Council of the
European Union

Brussels, 19 December 2023
(OR. en)

17002/23

JAI 1704
ENFOPOL 554
CRIMORG 207
IXIM 247
DATAPROTECT 384
CYBER 326
COPEN 461
FREMP 382
TELECOM 403
COMPET 1303
MI 1154
CONSOM 504
DIGIT 303

COVER NOTE

From:	Secretary-General of the European Commission, signed by Ms Martine DEPREZ, Director
date of receipt:	19 December 2023
To:	Ms Thérèse BLANCHET, Secretary-General of the Council of the European Union
No. Cion doc.:	COM(2023) 797 final
Subject:	REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the implementation of 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

Delegations will find attached document COM(2023) 797 final.

Encl.: COM(2023) 797 final

EUROPEAN
COMMISSION

Brussels, 19.12.2023
COM(2023) 797 final

REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL

on the implementation of 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

CONTENTS

LIST OF TERMS AND ACRONYMS

Term/Acronym	Definition
AI	Artificial Intelligence
API	Application Programming Interfaces
CG-CSAM	Computer-generated Child Sexual Abuse Material
ChatGPT	ChatGPT (Chat Generative Pre-trained Transformer) is a form or generative AI. It is a large language model-based chatbot, developed by OpenAI, that enables users to refine and steer a conversation towards a desired length, format, style, level of detail, and language.
Classifiers	A form of artificial intelligence, an algorithm that sorts data into labelled classes or categories
Content Safety API classifier	Google's Content Safety API classifier uses programmatic access and artificial intelligence to help classify and prioritise billions of images for review
CSA	Child Sexual Abuse
CSAI Match	CSAI Match is a technology developed by YouTube engineers to identify re-uploads of previously identified child sexual abuse in videos
CSAM	Child Sexual Abuse Material, e.g. images and videos depicting CSA
CSA Directive	Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA, OJ L 335, 17.12.2011, p. 1–14
CSEA	Child Sexual Exploitation and Abuse
CSA online	The common term used for the three types of child sexual abuse, defined in the CSA Directive, namely: child pornography, pornographic performance and solicitation of children ('grooming'), as defined in the Interim Regulation (Article 2(4))
EU	European Union
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016, p. 1–88
Grooming	Offenders building trust and a relationship with a child in an effort to gain access to the minor for sexual exploitation or abuse. Formally known as solicitation of children, as defined in Article 6 of the CSA Directive

Hash	A unique digital code created by a mathematical algorithm ("hashing") that becomes this file's signature, or its hash value
Interim Regulation	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, OJ L 274, 30.7.2021, p. 41–51
LLM	A large language model (LLM) is a type of artificial intelligence model that has been trained through deep learning algorithms to recognise, generate, translate, and/or summarise vast quantities of written human language and textual data.
MD5	Cryptographic message authentication code algorithm for use on the internet
Meta SSN++	An AI model, developed by Meta, that can detect near-exact duplicates.
NCMEC	National Centre for Missing and Exploited Children (US private, non-profit organisation) to which online service providers are required to report under US law instances of potential child sexual abuse that they find in their networks
PDQ and TMK+PDQF	Tools used by Facebook to detect harmful content. PDQ is a photo matching technology; TMK+PDQF is a video-matching technology.
PhotoDNA	The most widely used tool based on hashing technology, available free of charge, based on a licensing agreement tailored to avoid abuse and use for any other purpose than the detection of CSA

1. INTRODUCTION

Article 9 of the Interim Regulation (hereafter also: "the Regulation") obliges the Commission to prepare a report on its implementation, based on reports submitted by providers of interpersonal communications services (hereafter: "providers") on the processing of personal data and on the statistics provided by Member States. According to the aforementioned provision, in the implementation report, the Commission shall consider, in particular,

(a) the relevant conditions for the processing of relevant personal data and other data processed under the Regulation;

(b) the proportionality of the derogation provided for by the Regulation, including an assessment of the statistics submitted by the Member States pursuant to its Article 8;

(c) developments in technological progress regarding the activities covered by the Regulation and the extent to which such developments improve accuracy and reduce the numbers and ratios of errors (false positives).

This implementation report under the Interim Regulation builds on the data obtained through reporting by providers and Member States pursuant to its Article 3(1)(g)(vii) and Article 8, respectively. Such reporting brought to light significant disparities in the availability of data, the types of data collected, and thus also the comparability of the data collected by providers and Member States. As the Regulation does not provide a template for the reporting, providers shared different types of information which were not necessarily comparable; the Commission services therefore engaged in follow-up to ensure the correct interpretation of the data. The majority of Member States were unable to provide data in time and a number of them have been unable to provide any data until the publication of this report. This had a significant impact on the timing, completeness and usefulness of the report. Despite the efforts to ensure coherence of the data and comparability, significant disparities remain which are reflected in the below tables, which also do not contain data for all providers or Member States on all points.

This implementation report seeks to give a factual overview of the state of play in connection to the implementation of the Interim Regulation, based on the available data. The report does not contain any interpretations of the Regulation and does not take any position on the manner in which it has been interpreted and applied in practice.

2. IMPLEMENTATION MEASURES

2.1. Processing of personal data by providers (Article 3 (g) (vii))

Article 3(g)(vii) of the Interim Regulation lays down the conditions for providers acting under the derogation contained therein to publish and submit to the competent supervisory authority and to the Commission, by 3 February 2022, and by 31 January every year thereafter, a report on the processing of personal data under this Regulation. Google, LinkedIn, Meta, Microsoft, and X (former Twitter)[1] submitted reports for 2021 and 2022.

2.1.1. Type and volumes of data processed

Providers reported processing both content and traffic data.

As regards content data processed to detect online child sexual abuse, all aforementioned providers mentioned images and videos. Mostly they relied on the hash matching technologies PhotoDNA and MD5 to detect matches of previously identified child sexual abuse material (hereafter: "CSAM"). Google's CSAI Match tool was used to create digital fingerprints of videos on platforms and compare them with the files in Google/YouTube's fingerprint repository (LinkedIn). The use of automated technology (artificial intelligence machine learning) and human review was equally reported (e.g. by Google). Google and LinkedIn confirmed identifying also CSAM that did not match previously identified CSAM. None of the five providers that submitted data reported data on detecting solicitation of children via text detection under the scope of the derogation provided by this Regulation.

As for traffic data collected and the respective volumes of the different types of content and traffic data processed, the reports of providers varied substantially.

Traffic data collected by providers and included in CyberTipline reports to the U.S. National Center for Missing and Exploited Children (hereafter: "NCMEC") include (all or a selection) of the following data:

a) User/reportee/account related data (Google, LinkedIn, Microsoft, X);
b) Metadata related to content/transactional data (Google, LinkedIn, Microsoft);
c) Data related to a potential victim (Google);
d) Abuse operations data (Google).

In terms of volumes of data processed under the Interim Regulation, LinkedIn reported processing 8 million images and videos originating from the EU between 14 July and 31 December 2021, and 21 million images and 63 000 videos originating from the EU in 2022. Microsoft reported processing for the purposes of the Regulation 8.9 billion images and videos globally between July and December 2021, and 12.3 billion content items globally in 2022; EU figures were not available, hence it is not possible to draw any conclusions for the purposes of

[1] Twitter submitted their contribution before its renaming; it will be referred to as "X" throughout the rest of this report.

this report. The other providers did not provide information on volumes of data processed. Of the five reporting providers, therefore, only one provided data at the requisite level of granularity.

To illustrate the overall context, NCMEC reported having received a total of 87.2 million images and videos globally and 5.1 million images and videos concerning the EU in 2022, and 84.8 million images and videos globally and 1.8 million images and videos concerning the EU in 2021. These are only the materials that have been identified as potential CSAM by a provider and therefore cannot be taken as indicative of the overall volumes of data processed under the Interim Regulation.

2.1.2. Grounds for processing pursuant to Regulation (EU) 2016/679

The providers reported relying on the following specific grounds pursuant to Regulation (EU) 2016/679 (hereafter: "GDPR"):

- Article 6(1)(d) of the GDPR, i.e. processing necessary in order to protect the vital interests of children and those who are the victims of online child sexual abuse (Google, Meta, X[2]);
- Article 6(1)(e) of the GDPR, i.e. processing necessary for the performance of tasks carried out in the public interest (LinkedIn, Microsoft, Meta, X[3]);
- Article 6(1)(f) of the GDPR, i.e. processing necessary for the purposes of the legitimate interests of:
    i. the provider to detect, prevent or otherwise address online child sexual abuse on their services and to protect other users, customers, partners, and the public from this form of illegal content (Google, Meta);
    ii. victims of child sexual abuse and the organisation to whom the provider reports online child sexual abuse (e.g. NCMEC) to detect, prevent and remove online child sexual abuse from their services (Google).

2.1.3. Ground for transfers of personal data outside the Union pursuant to Chapter V of the GDPR, where applicable

All providers reported relying on standard contractual clauses approved by the Commission pursuant to Article 46(2)(c) of the GDPR. For transfers of personal data to NCMEC, LinkedIn also reported relying on a derogation, to the extent applicable, permitted under Article 49(1) GDPR.

[2] Meta and X did not specify the concrete Article explicitly.
[3] Meta and X did not specify the concrete Article explicitly.

2.1.4. Number of cases of online child sexual abuse identified, differentiating between CSAM and solicitation of children

Table n.1: Number of cases of online child sexual abuse identified in 2021

Provider	Number of cases of CSAM identified in 2021	Comments
Google	33 Google Chat accounts	This refers to the number of Google Chat accounts for EU users where online child sexual abuse was identified through automated technologies in the period from 2 August 2021 to 31 December 2021. No data was provided on the number of content items identified.
LinkedIn	31 content items	31 content items were confirmed via manual review as CSAM and reported to NCMEC; 6 were known CSAM and the remaining 25 were unknown CSAM.
Meta	340 000 accounts	Number of accounts detected as sending at least one piece of media constituting CSAM, in message threads which included an EU user between 8 November 2021 to 31 December 2021.
Microsoft	6 600 content items	6 600 content items (single image or video) confirmed as CSAM detected from the European Union out of over 20 000 content items identified globally between July 2021 to December 2021.
X (former Twitter)	532 898 accounts	Accounts (unclear if only from the EU or globally) suspended for violating X's child sexual exploitation policy between 2 August 2021 to 31 December 2021.

In the case of X, it is unclear from the data provided whether it relates exclusively to services falling into the scope of the Interim Regulation (number-independent interpersonal communications services) or whether that number also comprises other services (such as information society services). This concerns all the data related to X in this report.

Table n.2: Number of cases of online child sexual abuse identified in 2022

Provider	Number of cases of CSAM identified in 2022	Comments
Google	2 045 content items	Content items identified and reported to NCMEC in 752 Google accounts through automated technologies for EU users.
LinkedIn	2 content items	LinkedIn detected 2 images and 0 videos confirmed as CSAM.
Meta	6.6 million content items	Pieces of media constituting CSAM actioned that were detected using Meta's media matching technology, in message threads which included an EU user.
Microsoft	12 800 content items	12 800 content items (single image or video) confirmed as CSAM detected from the EU out of over 50 000 content items identified globally in 2022.

X (former Twitter)	2 348 712 accounts	Accounts (unclear if only from the EU or globally) suspended for violating X's child sexual exploitation policy.

2.1.5. User redress and outcome

In accordance with Article 3(1)(g)(iv) of the Interim Regulation, providers are to establish appropriate procedures and redress mechanisms to ensure that users can lodge complaints with them. In addition, its Article 5 contains rules on judicial redress.

Providers reported implementing such internal redress procedures and mechanisms for users whose accounts have been restricted for sharing CSAM and/or content removed as CSAM so that they can appeal their restriction/removal and have their case reviewed for errors.

They reported cases in which a user has lodged a complaint with the internal redress mechanism or with a judicial authority regarding matters in scope of the Regulation within the EU, and the outcomes of such complaints. No provider except Microsoft (which reported 0 complaints in both channels in 2021 and 2022) reported separate statistics on internal redress and judicial redress; as a result the tables below cover both internal and judicial redress procedures.

Table n.3: Number of cases in which a user has lodged a complaint with the internal redress mechanism or with a judicial authority and the outcome of such complaints in 2021

Provider	Number of cases of user complaints	Number of reinstated accounts	Number of reinstated contents	Comments
Google	8	0	n.a.	Google Chat accounts disabled for online child sexual abuse where the user appealed: 8. None was reinstated.
LinkedIn	0	n.a.	n.a.	
Meta	4 900	n.a.	207	4 900 users appealed. Following the appeals process, 207 users had their content restored, and account actions reversed.
Microsoft	0	n.a.	n.a.	
X (former Twitter)	ca 90 000	ca 3 000	n.a.	Ca 90 000 appeals. X reinstated ca 3 000 of those accounts.

Table n.4: Number of cases in which a user has lodged a complaint with the internal redress mechanism or with a judicial authority and the outcome of such complaints in 2022.

Provider	Number of cases of user complaints	Number of reinstated accounts	Number of reinstated contents	Comments
Google	378	0	n.a.	Google accounts disabled for online child sexual abuse where the user appealed: 378. None were reinstated.
LinkedIn	0	n.a.	n.a.	
Meta	29 000	n.a.	3.700	Users appealed the actions of around 29 000 pieces of their shared media. Following the appeals process, around 3 700 pieces of content were restored, and account actions reversed.
Microsoft	0	n.a.	n.a.	
X (former Twitter)	Ca 430 000	Ca 4 000	n.a.	Ca 430 000 appeals. X reinstated ca 4 000 of those accounts.

2.1.6. Number and ratios of errors (false positives) of the different technologies used

In accordance with Article 3(1)(e) of the Interim Regulation, providers are to ensure that the technologies used are sufficiently reliable in that they limit to the maximum extent possible the rate of errors regarding the detection of content representing online child sexual abuse.

In this respect, providers reported that they do not apply each technology to detect online CSA in isolation. Rather, they implement a layered approach to detection of online CSA by combining different detection technologies to increase accuracy. To reduce errors or false positives, all providers complement these with human review. Providers did not provide the number and ratios of errors (false positives) for each of the different technologies used separately, but rather reported aggregate data for all technologies used.

Most providers measure the number and ratios of errors as enforcement decision reversals, i.e. as the overall appeal reinstate/reversal rate (e.g. the rate at which the provider reinstated disabled accounts or content upon the user appeal). The approach taken by the providers does not necessarily reflect the definition of ‘false positives’ in statistics.

The appeal reinstate/reversal rates reported are as follows:

Table n.5: Appeal reinstate/reversal rates

	2021*		2022		
Provider	**% of reinstated accounts vs number of appeals**	**% of reinstated vs suspended accounts**	**% of reinstated accounts vs number of appeals**	**% of reinstated vs suspended accounts**	**Comments**
Google	0 % (0 vs 8)	0 % (0 vs 33)	0 % (0 vs 378)	0 % (see comments)	The number of suspended accounts was not provided for 2022. Instead, it was provided the number of pieces of content identified and reported to NCMEC, 2045. No account was reinstated following the appeal.
LinkedIn	0 %	0 %	0 %	0 %	No appeals. For the period from 13 July 2021 to 31 December 2021, LinkedIn also reported that of the 75 files reviewed as potential CSAM originating from the EU, 31 were confirmed as CSAM by human review. LinkedIn did not provide such data for 2022.
Meta	4.22 % (207 vs 4.9k)	0.06 % (207 vs 340k)	See comments	See comments	No information was provided to determine precisely the scope, the content of the appeals and the reasons to reinstate. For 2022, the data provided was in terms of pieces of content, not in terms of accounts: - Number of pieces of content suspended ("actioned"): 6.6 million - Number of pieces of content appealed: 29k - Number of pieces of content reinstated: 3.7k Therefore: - % of reinstated pieces of content vs number of appeals: 12.8% (3.7k vs 29k) - % of reinstated pieces of content vs number of pieces of content suspended: 0.06% (3.7k vs 6.6 million)

Microsoft	0 %	-	-	-	Insufficient data to enable the calculation of the appeal reinstate/reversal rate. For 2022 17 total reversals of initial content moderation decision, no figures on total appeals provided.
X (former Twitter)	1.43% (100 vs 7k)	0.06 % (100 vs 166k)	2.17% (500 vs 23k)	0,10% (500 vs 501k)	For the second half of 2021 ca 166k users suspended for CSA via automated mechanisms. Of these users, ca 7k submitted appeals which resulted in ca 100 overturns. In 2022, 501k users suspended for CSA via automated mechanisms. Of these users, ca 23k submitted appeals which resulted in ca 500 overturns.

* The reporting periods in 2021 vary for each provider

2.1.7. Measures applied to limit the error rate and the error rate achieved

According to Article 3(1)(e) of the Interim Regulation, the technologies used must be sufficiently reliable and the consequences of any occasional errors must be rectified without delay. In addition, Article 3(1)(g)(ii) requires human oversight and, where necessary, human intervention.

All providers reported using a layered approach to detect and combat the spread of online CSA. This includes the use of hash-matching technologies (including PhotoDNA) to detect CSAM in combination with human review processes to confirm whether a media file (image and video) contains CSAM, as well as human oversight over the CSAM processing.

Providers reported applying different measures and safeguards to limit and reduce the error rate in their detection, reporting and removal of online CSA. These include (non-exhaustive list)[4]:

i. monitoring and quality assessment of the performance of CSA detection tools, both to fine tune precision (that they are detecting only online child sexual abuse) and recall (that they are not missing online child sexual abuse on their platforms) (Google, X);
ii. human review and oversight: samples of media detected as CSAM by hash-matching technologies are audited by human reviewers/trained analysts (Google, LinkedIn, Meta, Microsoft);
iii. flagging and review of high-volume clusters (Meta);
iv. deployment of further manual review processes as ongoing hash quality checks (LinkedIn, Microsoft);
v. human reviewers undergoing specialised robust trainings under guidance of counsel on how to recognise CSAM content to ensure accuracy of human review (Google);
vi. periodic quality control assessments of human reviewers and the verdicts that are applied (Google, X);
vii. other quality control processes to reduce errors and immediate remedy, such as independent hash verification (Google, LinkedIn), human review of each instance of never-before-seen CSAM prior to reporting (Google);
viii. development and regular review of policies and enforcement strategies by trained subject matter experts on online CSA (Google);
ix. engagement with NCMEC CyberTipline reports quality, and false positives, if any. (Google, LinkedIn, Meta, Microsoft, X).

2.1.8. The retention policy and the data protection safeguards applied pursuant to the GDPR

Article 3(1)(h) and (i) of the Interim Regulation require relevant personal data to be stored in a secure manner only for certain specified purposes and contain specifications regarding the storage period, respectively. In addition, the applicable requirements of the GDPR must be respected.

[4] The providers indicated in brackets are those that specifically reported the concrete measures. If some providers are not listed, it does not mean that they do not implement this measure, but only that they have not mentioned it in their report.

Providers reported having robust retention policies and personal data protection safeguards in place. Data retention policies vary depending on the type of data. They indicate that in each case the retention period is limited in time as deemed appropriate for the type of data and the purpose of processing and the data is deleted at the end of the retention period. Providers have more detailed information on data retention practices defined in their Privacy Policy/Statements and Service/User Agreements.

Most providers (LinkedIn, Meta, Microsoft) apply a retention policy of 90 days for media confirmed to contain CSAM detected within the scope of the Regulation. During this period, the content confirmed as CSAM is stored in separate and secure CSAM storage managed by specialised teams (e.g. Microsoft's Law Enforcement and National Security team). These storage systems automatically delete the stored CSAM content after 90 days unless the storage period has been extended upon receipt of lawful process requests generally related to law enforcement agencies following up on NCMEC reports.

Google reported that CSAM detected within the scope of the Regulation is stored no longer than strictly necessary for the relevant purposes under the Regulation and, in any event, no longer than 12 months from when the CSAM is identified and reported, with a possible extension based on a valid legal preservation request.

X (former Twitter) reported keeping profile information and content for the duration of the user account and personal data collected when users use their service for a maximum of 18 months. When an account is deactivated by the user, X generally keeps the data for an additional 30 days, then the account will go for deletion. User data related to complaints and policy violations, including the account information of violators (e.g. identifiers used to create the account: e-mail address or phone number), is retained indefinitely to prevent repeat policy offenders from creating new accounts and ensure that violators of X's policies cannot simply wait for the deletion period and then violate policies again.[5]

Personal data protection safeguards implemented by providers include industry standard measures (all or selection of), such as (non-exhaustive list)[6]:

i. Use of de-identification or pseudonymisation techniques and anonymisation of data (e.g. masking, hashing, differential privacy) (Goggle, LinkedIn, Meta, Microsoft);
ii. Provision of only hash values to third parties for the purpose of CSAM detection (Google, LinkedIn);
iii. Use of industry standard encryption (algorithms and protocols) for data in transit between privately owned infrastructure and public networks (Meta);
iv. Implementation of data governance strategies/comprehensive privacy programmes (X, former Twitter, Google) and of strict internal data access restrictions (Meta) (e.g. applied to staff, contractors, agents who need the information in order to process it), usage of

[5] X Privacy Policy, 4. How Long We Keep Information, Available at: https://twitter.com/en/privacy

[6] The providers indicated in brackets are those that specifically reported the concrete measures. If some providers are not listed, it does not mean that they do not implement this measure, but only that they have not mentioned it in their report.

Access Control Lists across CSAM review tools and hash bans (Meta), and strict contractual confidentiality obligations applied to those with access;

v. Review of anonymisation and data governance strategies, i.e. conducting privacy reviews to identify, access and mitigate potential privacy risks from the collection, processing, storing and sharing of personal data, review of protection practices (Microsoft);

vi. Maintaining security incident response plans for monitoring, detecting, and handling any possible security vulnerabilities and incidents across infrastructure (Google, Meta).

2.1.9. The names of the organisations acting in the public interest against child sexual abuse with which data has been shared pursuant to this Regulation.

In both reporting periods (July/August 2021 to December 2021 and January 2022 to December 2022), all providers reported sharing the data processed under this Regulation with NCMEC. All reporting providers also communicated to the Commission, in compliance with Article 7(1) of the Interim Regulation, that they reported online child sexual abuse under this Regulation to NCMEC.[7]

2.2. Member States' statistics (Article 8)

Member States are obliged to provide statistics pursuant to Article 8(1) of the Interim Regulation on the following:

(a) the total number of reports of detected online child sexual abuse that have been submitted by providers and organisations acting in the public interest against child sexual abuse to the competent national law enforcement authorities, differentiating, where such information is available, between the absolute number of cases and those cases reported several times and the type of provider on whose service the online child sexual abuse was detected;
(b) the number of children identified through actions pursuant to Article 3, differentiated by gender;
(c) the number of perpetrators convicted.

While most Member States provided at least partial information, the relevant data collection and reporting systems had not been set up in many of the Member States. As a result, where statistics were provided, they relate to very diverse reporting periods and differ significantly in terms of granularity. Some Member States submitted yearly statistics as of the date of entry into force of the Regulation. Most of them reported for calendar years as they might not have the technical means to distinguish the requested statistics per year as of the date of entry into force of the Regulation. A few Member States provided no data at all.

[7] The information on the organisations acting in the public interest to which providers report online child sexual abuse under this Regulation has been published at https://home-affairs.ec.europa.eu/policies/internal-security/child-sexual-abuse/legal-framework-protect-children_en, in line with the Commission's obligations under Article 8(2) of the Interim Regulation.

It should also be noted that in some cases the statistical data are obtained from so-called current databases, or journalisation and case management systems, i.e. not actual statistics systems. The numbers are at times provided based on dynamic data, which means the data are not final, i.e. they are subject to change. Changes occur, e.g., depending on the time of extraction (e.g. in Slovenia and Denmark) as more investigations and court cases are completed.

In several Member States, the competent authorities are creating new departments for investigating crimes related to child sexual abuse online, and creating central reporting for child sexual abuse online (Latvia, Czechia). This should help in having more accurate statistics in the future.

Germany stated that it could not provide any statistics according to Article 8(1) of the Interim Regulation, as it considered it had no legal basis for voluntary detection.[8] However, the German Federal Criminal Office (BKA) on its website reports receipt of 89 844 reports from NCMEC in 2022, and NCMEC reports sending 138 193 reports to the German authorities.[9] Three Member States did not provide any data or justification for not reporting pursuant to that provision (Malta, Portugal and Romania).

2.2.1. The total number of reports of detected online child sexual abuse

Most Member States provided some statistics on the total number of reports of online child sexual abuse pursuant to Article 8(1)(a) of the Interim Regulation. As Member States provided data for differing reporting periods, it was not possible to calculate the total number of reports of detected online child sexual abuse received at EU level for any given period such as the time of implementation of the Regulation.

Member States mostly provided the total number of reports received from providers or other organisations acting in the public interest against child sexual abuse to the national law enforcement authorities. Given that most US-based providers report to NCMEC, most Member States reported receiving most or all of their reports from NCMEC. Member States did not indicate the number of actionable reports, i.e. reports suitable for investigation, but some indicated the number of cases launched, which is significantly lower. The difference between reports received and investigated cases were attributed to several reasons, e.g. that the report contained CSAM, but it did not contain sufficient information to open an investigation; the merging of reports when more than one report applies to a certain suspect; or that the material, while showing exploitative situations, was not assessed as criminal under national law. In addition, Member State mostly did not differentiate the absolute number of cases and those reported several times. Where reports were provided by NCMEC, NCMEC already pre-

[8] Report submitted by Germany in line with article 8 of the Regulation (EU) 2021/1232, received on 18. October 2022. NCMEC publishes all data on reports received and related to the EU Member States, including Germany, in their CyberTipline Reports by Country. See: NCMEC, 2021 CyberTipline Reports by Country, accessed in July 2023; NCMEC, 2022 CyberTipline Reports by Country, accessed in July 2023.

[9] https://www.bka.de/SharedDocs/Kurzmeldungen/DE/Kurzmeldungen/230623_Mindestspeicherfristen_IP-Adressen.html

categorised reports it had received from providers as “actionable” or “informational”. NCMEC defines an actionable report as one containing sufficient information to launch an investigation. This typically includes user details, imagery, and a possible location. The report is categorised as “informational” when it contains insufficient information or where the imagery is considered viral and has been reported many times. NCMEC designated 49% of the reports as “actionable” in 2022, while 51% were designated as “informational”.

Only very few Member States indicated the type of providers on whose services the online child sexual abuse was detected (e.g. Belgium, Czechia, Estonia, France, and Poland) and only one Member State provided a detailed breakdown (Belgium).

Slovenia indicated that it could not provide figures only on offences investigated because of reports submitted by providers and organisations, but rather could only provide figures for all investigations into online child sexual abuse regardless of the source of the information that led to the launch of the investigation.

Table n.6: The total number of reports of detected online child sexual abuse as reported by Member States

Country	Reporting Period	Total number of reports of online child sexual abuse	Source of reports	Comments
Austria	2021 to 2022	16 311	NCMEC	
Belgium	1 August 2021 to 31 July 2022	26 226	Reports originating from providers (social media) and Childfocus hotline	
Bulgaria	2021 to 2022	9 120	Providers and INHOPE hotline through "Safenet" and other	Out of these, 9 112 alerts were about webpages containing CSAM, hosted by Bulgarian providers.
Croatia	1 January 2021 to 31 October 2022	9 044	NCMEC	
Cyprus	1 July 2021 to 31 December 2022	3 570	NCMEC	
Czechia	1 January 2022 to 31 July 2022	13 279	NCMEC	
Denmark	2 August 2021 to 20 January 2023	10 744	NCMEC	
Estonia	-	-	NCMEC, Child Helpline 116 111	Estonia reported that the statistics of Police and Border Guard, including NCMEC, are not public. For 2021 they reported 360 non-contact sexual crimes against a child. Moreover, 86 % of all non-contact sexual crimes were committed in the Internet environment or using information technology tools.

Finland	2022	25 000	NCMEC and Save the Children	
France	1 August 2021 to 1 August 2022	120 000	NCMEC	
Germany	-	-	-	Data not available/not reported.
Greece	2021 to 2022	142	NCMEC, Greek Hotline for illegal Internet content – Safeline, National Telecommunications and Postal Commission, National Line SOS 1056 – The smile of the child, Greek Ombudsman	
Hungary	2022	0	None of the reports sent by providers were sent under the Interim Regulation	
Ireland	2021 to 2022	15 355	NCMEC	
Italy	2022	4 607	Not specified	
Latvia	1 August 2022 to 6 March 2023	Approximately 115 to 220 reports monthly	From non-Latvian providers and organisations acting in the public interest against child sexual abuse (mostly NCMEC) and Latvian providers and organisations (mostly Latvian Safer Internet Center)	
Lithuania	1 January 2021 to 30 June 2022	4 142	Not specified	
Luxembourg	2021 to 2022	2 491	Not specified	
Malta	-	-	-	Data not submitted/reported.
Netherlands	2021	36 537	Providers and organisations acting in the public interest against child sexual abuse	

Poland	3 August 2021 to 3 August 2023	227	Providers and organisations acting in the public interest against child sexual abuse	For the period from 3 August 2022 to 3 August 2023 Poland noted 1 report of child grooming and 105 reports of CSAM.
Portugal	-	-	-	Data not submitted/reported.
Romania	-	-	-	Data not submitted/reported.
Slovakia	1 August 2021 to 31 July 2022	7 206	Providers and organisations acting in the public interest against child sexual abuse	
Slovenia	1 January 2021 to 14 July 2023	452	This number indicates criminal offences related to activities on internet. At present, the existing statistical data do not allow Slovenia to separate statistical data on offences investigated on the basis of reports submitted by providers and organisations from statistical data on other reports.	
Spain	2022	31 474	Organisations acting in the public interest against child sexual abuse	
Sweden	August 2021 to 31 December 2022	32 830	Mostly NCMEC	

Given that NCMEC is the main source of reports, it is informative to compare the figures on reports received by Member States to those provided by NCMEC on reports sent to Member States. NCMEC received a total of 29 397 681 reports from industry globally in 2021, of which 99.7% (or 29 309 106) contained one or more child sexual abuse images or videos, 0.15% (or 44 155) related to grooming and 0.05% (or 16 032) to child sex trafficking. In 2022, NCMEC received a total of 32 059 029 reports, of which 99.5% (or 31 901 234) related to child sexual abuse images or videos, 0.25% (80 524) to grooming and 0.06% (or 18 336) to child sex trafficking. For the EU, the breakdown is as follows:

Table n.7: NCMEC reports of suspected online child sexual abuse provided to EU Member States in 2021 and 2022

Country	Total reports 2021[10]	% of EU total 2021	Total reports 2022[11]	% of EU total 2022	% of EU population
Austria	7 580	1.36 %	18 501	1.23 %	2.00 %
Belgium	15 762	2.84 %	50 255	3.34 %	2.60 %
Bulgaria	13 584	2.44 %	31 937	2.12 %	1.53 %
Croatia	4 744	0.85 %	11 693	0.78 %	0.86 %
Cyprus	2 657	0.48 %	7 361	0.49 %	0.20 %
Czechia	15 004	2.70 %	61 994	4.12 %	2.36 %
Denmark	5 891	1.06 %	30 215	2.01 %	1.31 %
Estonia	2 729	0.49 %	6 408	0.43 %	0.30 %
Finland	6 079	1.09 %	10 904	0.73 %	1.24 %
France	98 233	17.67 %	227 465	15.13 %	15.16 %
Germany	79 701	14.34 %	138 193	9.19 %	18.59 %
Greece	14 616	2.63 %	43 345	2.88 %	2.37 %
Hungary	31 710	5.70 %	109 434	7.28 %	2.16 %
Ireland	7 327	1.32 %	19 770	1.31 %	1.13 %
Italy	37 480	6.74 %	96 512	6.42 %	13.32 %
Latvia	1 537	0.28 %	3 688	0.25 %	0.42 %
Lithuania	3 509	0.63 %	16 603	1.10 %	0.63 %
Luxembourg	2 005	0.36 %	2 004	0.13 %	0.14 %
Malta	750	0.13 %	4 713	0.31 %	0.12 %
Netherlands	36 790	6.62 %	57 012	3.79 %	3.96 %
Poland	37 758	6.79 %	235 310	15.65 %	8.41 %
Portugal	34 415	6.19 %	42 674	2.84 %	2.31 %
Romania	32 765	5.89 %	96 287	6.40 %	4.25 %
Slovakia	7 275	1.31 %	39 748	2.64 %	1.21 %
Slovenia	3 162	0.57 %	14 795	0.98 %	0.47 %

[10] NCMEC, 2021 CyberTipline Reports by Country, accessed in November 2023.
[11] NCMEC, 2022 CyberTipline Reports by Country, accessed in November 2023.

Spain	33 136	5.96 %	77 727	5.17 %	10.60 %
Sweden	19 635	3.53 %	48 883	3.25 %	2.33 %
Total	555 834		1 503 431		

The significant disparities between the number of reports in 2021 and 2022, showing a steep increase of reports in 2022, is due in large part to the decrease in voluntary detection between January and August 2021, when the Interim Regulation did not yet apply.

NCMEC does not differentiate in its statistics per EU Member State according to the source of the report, in particular whether it stemmed from a number-independent interpersonal communications service. However, NCMEC does provide statistics about the overall number of reports concerning the EU stemming from number-independent interpersonal communications services. In 2021, 283 265 reports concerning Member States stemmed from a chat, messaging, or email service, that is, 51 % of total reports concerning the EU. An additional 164 645 (30% of the total) of reports stemmed from social media or online gaming platforms, which may also have integrated messaging or chat services. In 2021, 3 565 reports concerning the EU were about grooming. In 2022, 1 015 231 reports concerning to Member States stemmed from a chat, messaging, or email service, that is, 68 % of total reports concerning the EU. An additional 325 847 (22% of the total) of reports stemmed from social media or online gaming platforms, which may also have integrated messaging or chat services. In 2022, 7 561 reports concerning the EU were about grooming. Again, the disparities in the number of reports from number-independent interpersonal communication services in 2021 and 2022 is due to the decrease in voluntary detection between January and August 2021, when the Interim Regulation did not yet apply.

The proportion of reports per Member State roughly matches the proportion of the population of the Member State as compared to the EU population as a whole in many cases, which could point to a comparable incidence of child sexual abuse online across Member States. Notable deviations are visible in relation to Spain and Italy, whose percentages appear low compared to the percentage of the EU population across both years, while the proportion of reports for other Member States appear to fluctuate significantly (e.g. Germany, Poland, Netherlands, Slovakia). These changes are not reflected as such in reports on numbers of cases and it is therefore again difficult to draw conclusions on correlation between reports and investigations.

In view of the differing reporting periods, no direct match is possible but nonetheless there are significant disparities between the statistics provided by NCMEC and the figures reported by Member States. In addition, the NCMEC figures for the Member States also cannot fully be matched with those provided by industry reported in the previous section. While some of the differences may be due to reports of child sexual abuse online that come from sources other than interpersonal communications, this would require further analysis, as it is also possible that voluntary detection measures concerning the EU by providers other than those that have submitted reports to the Commission to date is taking place, given the list of providers that are reporting to NCMEC[12]. Nonetheless, the fact that for most Member States there appears to be a

[12] NCMEC data is available here.

significant disparity between the number of reports NCMEC lists as having sent to the Member State, and the number of reports the Member State lists as received, suggests that the Member States' data collection and reporting is not complete.

For each NCMEC report identified above, the associated images and videos of child sexual abuse were taken down and removed from circulation. This is important in particular for current victims and survivors of child sexual abuse. Studies have shown that the continued circulation of images and videos depicting their abuse limits victims' ability to overcome the psychological effects of the abuse and creates a secondary form of victimisation.

2.2.2. The number of children identified

Most Member States provided complete or partial statistics on the number of children identified, differentiated by gender, pursuant to Article 8(1)(b) of the Interim Regulation. However, several Member States did not provide any data or justification for not reporting pursuant to this provision.

Several Member States did not report any or reported only partial statistics for the reporting period but provided reasons for this. The reasons provided include:

- child victims of online CSA cannot be counted (France);
- data not available as not collected as part of the national statistical data collection / the national authorities did not register these statistics (Denmark, Lithuania);
- data not disaggregated by gender in the national statistical data collection (Belgium, Cyprus, Czechia, Greece, Ireland, Italy, Lithuania, Netherlands);
- information is not available with the requested level of detail in the existing information systems (Finland);
- the information is not collected (Germany).

A few of those Member States that indicated that they were not able to provide statistics confirmed that their national authorities were asked to alter their registration procedure for voluntary reports and investigations and statistics collection (Denmark) and/or are deploying new information systems that should allow reporting at the required level of detail (Finland).

In one Member State, the data below do not differentiate between child victims of CSA online and offline (Hungary). In some instances, the statistics also include children that were identified as having produced and uploaded this material themselves (self-generated material, mostly video) (Czechia, Estonia).

As Member States mostly reported for differing reporting periods, it was not possible to calculate the total number of children identified as victims of online child sexual abuse in the EU, per year and/or the same reporting period.

Table n.8: Number of children identified, differentiated by gender

Country	Reporting period	Female	Male	Total	Comments
Austria	2021 to 2022	11	6	17	
Belgium	2021 to 2022	-	-	63	Data differentiated by gender not available.
Bulgaria	2022	50	12	62	
Croatia	1 January 2021 to 31 October 2022	20	0	20	
Cyprus	2022	-	-	102	Data differentiated by gender not available.
Czechia	2022	-	-	30	Data differentiated by gender not available.
Denmark	-	-	-	-	Data not available.
Estonia	2021	6	12	18	
Finland	-	-	-	-	Data not available.
France	-	-	-	-	Data not available.
Germany	-	-	-	-	Data not available.
Greece	2021 to 2022	-	-	4	Data differentiated by gender not available.
Hungary	2021 to 2022	379	47	426	Impossible to differentiate between victims of CSA online and offline. Only children below the age of 16 included.
Ireland	2021 to 2022	-	-	101	Data for 2021 (50 victims) cannot be differentiated by gender. Differentiated data by gender for 2022 are: 25 female and 26 male children identified.
Italy	2022	-	-	385	Data differentiated by gender not available.
Latvia	1 August 2022 to 6 March 2023	1	-	1	
Lithuania	-	-	-	-	Data not available.
Luxembourg	2021 to 2022	0	0	0	
Malta	-	-	-	-	Data not submitted/reported.
Netherlands	2021	-	-	222	Data differentiated by gender not available.
Poland	2022	2368	487	3014	In 2022, data from the National Police Information System in Poland provide the data on 3 014 victims of CSA related offences (2 368 female, 487 male, 159

					where the gender is not stated).
Portugal	-	-	-	-	Data not submitted/reported.
Romania	-	-	-	-	Data not submitted/reported.
Slovakia	August 2021 to July 2022	13	8	21	
Slovenia	1 January 2021 to 14 July 2023	220	85	305	
Spain	2022	80	39	119	
Sweden	2022	8	4	12	
TOTAL FOR ALL MEMBER STATES	**1 January 2021 to 6 March 2023**	**3 156**	**700**	**4 922**	

The data above are subject to a number of additional caveats. The existing statistical data do not always allow Member States to separate data on victims who were identified based on a report from a provider from those where, for example, the victim himself or herself may have reported the criminal offence or someone else who knew the victim or detected the criminal offence may have reported it (as mentioned by Slovenia). Sweden reported that children who have been identified through chat logs are also part of the reporting although photos or videos of the abuse never have been found or determined to originate from that specific victim.

Overall, the data in the table n.8 do not necessarily correspond to the reporting obligations in the Interim Regulation, which refer only to victims rescued thanks to the reports submitted by providers and by organisations acting in the public interest against child sexual abuse under the Regulation. The data provided in some cases include victims identified due to a variety of other reasons and means.

The data thus do not enable a comprehensive overview on the number of children identified as victims of online child sexual abuse in the EU.

In addition, even where a victim was identified, it does not necessarily mean that there has been a conviction linked to that identification. In some cases, the victim was identified, but the investigation could not establish a suspect or lead the investigation to a conviction (Sweden).

Nonetheless, it can be inferred from the data that a significant number of victims have been identified with the help of voluntary reporting in accordance with the Interim Regulation. This is confirmed by reports of cases from law enforcement authorities, which are often launched only on the basis of voluntary reporting[13].

[13] See for example a list of sample cases across the EU that were launched thanks to voluntary reporting by the companies in the impact assessment accompanying the Proposal for a Regulation to prevent and combat child sexual abuse (see in particular Annex 7).

2.2.3. The number of perpetrators convicted

While most Member States complied with their obligations, two Member States did not provide any data or justification for not reporting pursuant to Article 8(1)(c) of the Interim Regulation.

Several Member States did not provide any statistics for the reporting period pursuant to that provision and provided the following reasons:

- data not available yet (Belgium and Spain);
- the central database used to record crime did not require to record the nature of the initial referral (Ireland);
- the data is not collected (Germany).

Member States reported very varied data on the number of perpetrators convicted and there was no coherence in the reporting periods covered, as shown by table n.8 below.

Table n.9: The number of perpetrators convicted

Country	Reporting period	Number of convictions	Comments
Austria	2021	850	The data do not differentiate between offences committed online and offline.
Belgium	-	-	Data not available.
Bulgaria	2021 to 2022	52	
Croatia	-	-	Data not available.
Cyprus	2022	0	No convictions so far.
Czechia	1 January 2022 to 31 July 2022	20	
Denmark	2 August 2021 to 20 January 2023	224	
Estonia	2021	2	Includes only convictions resulting from NCMEC reports.
Finland	2021	240	
France	4 August 2021 to 3 August 2022	820	
Germany	-	-	Data not available.
Greece	2021 to 2022	62	
Hungary	2021 to 2022	126	
Ireland	-	-	Data not available.
Italy	2021 to 2022	5 835	The data do not differentiate between offences committed online and offline.
Latvia	2021 to 2022	33	The data do not differentiate between offences committed online and offline.
Lithuania	1 January 2021 to 30 June 2022	10	
Luxembourg	2022	11	The data do not differentiate between offences

			committed online and offline.
Malta	-	-	Data not submitted/reported.
Netherlands	2021	217	
Poland	Second half of 2021 to first half of 2022 / 2022 / First half of 2023	185 / 194 / 81	
Portugal	-	-	Data not submitted/reported.
Romania	-	-	Data not submitted/reported.
Slovakia	2021	10	
Slovenia	2021 to 2022	45	
Spain	2022	-	Data not available.
Sweden	2022	55	

It is important to note that the number of convictions does not equal the number of perpetrators convicted, as a person might be convicted for one or more offences of child sexual abuse online.

What is more, the statistics on convictions reported for a certain period are not necessarily linked to the reports that have been received in that given period (i.e. a conviction in 2022 could for instance be linked to a report from 2021 or 2020 and a report from 2022 might only lead to a conviction in 2023 or later). This fact was explicitly highlighted by several Member States in their reports (Ireland, Luxembourg, Sweden).

In some instances, no statistics were collected on whether reports of suspicious activities (e.g. via NCMEC) led to convictions, or in other words that convictions resulted from the information provided by a provider or a public organisation (Austria, Latvia). Only Estonia explicitly confirmed that the statistics show only convictions resulting from NCMEC reports. It is also possible that the reports led to other offenders, who are investigated and convicted in the course of the investigations (Austria).

Mostly it is presumed that the number of convictions reported is from when a case is finalised after having been presumably appealed through the judicial system. In one Member State (Denmark), where the numbers are compiled according to the latest decision, the numbers are not final, as the decisions may have been appealed afterwards.

In certain instances, the data contained in the national IT systems and reported by the Member States do not differentiate between offences committed online and offline (Austria, Luxembourg, Latvia).

The very varied reports submitted by the Member States and the way statistical data are gathered at national level thus do not allow for a comprehensive overview of the number of perpetrators convicted for online child sexual abuse in the EU. It is also not possible at present – on the basis of the data available – to link these convictions clearly to reports submitted by providers and organisations acting in the public interest against child sexual abuse in concrete reporting periods in accordance with this Regulation.

2.3. Developments in technological progress

Technologies currently used to detect child sexual abuse online include technologies and tools to detect 'known' (i.e. previously detected) CSAM, 'new' (i.e. not previously detected) CSAM and solicitation of children (known as 'grooming').

The examples given below include some of the most widely used tools, and they do not represent an exhaustive list. Many of these tools are made available to providers, law enforcement authorities and other organisations where a legitimate interest can be shown. Typically, these tools are combined with human review to ensure the maximum possible accuracy.

This section also includes additional developments in technological progress related to artificial intelligence.

2.3.1. Known CSAM detection

Existing technologies to detect known CSAM rely solely on automatic analysis of content[14] and are typically based on hashing. Hashing technology is a type of digital fingerprinting. It creates a unique digital signature (known as a "hash") of an image which is then compared against signatures (hashes) of other photos to find copies of the same image This technology only detects matching hashes and does not 'see' any material which do not match the hash. Hash values are also not reversible, and therefore cannot be used to recreate an image.

Many variations and implementations of hashing technology exist. Tools identified as used for known CSAM detections include: (i) Microsoft PhotoDNA; (ii) Google CSAI Match ; (iii) Apple NeuralHash + Private Set Intersection ; (iv) Meta SSN++; (v) PDQ and TMK+PDQF; (vi) MD5 Hash generator (Skype); (vii) Safer (Thorn).

The most widely used tool is Microsoft PhotoDNA, used by over 150 organisations[15]. PhotoDNA has been in use for more than 10 years and has a high level of accuracy. The rate of false positives is estimated at no more than 1 in 50 billion, based on testing[16]. PhotoDNA's error rate remains exceedingly low because of the nature of the technology. The technology exclusively detects copies of previously identified content. While the original PhotoDNA detects known CSAM in images, a version for detecting CSAM in videos is also available.

The technology is continuously developing and being improved. In May 2023, Microsoft announced the deployment of new matching capabilities that enable swifter searching (around 350 times faster), while reducing the cost of the matching process with no loss of accuracy. According to Microsoft, the new library also enables more comprehensive detection of flipped or

---

14 Providers do not consider metadata as an effective tool in detecting CSAM. See in particular p.10-11 of Pfefferkorn, R., Stanford Internet Observatory, Content-Oblivious Trust and Safety Techniques: Results from a Survey of Online Service Providers, 9 September, 2021.

15 Microsoft, Digital Crimes Unit.

16 Testimony of Hany Farid, PhotoDNA developer, to House Committee on Energy and Commerce Fostering a Healthier Internet to Protect Consumers, 16 October 2019.

rotated images. In addition, the Internet Watch Foundation (IWF) reported recently enhancing its hashing technology[17].

2.3.2. New CSAM detection

Technologies currently used for the detection of new CSAM include classifiers and artificial intelligence (AI) that analyse images and videos to detect content patterns that match patterns generated on the basis of previously identified child sexual abuse materials. A classifier is an algorithm that sorts data into labelled classes, or categories of information, through pattern recognition. Classifiers need data to be trained on and their accuracy improves the more data they are fed.

Tools to detect new CSAM include: (i) Safer (Thorn); (ii) Google Content Safety API; (iii) Facebook's AI technology[18]; (iv) Amazon Rekognition ; (v) Hive AI for visual content.

Research has shown that automated tools and systems such as classifiers are the most useful means of detecting CSAM.[19] For the detection of new CSAM, the accuracy rate currently lies significantly above 90%. For example, Thorn indicates that its CSAM Classifier can be set to 99% precision rate (for both known and new CSAM), meaning a 0,1% false positive rate[20]. These metrics are likely to improve with increased usage and feedback.

2.3.3. Grooming detection

Tools to detect grooming (solicitation of children) in text-based communications make use of technologies solely to detect patterns which point to possible concrete elements of suspicion of online child sexual abuse, without being able to deduce the substance of the content. The technique is applied to text-based chat conversations. Conversations are rated on a series of characteristics and assigned an overall probability rating, indicating the estimated probability that the conversation constitutes grooming. These ratings serve as a determinant, set by individual companies, to flag conversations for additional human review.

Tools used for text detection operations include: (i) Microsoft's Project Artemis[21]; (ii) Amazon Rekognition; (iii) Twitch's Spirit AI technology (based on NLP, text classifiers)[22]; (iv) Meta in-house internally built machine learning 'ranking' classifier (combining internal language analysis

---

17 Internet Watch Foundation (IWF), The Annual Report 2022, p.129–133.

18 See here and here for more information on Facebook's tool to proactively detect child nudity and previously unknown child exploitative content using artificial intelligence and machine learning.

19 Pfefferkorn, R.: Content-Oblivious Trust and Safety Techniques: Results from a Survey of Online Service Providers, Journal of Online Trust and Safety, February 2022, p. 1-38.

20 Thorn, Thorn's Automated Tool to Remove Child Abuse Content at Scale Expands to More Platforms through AWS Marketplace, 24 May 2021.

21 Microsoft's Project Artemis was developed in collaboration with The Meet Group, Roblox, Kik and Thorn.

22 For more information see: https://safety.twitch.tv/s/article/Our-Work-to-Combat-Online-Grooming?language=en_US

tech with meta data); (v) Roblox chat filtering[23]; (vi) Thorn and the Tech Coalition's technical solution based on machine learning and classifiers[24].

As in the case of identification of new CSAM, identifying grooming content requires training the technology with such content. Access to such training data remains the biggest challenge to the development and improvement of such technologies.

Thorn, in partnership with the Tech Coalition and its members, has launched a new initiative aimed at developing technical solution to identify and address attempts of online grooming that will be useful and usable for a range of platforms offering text-based communications. It will be based on Thorn's team's work on an NLP (natural language processing) classifier, or machine learning model, that detects and categorises when online content or behaviour falls into defined "classes" related to grooming (such as exposure to sexual material or seeking an in-person meetup with a minor) as well as an overall score for how related a conversation is to grooming.[25]

2.3.4. New challenges raised by Artificial intelligence chatbots and art/image generators

The development and release of AI chatbots such as ChatGPT (a large language model (LLM) developed by OpenAI) and art/image generators such as DALL-E[26] and Midjourney[27], has generated significant public attention, mainly due to their ability to quickly provide ready-to-use answers or create realistic images that can be applied to a vast number of different contexts. These new tools rapidly gained widespread popularity and use. Leading products are being funded and developed by tech companies including Microsoft and Google and the new technologies are being refined and improved versions rolled out on a regular basis.

While these technologies offer great opportunities to businesses and the public alike, they can also pose a risk for them. Concerns about such products include how criminals may wish to exploit them for their nefarious purposes including child sexual exploitation.

As reported by Europol, while all the information ChatGPT provides is freely available on the internet, the tool makes it significantly easier for malicious actors "to learn about a vast number of potential crime areas with no prior knowledge, ranging from how to break into a home, to terrorism, cybercrime and child sexual abuse". This enables said persons to better understand and subsequently carry out these types of crimes[28].

---

23 Roblox filters posts and chats for players age 12 and younger for inappropriate content and to prevent personal information from being posted, e.g. home addresses. This filtering system covers all areas of communication on Roblox, public and private. Roblox, Safety Features: Chat, Privacy & Filtering, accessed in July 2023.

24 Tech Coalition, New Technology to Help Companies Keep Young People Safe, 20 June 2023.

25 Tech Coalition, New Technology to Help Companies Keep Young People Safe, 20 June 2023.

26 DALL-E is an AI system that can create realistic images and art from a description in natural language.

27 Midjourney is a generative artificial intelligence program and service generates images from natural language descriptions.

28 Europol, ChatGPT - The impact of Large Language Models on Law Enforcement, 2023, ISBN 978-92-95220-57-7, page 7.

OpenAI's rules restrict ChatGPT's capability to respond to prompts for sexual, hateful, violent content or content promoting self-harm. Nonetheless, these safeguards can be circumvented fairly easily through prompt engineering.[29] Recent deployment of AI chatbots (e.g. by Snapchat) shows how these can cross the line into offensive or dangerous interactions, including child sexual abuse[30]. With more companies now considering testing AI chatbots on their platforms (Instagram, potentially WhatsApp and Messenger), the impact on users, especially children and young people, has to be carefully assessed.

These new tools also require adequate safeguards so that they are not misused to produce AI generated deep-fake child sexual abuse material.[31] With the pace of development of AI tools, it is likely that it will soon become significantly easier to generate images that are indistinguishable from actual images. This presents several key challenges in fighting CSA, as the ability for law enforcement to investigate and prosecute CSAM cases and identify real victims may become severely hindered if highly realistic computer-generated CSAM become highly prevalent online[32].

Research has shown that accessing child sexual abuse material is often the first step towards hands-on abuse, regardless of whether the material depicts real or realistically looking abuse and exploitation[33]. Limiting the dissemination of AI generated deep-fake child sexual abuse material is therefore crucial as a form of offender-side prevention. Another key concern is that groomers can use the advanced text-generation powers of ChatGPT, combined with existing free text-to-image AI to generate quickly and easily content for fake profiles and plausible conversations with young people to target children online. "Whilst ChatGPT in itself won't encourage people to become online groomers, it does allow anyone to feed the conversations they are having with children online through AI technology to make themselves more persuasive and credible to their victims, aiding manipulation."[34] Generative AI could have the potential to contribute to a rise in online grooming cases and even to "automate child grooming at scale"[35].

---

29 Swanson, S. M., ChatGPT Generated Child Sex Abuse When Asked to Write BDSM Scenarios, Vice, 6 March 2023; Mitchell, A., ChatGPT gives sick child sex abuse answer, breaking its rules, New York Post, 24 July 2023; Europol, ChatGPT - The impact of Large Language Models on Law Enforcement, 2023, ISBN 978-92-95220-57-7, page 5.

30 Fowler, G.A., Snapchat tried to make a safe AI. It chats with me about booze and sex, The Washington Post, 14 March 2023; Vincent, J., Instagram is apparently testing an AI chatbot that lets you choose from 30 personalities, The Verge, 7 July 2023.

31 Crawford, A., Smith, T., Illegal trade in AI child sex abuse images exposed, BBC, 27 June 2023.

32 Thiel, D., Stroebel, M., and Portnoff, R. (2023). Generative ML and CSAM: Implications and Mitigations. Stanford Digital Repository. Available at https://purl.stanford.edu/jv206yg3793. https://doi.org/10.25740/jv206yg3793. P. 2.

33 Protect Children, Protect Children's research in the dark web is revealing unprecedented data on CSAM users, 6 June 2021; RAINN, What is Child Sexual Abuse Material (CSAM), 25 August 2022.

34 Breck Foundation, Is artificial intelligence putting children at risk?, 9 February 2023, updated 3 April 2023.

35 Butler, J., AI tools could be used by predators to 'automate child grooming', eSafety commissioner warns, The Guardian, 19 May 2023.

3. CONCLUSIONS

Implementation measures taken by providers

Providers' reporting showed that they have been carrying out detection and reporting of child sexual abuse online under the Interim Regulation using a variety of detection technologies and processes. All providers reported sending these reports to NCMEC. In terms of type and volumes of personal data handled by providers, the reporting showed an array of traffic data collected and a varied level of granularity of volumes of data processed which prevents the Commission from obtaining unified EU level data relating to providers for the reporting period in question (July 2021 to 31 January 2023).

Providers did not submit the number and ratios of errors (false positives) of the different technologies used broken down by technology employed, indicating that they use a layered approach to detection of online CSA complemented by human review. At the same time, providers put in place a wide range of measures and safeguards to limit and reduce the error rate in their detection. What is more, providers reported having data retention policies and data protection safeguards in place, defined in their Privacy Policies or Statements and supported by industry standard data protection safeguards and measures.

Implementation measures taken by Member States

The Interim Regulation also obliges Member States (pursuant to its Article 8) to provide key statistics on cases of online child sexual abuse detected and reported to their law enforcement authorities, the number of children victims identified, and the number of perpetrators convicted. As Member States mostly provided data for differing reporting periods, it was not possible to calculate from the data submitted the total number of reports of detected online child sexual abuse received at EU level. In addition, the reports received and reported by Member States might differ from actionable reports, i.e. from reports that could be used for investigations, or number of cases reported. Only a few Member States indicated the type of providers on whose services the online child sexual abuse was detected. In some cases, the national statistical data do not differentiate between offences investigated on the basis of reports submitted by providers, and other organisations acting in the public interest against child sexual abuse and offences investigated on the basis of other reports.

Correspondingly, it was not possible to extract from the reports received the total number of children identified as victims of online child sexual abuse in the EU, differentiated by gender. The reasons include, for example: reporting data for differing periods; different age limits used to define child victims of CSA online; non-collection of statistics at such level of detail at the national level due to technical or other limitations; no differentiation between child victims of CSA online and offline etc. Some Member States include in their statistics also children that produced self-generated material. More importantly, the statistics often do not differentiate between victims identified based on reports submitted by providers and organisation acting in the public interest against child sexual abuse under the Regulation and those identified based on other reasons and means.

The overview of the numbers of perpetrators convicted is also fragmented. In certain cases, such data are not available as the databases did not record the source of the initial referral and thus the data do not differentiate between perpetrators convicted because of reports submitted pursuant to the Regulation and other reports. In some instances, the data contained in national IT systems also do not differentiate between offences committed online and offline. What is more, the statistics on convictions reported for a certain period are not necessarily linked to the reports that have been received in that given period but might relate to reports from earlier periods. Collected statistics on the number of convictions might also differ from the number of perpetrators convicted (as one perpetrator might have more convictions).

The heterogenous statistics submitted by the Member States, which appear to not always systematically and properly collect the data, and all the above-mentioned factors thus do not enable a comprehensive overview on the reports on online CSA received, the number of children identified as victims of this crime, or the number of perpetrators convicted at EU level pursuant to the Regulation. The fact that for most Member States there appears to be a significant disparity between the number of reports NCMEC lists as having sent to the Member State, and the number of reports the Member State lists as received, suggests that the data collection and reporting is not complete. Some Member States confirmed that their competent authorities are undergoing structural changes or reorganisation linked to the creation of new departments responsible for investigating crimes related to child sexual abuse online. New IT systems are also being put in place and national authorities were asked to alter their registration procedures and statistics in some Member States. This should create favourable conditions for having more accurate statistics from Member States in the future. In any case, the Commission will make use of its powers under the Treaties as needed to ensure that Member States comply with their reporting obligations under the Interim Regulation.

General considerations

Overall, this report shows considerable disparities in the reporting on data relating to combating online CSA under the Interim Regulation by both providers and Member States. Greater standardisation of available data and the reporting thereof, such as those provided in the proposal for a Regulation to prevent and combat child sexual abuse[36], would contribute to obtaining a better picture as regards relevant activities in the fight against this crime. It appears that additional efforts by providers and Member States are needed to ensure data collection and reporting in line with the requirements of the Interim Regulation.

The available data shows that under the current voluntary detection and reporting system it is possible that materials automatically flagged as possible CSAM result, upon human review, to be not CSAM. This can be due to the lack of a common set of hashes and other indicators to detect CSAM, confirmed as illegal in the EU, or differing legal standards across jurisdictions, notably between the EU and the US, in particular on the relevant definitions. The data also suggests large variations in the number of review requests, and review success rates, from which it is not

[36] 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.

possible to extract conclusions, given the lack of information on notably the scope of the review requests and the reasons to reinstate.

As regards the requirements of Article 9(2) on the conditions for the processing, the information provided indicates that the technologies used correspond to technological applications designed for the sole purpose of detecting and removing online child sexual abuse material and reporting it to law enforcement authorities and to organisations acting in the public interest against child sexual abuse. No information was provided in relation to whether the deployment of the technologies was in accordance with the state of the art and in the least privacy-intrusive way, and on whether a prior data protection impact assessment as referred to in Article 35 of Regulation (EU) 2016/679 and a prior consultation procedure as referred to in Article 36 of that Regulation had been conducted.

As regards the proportionality of the Regulation, the central question is whether the Interim Regulation achieves the balance that the EU legislature sought to strike between, on the one hand, achieving the general interest objective of effectively combating the extremely serious crimes at issue and the need to protect the fundamental rights of children (dignity, integrity, prohibition of inhuman or degrading treatment, private life, rights of the child) and, on the other hand, safeguarding the fundamental rights of the users of the services covered (privacy, personal data protection, freedom of expression, effective remedy). The available data is insufficient to draw definitive conclusions in this respect. It is not possible nor would it be appropriate to apply a numerical standard when assessing such proportionality in terms of number of children rescued, given the significant negative impact on a child's life and rights that is caused by sexual abuse. Nonetheless, in light of the foregoing, there are no indications that the derogation is not proportionate.

Despite the shortcomings of the available data, which do not allow insight into the use of voluntary reports in a significant number of Member States, it is clear from the data that is available that thousands of children were identified in the reporting period, more than two thousand convictions were obtained, and millions of images and videos were removed from circulation, reducing secondary victimisation. Therefore, it can be concluded that voluntary reporting contributed significantly to the protection of a large number of children, including from ongoing abuse, and it appears that the Interim Regulation is effective.