protection Regulation* (commonly known as "Chat Control").

[27] *Proposal for a Regulation of the European Parliament and of the Council Amending Regulation (EU) 2021/1232.*

[28] *Commission Proposal for a Regulation 2022/0155, supra note 14, at Introductory Memorandum pt. 2 ("Legal basis") (citing TFEU art. 114).*

[29] *Commission Proposal for a Regulation 2022/0155, supra* note 14, at art. 1.

[30] *Commission Proposal for a Regulation 2022/0155, supra note 14, at recital 14, art. 3 ("Risk Assessment").*

service providers regarding CSAM, as outlined in Article 7. This entails that EU Member States can issue detection orders to internet service providers that require these companies to "take measures" to detect CSAM on their services[31].

The first issue encountered is the widespread applicability of these provisions: by definition the proposed rules only apply to providers of online services which have proven to be vulnerable to misuse for the purpose of dissemination of child sexual abuse material or solicitation of children, principally by reason of their technical features[32]: nowadays, this term might cover all major communication platforms. In general, these would include encrypted and non-encrypted social media platforms and instant messaging apps like WhatsApp, Snapchat, and Messenger[33].

By imposing this obligation, the Regulation requires service providers to assess the vulnerability of their platforms to the dissemination of child sexual abuse material. However, carrying out such assessments would likely require providers to identify or verify the age of their users more systematically, potentially raising concerns regarding compliance with GDPR principles[34].

Moreover, the detection of data is allowed solely in communications involving a minor. This will require the use of age verification systems, which nowadays rely on biometric processing and/or profiling[35].

---

[31] *Commission Proposal for a Regulation 2022/0155, supra note 14, at § 2, art. 7*

[32] Commission Proposal for a Regulation 2022/0155, supra note 14, at Introductory Memorandum pt. 2 ("Proportionality").

[33] See Desara Dushi, *Does the End Justify the Means? The European Commission's Proposed Regulation Laying Down Rules to Prevent and Combat Child Sexual Abuse*, 32 *International Journal of Law and Information Technology* (2024), available at https://academic.oup.com/ijlit/article/doi/10.1093/ijlit/eaae027/7887512?guestAccessKey= (last visited April 12, 2026)

[34] *Regulation (EU) 2016/679*, recital 38 ("Children merit specific protection with regard to their personal data, as they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data").

[35] CNIL, *Online Age Verification: Balancing Privacy and the Protection of Minors*, available at https://www.cnil.fr/en/online-age-verification-balancing-privacy-and-protection-minors (last visited March 18, 2026).

Thus, the implications of this proposed Regulation are essentially twofold: on the one hand, it raises the question of which technologies media companies would be required to use to scan private communications; on the other hand, it concerns the potential risks arising from a failure to ensure adequate data security.

3.1. *End-to-End Cryptography, Scanning Technologies and False Positives*

The first issue concerns the technological means that providers are expected to deploy for detecting CSAM.

While the Regulation claims to promote innovation, proportionality, and technological neutrality by refraining from prescribing an exhaustive list of mandatory mitigation measures, this purported flexibility effectively authorises providers to select any detection technology so long as the regulatory requirements are nominally met: this allows them a degree of flexibility to design and implement measures tailored to the risk identified and the characteristics of the services they provide and the manners in which those services are used.

For the purpose of giving a more straightforward answer, providers are not required to use any specific technology, as long as the requirements set are met[36].

Such an open-ended approach has been criticized by scholars, who warn that the absence of defined standards invites inconsistent interpretations and potential misuse of surveillance tools, thereby heightening cybersecurity and privacy risks[37]. In particular, according to the EDRi open

[36] *Proposal for a Regulation*, supra note 14, art. 10(2).

[37] "Passing this legislation undermines the thoughtful and incisive work that European researchers have provided in cybersecurity and privacy, including contributions to the development of global encryption standards. Such undermining will weaken the environment for security and privacy work in Europe, lowering our ability to build a secure digital society", European Digital Rights (EDRi), *Joint Statement of Scientists and Researchers on EU's Proposed Child Sexual Abuse Regulation* (2023), available at https://edri.org/wp-content/uploads/2023/07/Open-Letter-CSA-Scientific-community.pdf (last visited March 18, 2026).

academia letter "the effectiveness of the law relies on the existence of effective scanning technologies. Unfortunately, the scanning technologies that currently exist and that are on the horizon are deeply flawed. These flaws mean that scanning is doomed to be ineffective. Moreover, integrating scanning at a large scale on apps running in user devices, and particularly in a global context, creates side-effects that can be extremely harmful for everyone online, and which could make the Internet and the digital society less safe for everybody"[38].

De facto, any generalised monitoring obligation inevitably interferes with encryption, as the detection of CSAM within interpersonal communications presupposes widespread access to content.[39]

It is worrisome that the Regulation would erode the confidentiality guarantees of end-to-end encryption[40] (E2EE), a technology widely used by communication services such as WhatsApp[41] to ensure that only intended recipients can decrypt messages, given that the Regulation will serve as a mandatory basis for indiscriminate interception of content communications. To further explain this concept, encryption is the scrambling of plaintext messages, turning them into unreadable code that can only be deciphered by those who have the secret key. End-to-end encryption is one of the most commonly used technologies to secure and send information across the internet. Hardware embedded into phones or phone applications allow for the random locks and keys that make E2EE only

[38] Ibid.

[39] Leonore Ten Hulsen, *Digital Fixes and Techno-Solutionism*, supra note 12, 154–75.

[40] "It is a means to protect individuals, civil society, critical infrastructures, media and journalists, industry, and governments by ensuring the privacy, confidentiality, data integrity and availability of communications and personal data; it is evident that all parties benefit from encryption technology." Council of the European Union, *Resolution on Encryption* (2020), available at https://data.consilium.europa.eu/doc/document/ST-13084-2020-REV-1/en/pdf (last visited March 18, 2026)

[41] WhatsApp, *FAQ on Encryption*, available at https://faq.whatsapp.com /820124435853543/?helpref=uf_share (last visited March 18, 2026).

work on the devices involved in the conversation[42]. By design, E2EE protects users from eavesdropping and unauthorised access, including by service providers themselves.

These characteristics are the reasons why critics fear that the Regulation implicitly requires the dismantling of these safeguards, on the grounds that encryption purportedly offers excessively robust protection[43].

Some deemed this type of protection granted to the user far too safe. In fact, with this proposal it is understood how the EU is asking to dismantle these safeguards[44]: there is a growing risk to public safety as criminals, and in this regard CSAM, are drawn to the use of E2EE apps that are technically impossible to access.

Furthermore, this technology also provides users with a sense of security, as their data stays safe during transmission, and neither third parties nor the provider itself can access it.

---

[42] Robert E. Endeley, *End-to-End Encryption in Messaging Services and National Security—Case of WhatsApp Messenger*, 9 *Journal of Information Security* 95, 95–99 (2018).

[43] "End-to-end encryption is designed so that only the sender and recipient can view the content of a message or other communication. Encryption is the only tool we have to protect our data in the digital realm; all other tools have been proven to be compromised. The use of link encryption (from user to service provider and from service provider to user) with decryption in the middle as used in the mobile telephone system is not an acceptable solution in the current threat environment. It is obvious that end-to-end encryption makes it impossible to implement scanning for known or new content and detection of grooming at the service provider. In order to remedy this, a set of techniques called "Client-Side Scanning" (CSS) has been suggested as a way to access encrypted communications without breaking the encryption. Such tools would reportedly work by scanning content on the user's device before it has been encrypted or after it has been decrypted, then reporting whenever illicit material is found. One may equate this to adding video cameras in our homes to listen to every conversation and send reports when we talk about illicit topics", European Digital Rights (EDRi), *Joint Statement of Scientists and Researchers on EU's Proposed Child Sexual Abuse Regulation* (cited in note 37).

[44] *Id.* ("Governments and secret services, on the other hand, are asking encrypted messaging services like WhatsApp to allow them access to their users' data.")

One notable example that also addresses user trust is the Skype case in the USA. The Microsoft Corporation service users believed that Skype offered a full end-to-end encryption feature. In 2013, however, thanks to Edward Snowden's disclosures, it was revealed that the platform contained a backdoor to the system: this revelation led to a protest of Skype users and an eventual loss of credibility of the application. In a response regarding the US government's position in seeking an encryption backdoor, Senator Wyden said that, "the US government does not need the approval of its secret surveillance court to ask a tech company to build an encryption backdoor"[45].

The author fears that the EU proposal risks replicating this dynamic by compelling providers to guarantee effective CSAM detection[46] and minimise false positives while simultaneously ensuring service security[47]: a combination that, in practice, pressures them to implement *de facto* backdoors or equivalent access mechanisms.

Moreover, current scanning technologies further demonstrate the tension between detection and data protection.

Experts have noted that, at present, scanning for CSAM is only feasible through a type of virus scanner within a chat application, which detects and reports any CSAM[48]. Scanning technologies for detecting child

---

[45] Z. Whittaker, *US Says It Doesn't Need Secret Court's Approval to Ask for Encryption Backdoors* (2017), available at https://www.zdnet.com/article/us-says-it-does-not-need-courts-to-approve-encryption-backdoors/ (last visited Aprile 12, 2026)

[46] *Proposal for a Regulation*, supra note 14, recital 28.

[47] Regulation (EU) 2016/679, art. 32 ("Taking into account the state of the art, the costs of implementation, and the nature, scope, context, and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia: (1) pseudonymization and encryption of personal data; (2) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services; (3) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident").

[48] See Hal Abelson et al., *Bugs in Our Pockets: The Risks of Client-Side Scanning* (2021),

sexual abuse material (CSAM) rely on an integrated set of perceptual hashing,[49] machine-learning classifiers[50], and behavioral analysis to identify both known illegal content and high-risk interactions.

To be more specific in addressing the preoccupations, we must highlight how recent debates have centered on client-side scanning (CSS), a technique enabling the on-device analysis of content before encryption. In practice, CSS is embedded directly into messaging applications so that every image a user intends to send is matched on the device itself against a database of CSAM fingerprints (hashes). The analysis occurs prior to any data being uploaded to an encrypted platform. If a threshold number of images corresponds to hashes contained in the CSAM database, the system automatically informs the competent authorities; otherwise, no information leaves the user's device. CSS, therefore, functions as a built-in scanning mechanism comparing outgoing content with CSAM hash databases and reporting suspected matches once predefined criteria are met[51].

---

available at https://www.researchgate.net/publication/355233857_Bugs_in_our_Pockets_The_Risks_of_Client-Side_Scanning (last visited 12 April, 2026)

[49] See Leon Twenning, Harald Baier and Thomas Göbel, *Using Perceptual Hashing for Targeted Content Scanning*, par. 3.1 ("Perceptual hashing is an approximate matching technique for comparing the similarity of objects. In this work, the focus is on the perceptual similarity of pictures (images). Perceptual similarity means that a human presented with two images would judge them as being the same or similar. Because images are often modified slightly during their use (e.g., compressed to reduce bandwidth and storage before being uploaded to a digital service), they cannot be identified by cryptographic hashes or other exact matching techniques. In such scenarios, perceptual hashing can be used effectively for approximate matching") (Bundeswehr University, Munich, Germany, available at https://www.researchgate.net/publication/374838370_Using_Perceptual_Hashing_for_Targeted_Content_Scanning (last visited March 18, 2026).

[50] M. Pereira, R. Dodhia, H. Anderson and R. Brown, *Metadata-Based Detection of Child Sexual Abuse Material*, 21 IEEE Transactions on Dependable and Secure Computing 3153, 3153–64 (2024).

[51] Abelson et al., *Bugs in Our Pockets* (cited in note 48); Internet Society, *Fact Sheet: Client-Side Scanning* (2020), available at https://www.internetsociety.org/resources/doc/2020/fact-sheet-client-side-scanning/ (last visited April 9, 2026).

In this regard, it could be argued that CSS allows authorities to investigate serious crimes without the need for centralized backdoors and therefore avoids weakening end-to-end encryption. However, despite this claim, the technology still enables continuous surveillance of user devices and exposes personal data to third parties, including law-enforcement bodies, at the precise moment when content is analysed on the device. In bypassing encryption at this pre-transmission stage, CSS compromises user privacy and introduces a persistent monitoring layer into everyday communications. Effective implementation would also require the technology to be mandated and installed by default on all new devices and software updates, thereby normalising pervasive, system-level scanning.

Challenging CSS on fundamental-rights grounds proves difficult because the sensitive nature of CSAM prevents users from accessing or reviewing the material that triggered the alert. This structural opacity means that individuals cannot verify whether a detection was accurate or whether the technology's scope has silently expanded beyond CSAM detection. Such limitations severely undermine transparency, accountability, and avenues for redress. The lack of user oversight allows authorities to justify actions on potentially false or unreviewable grounds, leaving citizens unable to contest erroneous or abusive determinations[52]. Ultimately, CSS introduces an intrusive, surveillance-enabling architecture whose operation is shielded from meaningful public scrutiny, raising profound concerns regarding proportionality, legality, and democratic oversight.

Technical accuracy remains another critical limitation. Hash-based systems such as PhotoDNA[53] enable highly accurate matching of previ-

[52] Joined Cases C-293/12 and C-594/12, *Digital Rights Ireland Ltd v Minister for Communications, Marine & Natural Resources*, ECLI:EU:C:2014:238 (invalidating the Data Retention Directive because it entailed broad, indiscriminate interference without safeguards).

[53] Testimony of Hany Farid, PhotoDNA developer, before the House Committee on Energy and Commerce, Fostering a Healthier Internet to Protect Consumers (October

ously identified CSAM, while newer AI models assess imagery and language patterns to detect novel material or grooming behaviors. CSAM detection systems have, in general, proven effective, while grooming detection still faces challenges. In this regard, even detection systems claiming 90% accuracy would generate millions of false positives when applied to the enormous volume of daily communications within the EU. Such an error rate is particularly problematic given the gravity of the allegations involved and the legal consequences that may follow. False positives may result from misinterpretation of benign family images, medical photographs, or ambiguous text, and evidence indicates that these errors disproportionately affect marginalised linguistic or cultural groups whose communication patterns differ from those represented in training datasets[54].

To be more precise, while image-based detection often has a 100% accuracy for known CSAM, grooming-detection models remain far less reliable, as they misclassify approximately one in ten flagged interactions[55].

16, 2019) ("There are claims that PhotoDNA has a false positive rate of 1 in 50 billion; but it is proprietary, thus these claims cannot be independently verified"), Available at: https://www.congress.gov/116/meeting/house/110075/witnesses/HHRG-116-IF16-Wstate-FaridH-20191016.pdf (last visit December 18, 2025).

[54] See Desara Dushi, *Does the End Justify the Means? The European Commission's Proposed Regulation Laying Down Rules to Prevent and Combat Child Sexual Abuse*, 32 *International Journal of Law and Information Technology* (2024), available at https://researchportal.vub.be/en/publications/does-the-end-justify-the-means-the-european-commissions-proposed-/ (last visited March 18, 2026).

[55] Hany Farid, PhotoDNA Developer, Testimony before the House Committee on Energy and Commerce, *Fostering a Healthier Internet to Protect Consumers* (October 16, 2019), available at https://www.congress.gov/116/meeting/house/110075/witnesses/HHRG-116-IF16-Wstate-FaridH-20191016.pdf (last visited April 1, 2026).

This argument is further aggravated by the fact that these errors trigger human review[56]. While this human-review requirement may seem a welcome safeguard, it also raises serious challenges.

Firstly, because the material in question is extremely sensitive and typically cannot be disclosed to users, the process remains largely opaque: users would rarely, if ever, have the ability to examine the content attributed to them, contest wrongful flagging, or to verify whether human reviewers acted correctly.

Secondly, given the scale of communications and the potentially vast volume of flagged items, the human workload could become enormous, raising concerns regarding practical feasibility, delays, or superficial review.

Finally, even well-trained reviewers may still diverge in their judgments, especially when evaluating borderline cases such as reclaimed slurs, sarcastic remarks, culturally specific communication styles, or politically ambiguous statements. These variations, often shaped by linguistic and cultural differences, risk embedding systemic biases into human moderation processes[57].

---

[56] Proposal Regulation 2022/0155 (COD), recital 28 ("With a view to constantly assess the performance of the detection technologies and ensure that they are sufficiently reliable, as well as to identify false positives and avoid, to the extent possible, erroneous reporting to the EU Centre, providers should ensure human oversight and, where necessary, human intervention, adapted to the type of detection technologies and the type of online child sexual abuse at issue. Such oversight should include regular assessment of the rates of false negatives and positives generated by the technologies, based on an analysis of anonymised representative data samples. In particular where the detection of the solicitation of children in publicly available interpersonal communications is concerned, service providers should ensure regular, specific and detailed human oversight and human verification of conversations identified by the technologies as involving potential solicitation of children."). See also Proposal Regulation 2022/0155, art. 10(4)(c) (cited in note 14).

[57] See Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi and Noah A. Smith, "The Risk of Racial Bias in Hate Speech Detection," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 1668, 1668–78 (Association for Computational Linguistics 2019); see also Vinay Koshy, Frederick Choi, Yi-Shyuan

Despite these concerns, supporters of the Regulation maintain that scanning is indispensable to counter the increasing misuse of encrypted and anonymized services, such as VPNs and proxy servers, by perpetrators seeking to evade detection[58].

The debate, therefore, highlights a fundamental dilemma: while technological tools may assist in combating child sexual abuse, their current form introduces severe risks of systemic surveillance, mass data collection, inaccurate reporting, and erosion of digital security, all of which challenge the proportionality and legality of the proposed framework.

3.2. *Safekeeping of Data and Cybersecurity Issues*

Safekeeping the vast quantities of data generated through CSAM detection orders raises an additional set of profound legal, technical, and cybersecurity concerns, particularly given that the Regulation remains insufficiently clear, precise, and complete in defining the risks associated with such intrusive data processing. This ambiguity is itself problematic: under the EU Charter of fundamental rights[59], any interference with the rights to privacy and data protection must be provided for by law, meaning that individuals must be able to foresee with reasonable certainty under which circumstances their data may be accessed, processed, retained, or disclosed.

Chiang, Hari Sundaram, Eshwar Chandrasekharan and Karrie Karahalios, "Venire: A Machine Learning-Guided Panel Review System for Community Content Moderation," 9 *Proceedings of the ACM on Human-Computer Interaction* 1, 1–35 (2024).

[58] Europol, *European Union Serious and Organised Crime Threat Assessment* (SOCTA 2021, no. 6), available at https://www.europol.europa.eu/cms/sites/default/files/documents/socta2021_1.pdf (last visited April 12, 2026)

[59] Charter of Fundamental Rights of the European Union, art. 52(1) ("Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others").

In its current form, however, the Regulation leaves significant ambiguity regarding the nature of the data to be stored, the security guarantees expected from providers, and the safeguards to prevent misuse or secondary processing. The obligation to screen all interpersonal communications through automated systems, irrespective of whether any wrongdoing is suspected, effectively amounts to systematic access to vast amounts of personal information. This constitutes an interference with Articles 7 and 8 of the EU Charter[60] regardless of how the resulting data is subsequently used[61], and whether weakening or circumventing end-to-end encryption (E2EE) may be required for detection further amplifies the risks. Weakening E2EE not only facilitates state-level access but also exposes communication infrastructures to exploitation by malicious actors, thereby introducing a real possibility of disproportionate surveillance practices[62].

These tensions become even more complex within the EU's multilevel legal framework, where criminal-procedural rules may vary significantly between Member States but are nonetheless bound by the minimum standards of fairness set out at the European level[63]. Data obtained through detection orders could be used in criminal proceedings, raising difficult questions regarding the reliability of CSAM-related evidence, the

[60] Charter of Fundamental Rights of the European Union, arts. 7, 8 ("Everyone has the right to respect for his or her private and family life, home and communications"; "Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. Compliance with these rules shall be subject to control by an independent authority").

[61] Council of the European Union, Legal Service, *Opinion of the Legal Service on the Proposal for a Regulation Laying Down Rules to Prevent and Combat Child Sexual Abuse*, Doc. 8787/23 (April 26, 2023).

[62] *Open Letter from Security and Privacy Researchers in Relation to the Online Safety Bill* (2023), available at https://haddadi.github.io/UKOSBOpenletter.pdf (last visited April 12, 2026)

[63] Roberto Kostoris, *Handbook of European Criminal Procedure* at 67–98 (Springer 2018).

chain of custody, and the ability of defendants to meaningfully challenge the material relied upon against them. Despite the invasive nature of such surveillance, EU principles on the exclusion of unlawfully obtained evidence[64] are not consistently robust enough to prevent courts from admitting detection-order data collected in violation of privacy and data-protection norms. The absence of clear EU-wide guidelines on disclosure obligations[65], particularly regarding whether and how a defendant may access or contest the data that triggered the detection, further undermines procedural guarantees and risks creating fragmented standards among Member States. This lack of foreseeability threatens the rule of law requirement for strict legal safeguards and effective judicial oversight when it comes to interferences with fundamental rights.

Beyond legal uncertainty, the technological tools used for CSAM detection introduce their own vulnerabilities.

Research into perceptual-hashing and client-side scanning tools shows that these systems face serious technical weaknesses, including evasion, hash-collision, and data-leakage attacks[66]. Skilled users can edit images or their metadata in ways that bypass detection altogether, undermining the system's ability to identify the very offenders it is meant to capture. Hash-collision attacks make it possible to generate harmless images that share the same hash as known CSAM, allowing attackers to overwhelm detection systems with false alerts. Such overloads can effectively create denial-of-service situations, drain enforcement resources, and diminish the overall reliability of CSAM-monitoring infrastructures. Even more troubling is the risk that fabricated collisions could be used to

[64] Charter of Fundamental Rights of the European Union, arts. 47 (right to an effective remedy and to a fair trial), 48 (presumption of innocence and right of defence), 52(1) (scope and interpretation; limits on Charter rights, proportionality and necessity).

[65] Isadora Neroni Rezende, *The Proposed Regulation to Fight Online Child Sexual Abuse: An Appraisal of Privacy, Data Protection and Criminal Justice Issues*, 38 *International Review of Law, Computers & Technology* 369, 369–90 (2024).

[66] Leon Twenning, Harald Baier and Thomas Göbel, *Using Perceptual Hashing for Targeted Content Scanning*, in *IFIP Advances in Information and Communication Technology*, vol. 687 (Springer 2023).

frame individuals by producing false "matches" that falsely signal possession or distribution of CSAM, potentially subjecting innocent users to invasive scrutiny or criminal investigation.

Data-leakage attacks present another layer of harm: by reversing perceptual hashes, attackers may reconstruct blurred or partial versions of the original images[67]. This threatens not only the privacy of the children depicted but also the privacy of ordinary users whose benign content is scanned and hashed. The presence of these vulnerabilities shows that large-scale CSAM-detection systems inherently generate and concentrate highly sensitive data, which must then be protected against compromise, expanding the attack surface available to malicious actors.

Compounding these risks is the lack of transparency surrounding many of the technologies, policies, and operational practices behind such systems. Companies and law enforcement bodies might withhold technical information out of concern that disclosure might help offenders adapt. While this rationale is understandable, it also limits public debate, restricts independent research, and undermines users' trust in both service providers and public authorities. Without meaningful transparency, oversight, and external auditing, it becomes difficult to determine whether these systems operate lawfully, whether they introduce disproportionate risks, or whether they can realistically balance child protection with cybersecurity and fundamental rights guarantees.

Ultimately, a Perceptual Hashing and Client-Side Scanning (PHTCS) framework capable of supporting effective enforcement while respecting democratic principles can only be realised through mutual trust, rigorous research, strong procedural safeguards, and a legal architecture that sets out clear, predictable, and rights-compatible standards for data security, access, and accountability.

[67] *Ibid.*

4. *Conclusion*

In conclusion, while the protection of minors online constitutes a fundamental obligation, the tools adopted to achieve this objective must not compromise the essential rights to privacy, data protection, and secure communications. Given the limitations of current scanning technologies and the legislative uncertainties surrounding their deployment, any future framework should establish clear operational boundaries, meaningful human oversight, robust safeguards for data security, and strict limits on data retention[68].

At the same time, the large-scale deployment of CSAM detection infrastructures raises broader structural concerns. Systems initially designed to detect child sexual abuse material could, with limited technical modification, be repurposed for wider monitoring purposes[69]. The particularly sensitive nature of child protection risks facilitating the gradual normalisation of intrusive surveillance practices, potentially undermining the guarantees of confidentiality, integrity, and security that technologies such as end-to-end encryption are designed to protect[70].

---

[68] *Ekimdzhiev and Others v. Bulgaria*, ECHR 70078/12 (2022) (holding that Bulgarian legislation on secret surveillance and communications-data retention violated Article 8 ECHR because it failed to meet the "quality-of-law" standard, lacking clarity, precision, adequate safeguards, and strict necessity requirements, and emphasizing that any system allowing state access to communications data must clearly define its scope, the categories of persons who may be targeted, the conditions and duration of retention, and independent oversight to prevent arbitrary or mass surveillance). See also C-623/17, *Privacy International v Secretary of State for Foreign and Commonwealth Affairs* (2020) (holding that national laws requiring general and indiscriminate retention or transmission of traffic and location data to security or intelligence agencies for national-security purposes fall within the scope of Directive 2002/58/EC and are incompatible with arts. 7, 8, and 52(1) of the Charter of Fundamental Rights of the European Union, and confirming that metadata is subject to strong EU fundamental-rights protection and that bulk, untargeted data-collection regimes exceed the permissible limits of interference with privacy and data protection).

[69] *Id*. at 52

[70] Art. 21, para. 1, EU dir. 14 December 2022, no. 2555 ("Member States shall ensure that

The urgency of addressing online child sexual abuse is undeniable, as in February 2024, the UN Special Rapporteur on the sale and sexual exploitation of children warned of the rapidly escalating scale and sophistication of online abuse, calling for concrete, coordinated action from states worldwide[71].

These warnings, while they underscore the gravity of the problem, they also reinforce the need for solutions that truly enhance child protection without eroding the fundamental rights and digital security upon which individuals and democratic societies rely. In fact, effective child protection cannot be achieved through measures that introduce systemic cybersecurity vulnerabilities or enable disproportionate access to private communications.

A sustainable regulatory response must therefore protect children without eroding the fundamental rights and digital security on which democratic societies depend.

essential and important entities take appropriate and proportionate technical, operational and organisational measures to manage the risks posed to the security of network and information systems which those entities use for their operations or for the provision of their services, and to prevent or minimise the impact of incidents on recipients of their services and on other services."). See also art. 21, para. 2, lett. h, EU dir. 14 December 2022, no. 2555 (explicitly mentioning encryption or cryptographic protections as a means to secure sensitive information).

[71] *OHCHR, UN Expert Alarmed by New Emerging Exploitative Practices of Online Child Sexual Abuse* (2024), available at https://www.ohchr.org/sites/default/files/documents/issues/children/sr/cfis/existing-emerging/subm-existing-emerging-sexually-cso-suojellaan-lapsia-ry.pdf (last visited April 12, 2026).