A Rapid Evidence Assessment of Technical Tools for the Detection and Disruption of Child Sexual Abuse Media (CSAM) and CSAM Offenders in The ASEAN Region

Corinne May-Chahal, Claudia Peersman, Awais Rashid, Maggie Brennan, Emma Mills, Peidong Mei, John Barbrook

March 2022

Introduction

Taking a public health approach, this review contributes to online child sexual abuse tertiary prevention (post-event), assessing the effectiveness of automated tools used to identify and support the removal of CSAM. International research is first appraised for its relevance to CSAM prevention before examining research specific to the ASEAN region.

Gaining accurate data on the prevalence of CSAM is challenging; CSAM media is constantly being produced and identification of it is dependent on reporting. It is also constantly evolving, for example, there is no reliable data on the prevalence of web streaming (Quayle and Koukopoulos, 2019). The main source of data on current trends is from NGO reporting, such as NCMEC, IWF and the Canadian Centre for Child Protection, or from international law enforcement (Interpol, Europol). A wide range of sources report CSAM to these agencies including ISPs, civil society organisations and members of the public. Reports have increased significantly over the last decade and particularly since 2020, as more day-to-day activity has transferred online during the Covid 19 outbreak (Parks et al., 2020). This continues growth observed before the pandemic. Analysis of 23.4M NCMEC reports (Burszstein et al, 2019) finds that 40% of CSAM produced up to 2019 emerged in 2017. This analysis, along with others (IWF 2015, 2016, 2017, 2018, 2019, 2020 and 2021) also find that CSAM distribution through Peer to Peer (P2P) networks is decreasing

Deeper analysis of trends finds that the relationship between absolute numbers of CSAM perpetrators and content in P2P is not straightforward. Bissias et al. (2016) report on activity in 5 P2P networks across 2012-2014, finding that individual CSAM sharers reduced over the period, but content and content severity increased. In the US the total number of files shared increased to 122,000 per month in 2014, up from 42,000 per month in 2012. Furthermore, an association was found between severity of content, including the sexual abuse of infants and toddlers, and contact offending. Peers sharing the most severe content (Level 8) had a significantly higher rate of contact offences (28.8%) than those sharing Level 1-4 content (15.4%). The authors conclude that CSAM in P2P still outpaces law enforcement capacity or resources to adequately respond and suggest that tool development should be focussed on detecting severe content.

Methods of production and dissemination have evolved over time, meaning that P2P has become part of an expanding CSAM file sharing system. Steele et al (2020) identify five distinct phases, beginning in the late 1980's key technologies in use by offenders included floppy disks, email, scanners and modems. The diversity of images was limited as CSAM producers transitioned from print sources and had less access to computers. From the mid 1990's access to the internet became easier and image collections expanded. Early CSAM video material and use of encryption were both observed during this period. Peer to peer file sharing emerged in the mid 2000's as platforms such as Bittorrent and eDonkey became popular, marking a significant increase in CSAM dissemination. Access to the dark web followed along with the introduction of cryptocurrencies, cyberlockers and far greater anonymisation options. The 'mobile era' has expanded CSAM production and dissemination capacity on a range of dimensions; countermeasures are increasingly built into the software and there has been massive expansion in children's direct and unsupervised access to tablets and mobile devices increasing opportunities for abuse content creation, both live and stored. Thus, rapid technological evolution means that a much wider range of methods can be applied, whilst older methods are retained with new functionalities in a constantly evolving CSAM system. For example,

"Web-based forums can be used to share Bittorrent links, and encrypted files can be shared from public cyberlockers. Peer-to-peer software that shares encrypted (and innocuously labelled) binaries can be run over the Tor network, and the decryption passwords and pointers to the content shared on Usenet newsgroups accessed via the web. The Ares peer-to-peer network client now includes Bittorrent link capabilities, an integrated image/movie viewer (allowing it to be used to view and not just download content), and an integrated chat function" (Steel et al, 2020 p14).

Related work

Research that aims to create automated tools to detect CSAM has expanded in parallel with the pace of change in technology and offender behaviours. Two reviews have been published since 2016 summarising technical approaches relevant to CSAM detection. Lee, Ermakova, Ververis and Fabian (2020) report a comprehensive review across three databases (IEEE Xplore, Google Scholar and ScienceDirect) using four related search terms. This retrieved 21 results, 12 of which pre-dated 2016. Five concepts were analysed: policy and legal frameworks, distribution channels, technology, implementations, and challenges. Regarding policy, drawing on ICMEC's Global Review (ref) Lee et al note that legal frameworks are not comprehensive, in 2018 less than two thirds (n=118) of 196 countries had sufficient legislation, with wide variation in definitions and coverage. Blocking efforts (e.g., by Google and Microsoft) have a deterrence effect but this may only displace CSAM as it moves to platforms and jurisdictions that fail to act. A range of distribution channels were covered by the research including peer to peer networks (n=10), darknet (n=2), websites and website engines (n=3), mobile devices (n=1) and social media (n=2). Across the studies, four technologies were applied in CSAM detection: image hash databases (n=6), web crawlers (n=4), visual detection algorithms (n=12) and AI/deep learning methods (n=6). The latter outperform other methods, but multiple approaches combined were found to be most effective. Three implementation approaches are presented: Google's AI implementation, Microsoft's PhotoDNA (with extension to video), and IWF and the Canadian Child Protection Centre's Project Arachnid web crawlers. Challenges continue to be the lack of access to CSAM databases to test tools, focused research on live streaming (of which there is very little), prosecution of perpetrators and transparency (explanations of how tools work may enable perpetrators to develop countermeasures). The authors note that despite all research efforts, CSAM continues to grow.

Over 40% of sexually explicit material is self-produced, either voluntarily or coercively (UK Quayle et al, 2017); a distinction that is not always made by the child involved who may be a victim of deception, coercion or loss of image control. A comprehensive metareview on 'sexting' (39 studies) found 1 in 7 adolescents send sexts and 1 in 4 receive them. Images were included in all 39 studies, video was reported in over 50% (21 studies) and sexually explicit messages included in over a third (14 studies) (Madigan et al, 2018). Tariq et al (2019) review adolescent sexting from a 'human lens' perspective. They critique evaluation measures used by software engineers and technical researchers that only report on the success of the selected algorithmic approach, using measures such as precision, accuracy, recall and F1. A key finding from their analysis of 45 papers published between 2008-2018 is that none report user evaluation by the child. The primary model applied to the detection of self-produced media across the decade was nudity/skin detection, with 86.4% of papers deploying computer vision techniques and only 9% using natural language processing (NLP). Two of the 45 papers focused on mobile device detection, with the remainder producing general solutions. A significant conclusion is that all the approaches are applied post-production, detecting risk, rather than risk mitigation that prevents production. Suggestions for risk mitigation strategies include in device mobile applications that warn or block adolescents before an image is sent, or skin detection in live video (see Tariq et al, 2018 for proof of concept).

Given the pace at which technology evolves and the continuing expansion of CSAM, the current review provides an updated state of the art account of CSAM research.

In addition to supplementing previous reviews, it has a specific emphasis on the effectiveness of the research for CSAM prevention as it applies to the ASEAN region, where a large proportion of live streamed CSAM originates.

Research Questions

The research question for the evidence assessment was divided into two parts:

1. Are tools designed to prevent child sexual abuse media (CSAM) effective in prevention?
    a) What tools for the disruption of child sexual abuse media in digital environments are effective in prevention?
    b) What tools for the disruption of live streaming of child sexual abuse are effective in prevention?
2. What in 1) is specific to the ASEAN countries?

Method

The substantive areas of interest were translated into a standardized systematic search framework that takes account of sample, phenomena of interest, design, evaluation and research type SPIDER (Cooke, Smith and Booth, 2012) as it applied to the review questions (see Figure 1).

SPIDER Tool	**Search Focus**
Sample (Population of Interest)	Digital/Forensic Tools; Services
Phenomenon of **I**nterest	Online CSA/E; ASEAN countries
Design	All relevant (e.g. meta-analysis, survey, interview, case study)
Evaluation	Effectiveness, measures of prevention
Research	Technical, qualitative, quantitative, mixed methods

Figure 1: SPIDER Search Strategy

A scoping search was conducted by an information specialist to explore multiple research areas of interest and then filtered with Boolean operators. For example, searches were first constructed in the fields of digital tools and CSAM separately and then combined (e.g. digital tools AND CSAM) as the search fields. This scoping search allowed further identification of specific key words that were frequently referred to in relevant literature, which were then added to search strings. The search process involved testing terms, strings and combinations of strings, for example:

S1 “Digital” OR “Technical” OR “Comput*” AND “Tools” OR “Forensics”
S2 “Law enforcement” OR “Police” OR “Services”
S3 [“child” OR “young” OR “peer” OR “youth” OR “adolescent” OR “minor” OR “teenage*”] AND [“sexual exploitation” OR “sexual abuse” OR “sexual violence” OR “sexual slavery”] OR “Prostitution” OR “Child-Sex Tourism”
S4 [“online” OR “on-line” OR “live stream*” OR “groom*” OR “image*” OR “video” OR “media” OR “webcam” OR “website” OR “CSA material” OR “CSA content”]
S5 “cyberped*” OR “cyber-ped*” OR cyber-paed*”
S6 [“ASEAN” OR “Brunei” OR “Cambodia” OR “Indonesia” OR “Lao” OR “Laos PDR” “Vietnam” OR “Viet Nam” OR “Malaysia” OR “Thailand” OR “Singapore” OR “Myanmar” OR “Burma” OR “Philippines” OR “South-East Asia” OR “South East Asia” OR “Southeast Asia” OR “Asia Pacific” OR “East Asia”]
S7 “Prevention” OR “evaluation” OR “effective*” OR “protection” OR “Assessment”
S8 “Disrupt*” OR “Detect*” OR “Deter*”

Terms were exploded, where possible, in Academic Search Ultimate, and 10 final search strings were run on subject, title and abstract fields across nine databases: Academic Search Ultimate, ACM, CINAHL, IEEE, MEDLINE, PsycINFO, Scopus, SocINDEX and Web of Science. This returned 2944 results, reduced to 2497 following de-duplication. Taking account of previous reviews and the rapid pace of technology development, references dated before 2016 were excluded leaving 1008 papers for screening (Figure 2).

A title and abstract sift protocol (TAP) defined inclusion criteria and indexing strategy. The TAP was organized in the format of a flowchart and contained five main questions (see Appendix). Papers were sifted against all of the review questions in a sequential order until excluded or included. Three rounds of calibration involving 85 references from the initial search return were completed between two reviewers to test the TAS protocol and train the reviewers. In each calibration exercise, 30 references were randomly selected and sifted independently by two reviewers using the systematic review platform *Rayyan* (Ouzzani et al., 2016). Before moving to the next round, disagreement was resolved by discussion among the two reviewers and the first author, and a final decision was reached. The TAP was refined during each calibration exercise in light of the discussion. This calibration process was repeated three times until interrater reliability improved from Kappa 70.0% to 89.1%. Subsequently, references were divided between two reviewers to independently sift following the protocol.

Within the exclusions, 86.8% were not CSAM relevant and a further 8.4% mentioned CSAM but not any technical interventions. Another 0.9% were excluded as commentaries. This resulted in 188 references (18.7%) included for full-text review, of which 180 could be retrieved. Studies were excluded after full text reading for the following reasons:

- Tools developed and tested exclusively on adult data
- The technology applied to CSA interventions other than CSAM detection (e.g., offender management, education programmes, tools for parental supervision)
- CSAM detection mentioned as relevant to the research but not the primary focus
- Focused solely on victim grooming in social media
- Duplicated research (research published in different locations)

31 studies met the following final inclusion criteria, which were that the research reported was:

- Peer-reviewed
- Published between 2016 and 2021
- Focused primarily on computational/algorithmic approaches to CSAM detection
- The approach/model/tool was evaluated or was a systematic review

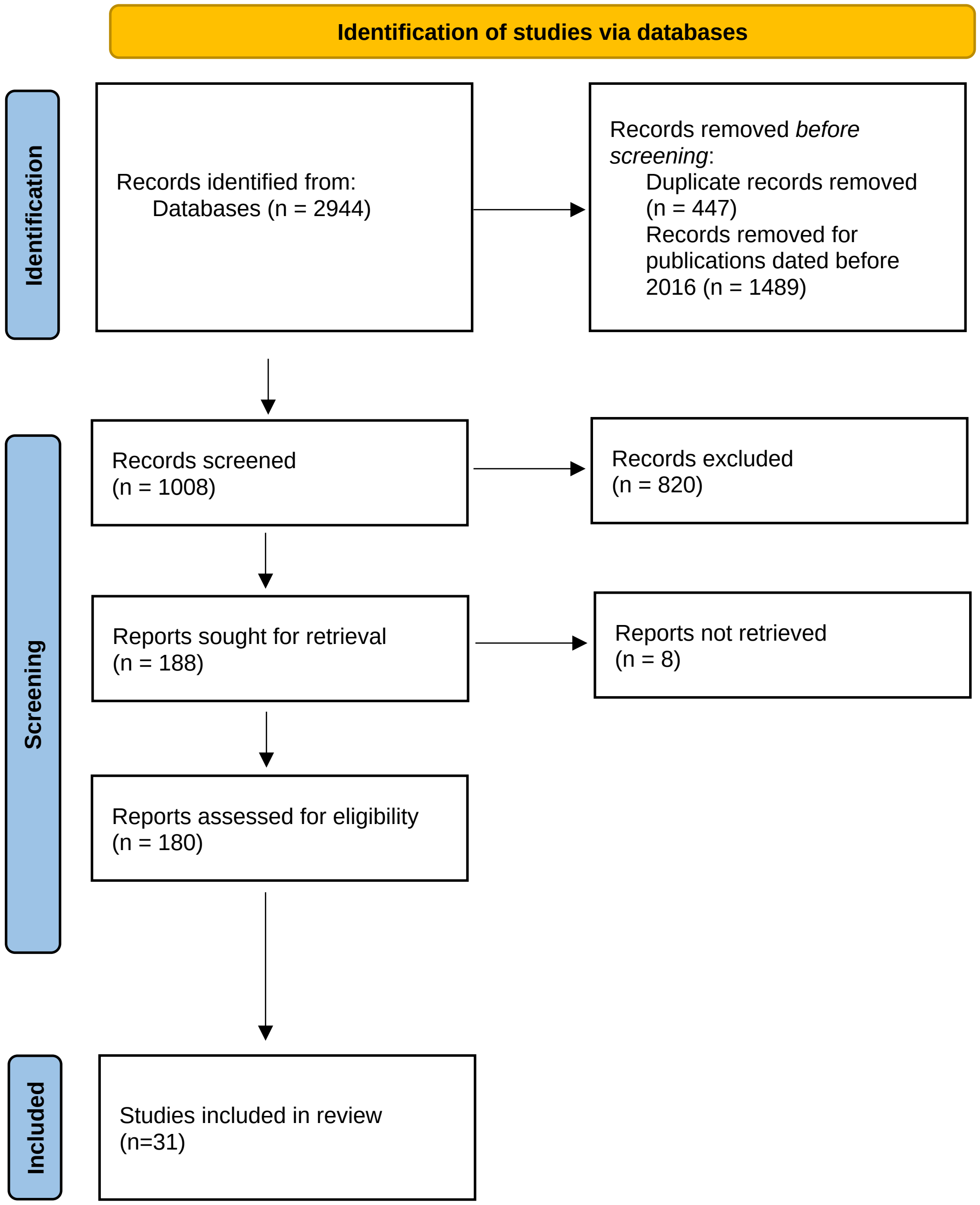

Identification of studies via databases
Identification
Records identified from:
Databases (n = 2944)
Records removed *before screening*:
Duplicate records removed (n = 447)
Records removed for publications dated before 2016 (n = 1489)
Screening
Records screened (n = 1008)
Records excluded (n = 820)
Reports sought for retrieval (n = 188)
Reports not retrieved (n = 8)
Reports assessed for eligibility (n = 180)
Included
Studies included in review (n=31)

Figure 2. Screening and Inclusion Process

Results

Publication characteristics

An average of five papers were published each year across the period, with above average outputs in 2019 and 2021 (Figure 3).

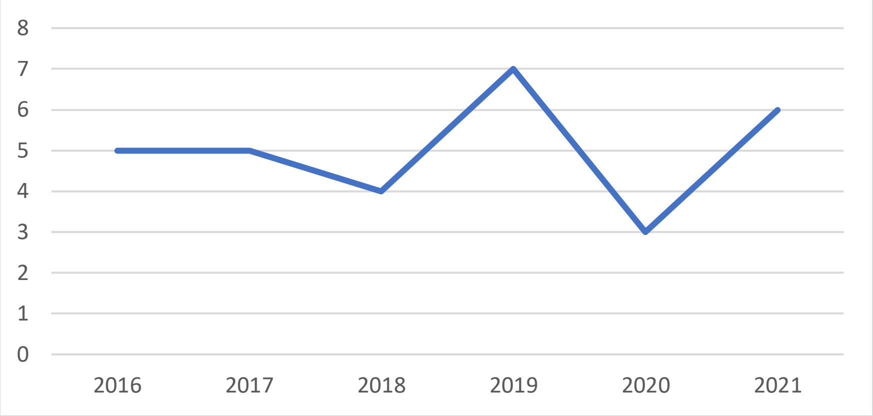


Figure 3: Date range of publication

CSAM research was reported across a range of peer reviewed forums; over half of the outputs were found in conference proceedings, 5 in technically oriented journals, 5 in social science journals, 4 in forensic journals with the remainder in open-source repositories. 30 papers were assessed as international in scope, with the highest number of researchers affiliated to institutions in Europe (n=13), which reflects EU investment in this field. 7 studies had US authors, with the remainder South American (n=4), Australian (n=3), SE Asian (n=2) and the UK (n=2) (Figure 4).

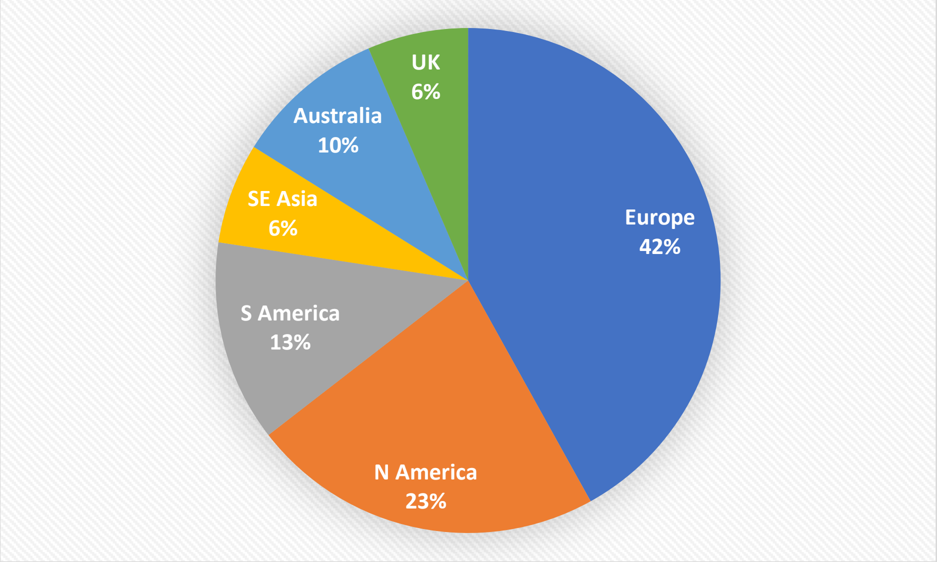


Figure 4: Researcher Location

Data sources in CSAM research

A wide range of data is drawn on in the reviewed research, dependent on method and models. Despite the known problem of access to 'real world' CSAM data sets, just over one fifth of research in our sample did have such access. This was either from collaboration with law enforcement (Brazil, Spain and the European Framework collaboration 4NSeek), or with reporting NGO's (NCMEC, IWF, CCCP, Project Vic). To address the lack of access for researchers unable to establish these links, a Brazilian team have offered access to their database, following an application procedure (ref). A further fifth of studies relied on textual samples, including posts in dark web forums, or social media platforms such as Omegle or Twitter. 18% sought to train algorithmic approaches for age and face recognition using generic images harvested from the web or known datasets such as IMBD and MORPH, with the explicit intention that the resulting technology could then be applied to support CSAM detection. The lack of validated child datasets was noted by several authors, with one study focused on constructing a generic child image data set that could be shared by researchers in the field (ref).

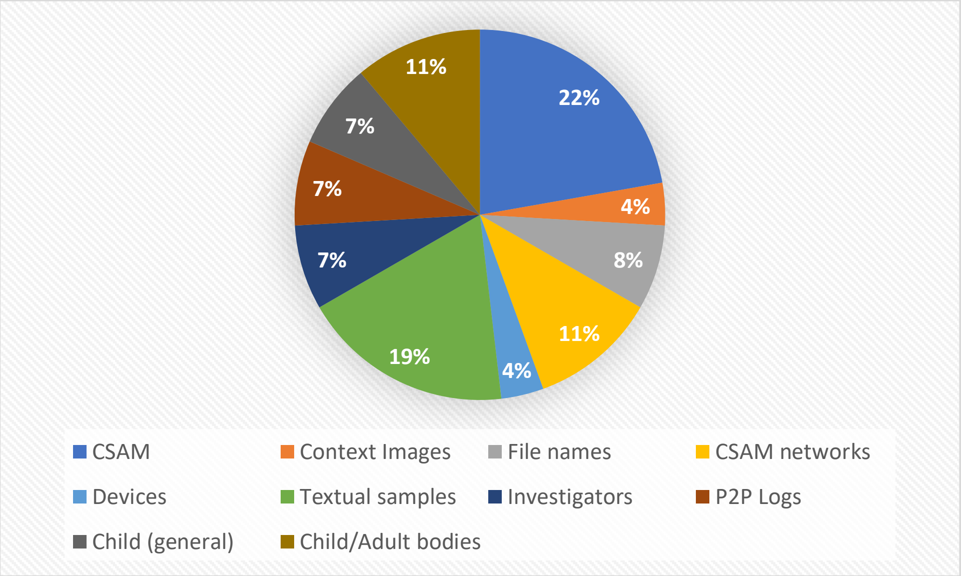


Figure 5: Data Sources

From a child centred lens, it is important to consider the characteristics of CSAM that are the focus of research attention. In just under a third of studies this was on perpetrator activity including analysis of posts in forums, social media and P2P networks, website activity and chatbot engagement. A fifth of studies focused directly on identifying child sexual abuse activity in media content. A quarter (n=8) sought methods to identify related visual data, such as child nudity, child age, place of abuse and text within images. 15% focused on media metadata, such as hash values, file names, file paths, and artefacts left in devices. Only three studies focused on the law enforcement task, all at a national level (Australia, US and Malaysia) (Figure 6).

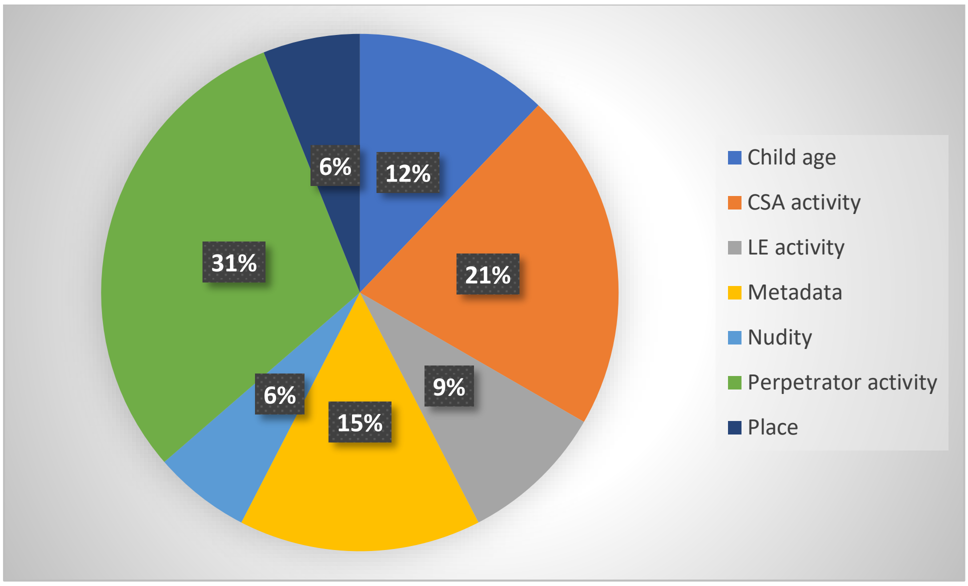


Figure 6: CSAM Characteristics Focused on in the Research

Methods and Models in CSAM research

A wide range of methods and models were covered by the research (Figure 7), which can be broadly distinguished as technical and social science approaches.

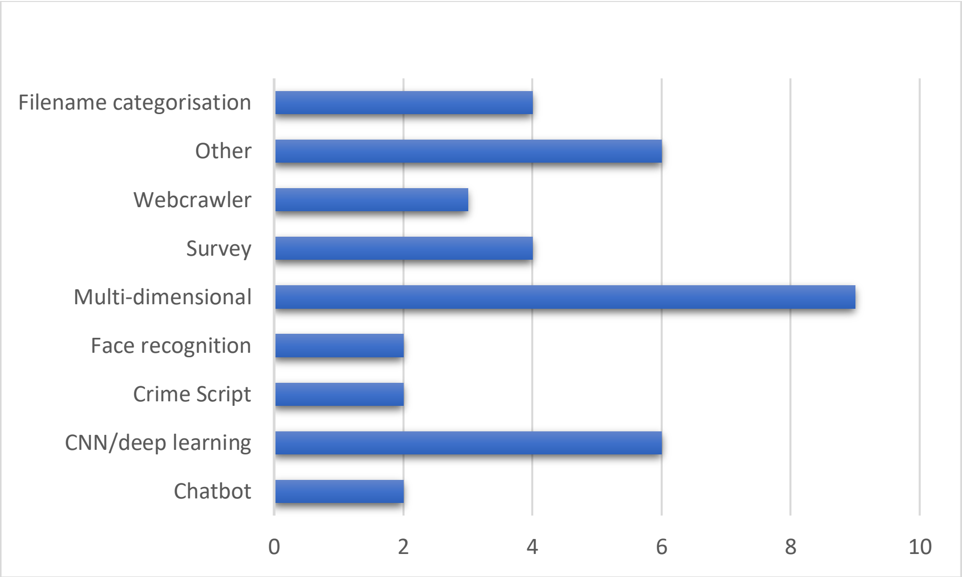


Figure 7: Methods and Models for CSAM Detection

Technical models

As Lee et al (2020) also found, all technical approaches were designed to detect CSAM post-production. Gangwar et al (2017) identify five salient approaches in previous research and test their effectiveness through five different models; skin detection by colour, nudity detection, HSVColor-SIFT, shallow CNN and deep learning. When tested on a dataset of real CSA images, deep learning methods achieved the highest overall accuracy at 88%. The authors note that some features were difficult to identify, such as where victims are clothed, and where children are not fully visible.

Several multi-dimensional approaches were identified that could overcome this problem, such as combining filename categorisation (keywords, n-grams, semantic features), colour-correlograms, RGB skin colour, visual Words and pyramids & audio words (Peersman et al, 2016). A further angle was to apply staged approaches to the problem, such Macedo et al (2018) which first identified pornography and then applied child face and age detection, an approach that achieved 79.84% accuracy and out-performed two tools currently used by the Brazilian Federal Police (NuDetective and LED) when applied to LE datasets. Taking a similar approach, Gangwar et al (2021) developed a deep CNN architecture, combining pornography detection and the age-group classification networks using decision level fusion for binary CSA classification and score level fusion for the rearrangement of suspicious images. This model was tested on a data set including 5,000 validated CSA images achieving 92.7% accuracy. Their work also introduced two new datasets: Pornographic-2M, containing two million pornographic images, and Juvenile-80k, including 80k manually labelled images with apparent facial age.

The majority of CSAM reports to LE come through independent clearing houses such as NCMEC, Canadian Centre for Child Protection and IWF. These NGOs are experiencing a challenge in processing the increasing volume of material. Bursztein et al (2019) suggest improvements in scene clustering and facial recognition technologies would help, incorporating new deep learning models, several of which are included here. They also reinforce the need to prioritise technology development on the detection of new content, which the main tool used by ISPs (Microsoft’s PhotoDNA) cannot achieve, across all digital media (image and video). Finally, they caution that as distribution technologies emerge, solutions beyond P2P are critical. A recent partnership between Microsoft and Project Vic may offer a solution, the first study to apply supervised learning to detect CSAM file paths (Pereira, Dodhia and Brown, 2021). A dataset of over 1 million filepaths attributed to four different labels (non-pertinent, CSAM, Child Exploitative/Age Difficult and CGI/Animation) was used to test three approaches: extracting bag of words, computing n-grams, and deep learning models. The CNN model outperformed all other approaches with accuracy, precision and recall all higher than any previous studies (97%, 94% and 94% respectively).

Westlake & Bouchard (2016; 2017) developed CENE, a webcrawler, to identify child sexual exploitation networks. Networks seeded from child exploitation websites were compared to networks seeded from similar non-CE sexuality websites and sports websites. They provide evidence that webcrawlers have the potential to provide valid CE data. Since this study was published, webcrawlers have been applied by NGO's such as the Canadian Child Protection Centre (Project Arachnid) and IWF to effectively detect and take down images. In recognition of the displacement of CSAM to the darkweb, Dalins et al (2018) introduced the Tor-use Motivation Model (TMM) which distinguishes between content (e.g. CSAM, drugs, weapons) and motivations (e.g. information sharing, marketplace) to enable law enforcement to better distinguish illegal activity in Tor sites. Their crawler, authorised by the Australian government, gathered 232,792 pages from 7651 Tor virtual domains. An analysis of 4000 unique Tor pages found child exploitation domains comprised approximately 1.75%, which was similar to illicit/illegal pornography 1.75% and adult pornography 1.4%.

Interest in the development of chatbots as a prevention measure has grown since the success of the Terre des Hommes operation "Sweetie" in 2013 which launched an online avatar of a Philippino child, leading to the arrest of over 1000 men (https://www.bbc.co.uk/news/uk-24818769). A second version is now in development, Sweetie 2.0 at the University of Tilburg (Hensler and Wolf, 2019). Two further teams have created and run chatbots since 2016. Zambrano et al (2017) present an experimental model for a distributed platform named BotHook, that contains a module designed to attract paedophile interest, whilst Triviño et al (2019) created C3-Sex, which connects to online chat sites to start conversations on child abuse images. The system differentiates between three categories of CSAM interactors ('indifferent, interested and perverts') and was evaluated on 326 chats in Omegle, finding 4 interested and 5 pervert individuals.

Finally, the 'other' category includes a crowdsourcing model to accumulate a dataset of hotel rooms (Stylianou et al, 2017), which contains 2.85 million images from over 254,000 hotels across the world to assist in CSAM location identification. In addition to Omegle, two further studies included in the present review focus on social media.

Riesco et al (2019) tested six different approaches to text classification on Pastebin, an anonymous online notepad service, of which there are now several. Whilst only a small proportion of content, CSAM posts were retrieved. Studies on Twitter also find CSAM relevant material. For example, using a classification model to detect sex trafficking of underage girls, a Spanish ML study detected 32453 suspicious tweets in data harvested over four days (Hernández-Álvarez, 2019).

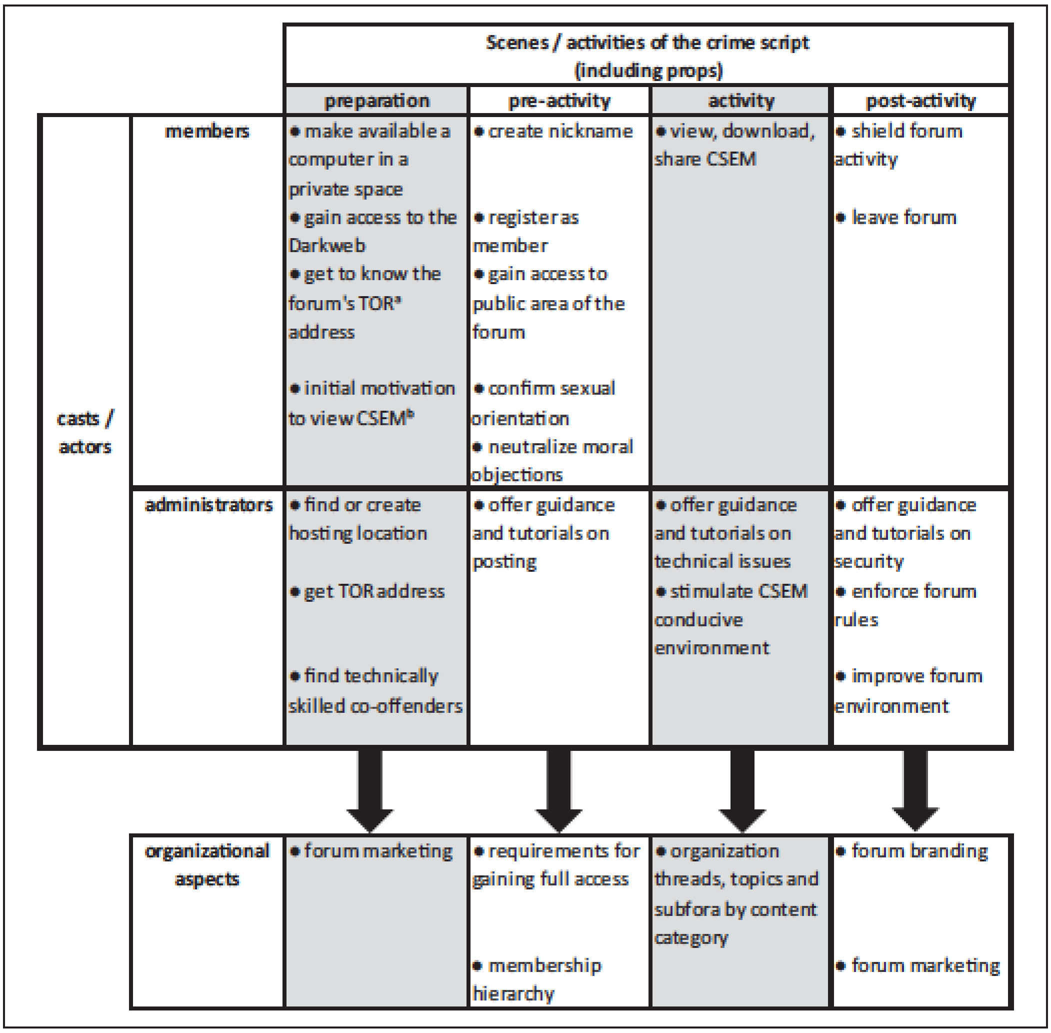
Also referenced is the only study to focus on CSAM live streaming (Horsman, 2018). Academic research on live streaming is scarce (Lee et al, 2020), primarily focused on the legal context (see, for example, Dushi, 2019). Much of what is known about live streaming derives from law enforcement sources (Acar, 2017; Europol 2019; GACSAO 2016) and NGO reports (UNODC, 2017; ECPAT International 2017; IWF, 2018). In summary, this establishes the live streaming of CSA as a growing problem, of two dimensions. Commercial live streaming is known to be prevalent in the Philippines and Thailand and known to be linked to child trafficking routed through Myanmar, Cambodia and Viet Nam (UNODC, 2017). According to UNICEF the Philippines is the largest source of live streamed CSAM. In 2017 it was estimated that 149 of every 10,000 IP addresses, a rise from 43 per 10,000 in 2014 (Merten, 2020). A second form is generally non-commercial, deriving from peer sharing of Self Generated Indecent Media (SGIM) online as streaming has become more accessible. Most current digital forensic detection methods do not apply to this problem as live streamed CSEM leaves no visual data unless captured. Because this form of CSEM frequently involves financial exchange research is beginning to focus on detecting patterns in financial activity. A rare study by Australian researchers sought to identify live streaming offender profiles and patterns of financial transactions using government agency data on 256 people who made 2,714 payments to CSA live streaming facilitators in the Philippines. Transactions were not evenly distributed; 50% were made by just 8 individuals and 48% made only one. The majority (75%) were worth AU$ 170 or less, although the total value was AU$ 1.32m (Brown, Napier and Smith, 2020). A further promising forensic direction is analysis of artifacts remaining on devices, which has only been evaluated in relation to CSAM by Horsman (2018). Horsman’s experimental study focused on Periscope finding it is possible to find some artifacts (emails, user ID, message pathways) but nothing that can directly point to the stream watched, unless data is captured in real time with the permission of the user (see also an experimental study focused on forensic analysis in Twitch, Al Zahrani, Ahtisham and Bhat, 2020).

However, Horsman finds that still images from videos are available and that cached imagery can enable a partial stream recovery, although it was not possible to identify whether these related to a livestream or viewed stream. Live streams can be captured, a process known as ‘capping’, and are increasingly found in P2P networks and other online or cloud locations (IWF, ref). Once capped, CSAM video detection methods can be applied.

Social science approaches

Crime script analysis (CS) is a criminological approach to understanding crimes and thereby preventing them. This method identifies elements, such as ‘scenes, paths, actions, roles, props, and locations’ (Cornish, 1994), at different stages in the commission of crime from preparation to exit. The aim is to identify intervention points and potential prevention actions. CS has been applied to offline CSA (Leclerc et al, 2011) and more recently advocated as an approach for studying cybercrime more broadly (Dehghanniri and Borrion, 2016). Subsequently, two research studies have reviewed CSAM through a crime script lens. Van der Bruggen and Blokland (2021) conducted a CS analysis from CSAM darkweb fora posts seized by the Dutch police. Criteria for inclusion were that the forum had over 10,000 posts and members. Up to 100 posts were randomly sampled from subforums in four fora, confirmed as representative by law enforcement (n=4905). The analysis focuses on four crime script stages: preparation, preactivity, activity, and postactivity for the two roles that were identified (members and actors) (see Figure 8). The crime scripted is that of access to, and sharing of, CSAM. Rich description of activity in each stage is provided, including what members do to join and how they rise through the fora hierarchy by posting new and more extreme media. The fora shed some light on the production of CSAM, containing hints and tutorials for this purpose. In terms of intervention, the authors recommend targeting administrators along with members providing technical support to disrupt maintenance of the fora. Understanding the crime script may also help in therapeutic interventions, thereby aiding prevention of further CSAM activity.

In Australia, Leclerc and colleagues (2021) also develop a crime script analysis of CSAM on the darknet through interviews with 29 law enforcement personnel. They simplify the four stages identified by van der Bruggen and Blokland into three: crime set-up, crime completion and crime continuation. Set-up involves learning about Tor and accessing the darknet, crime completion involves joining fora, identity protection and interaction with members to view and share CSAM. The continuation phase has three possible scenarios; the offender consumes and leaves, they consume and distribute or they increase their activity to move up the forum hierarchy. In both studies the identification of the crime script at this more general level appears to provide information that is already known by law enforcement. Leclerc et al recommend more detailed research at each stage. In a review structured by CS analysis, Fortin et al (2018) examine the script of CSAM offenders who commit offline CSA. Four episodes (or stages) are identified: the consumption of legal online pornography; possession of CSAM through searching, P2P networks and fora; CSAM distribution through chatting, use of specialized tools and advanced key words; contact sexual abuse and the production of CSAM (see Figure 9). It is acknowledged that not all CSAM offenders follow this metascript with only a minority reaching the final episode.



		Scenes / activities of the crime script (including props)			
		preparation	pre-activity	activity	post-activity
casts / actors	members	• make available a computer in a private space • gain access to the Darkweb • get to know the forum's TOR[a] address • initial motivation to view CSEM[b]	• create nickname • register as member • gain access to public area of the forum • confirm sexual orientation • neutralize moral objections	• view, download, share CSEM	• shield forum activity • leave forum
	administrators	• find or create hosting location • get TOR address • find technically skilled co-offenders	• offer guidance and tutorials on posting	• offer guidance and tutorials on technical issues • stimulate CSEM conducive environment	• offer guidance and tutorials on security • enforce forum rules • improve forum environment
	organizational aspects	• forum marketing	• requirements for gaining full access • membership hierarchy	• organization threads, topics and subfora by content category	• forum branding • forum marketing

Figure 8: Crime Script Analysis of CSAM in Dark Web Fora (Source: Van der Bruggen and Blokland, 2021, p959)

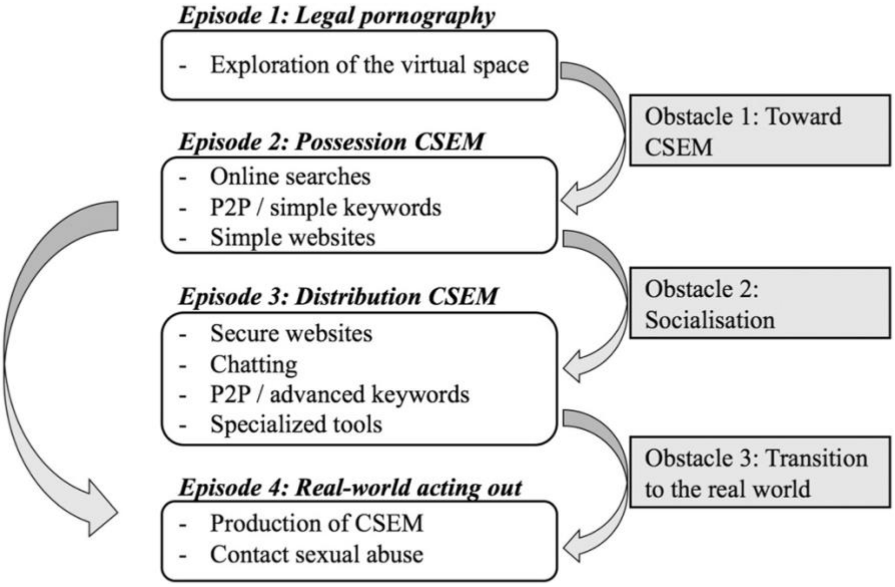

Episode 1: Legal pornography
- Exploration of the virtual space
Obstacle 1: Toward CSEM
Episode 2: Possession CSEM
- Online searches
- P2P / simple keywords
- Simple websites
Obstacle 2: Socialisation
Episode 3: Distribution CSEM
- Secure websites
- Chatting
- P2P / advanced keywords
- Specialized tools
Obstacle 3: Transition to the real world
Episode 4: Real-world acting out
- Production of CSEM
- Contact sexual abuse

Figure 9: Analysis of metascripts of consumers of child pornography (Source: Fortin et al, 2018, p35)

User studies on the effectiveness of digital forensic tools in terms of value to law enforcement are rare. One survey of digital investigators (N=153) found the main challenges are volume and limited resources, lack of up-to-date hash lists, evolution of the technological environment (cryptography, stenography), improvements in offender technology skills, lack of standardisation in legal frameworks (e.g., re COPINE levels) and managing the emotional stress of the work (Franqueria et al, 2018). The first study to address the effectiveness of digital tools in the investigation of CSAM directly with law enforcement also adopted an online survey methodology (Sanchez et al, 2019). All 106 respondents were based in the US and were digital forensic investigators (70%) or investigators (26%). This found commercial tools were used more than open source for triage and processing; for both image and video processing the top three tools selected were Cellebrite UFED/PA (17%), Magnet Forensics IEF/Axiom (16%), and Forensic Toolkit (12%). In terms of CSAM detection, of the 75% who responded to the question, 50% had used iCOP/iCAC COP, 3% had used Yahoo NSFW and 1% had used both, with the remainder not using such tools. The highest rated benefit was quickness and the ability of tools to identify explicit content with the most frequently stated limitation being that only known files can be detected (Figure 10).

Top 5 Benefits of iCOP/iCAC COP	N=42 48 responses	Top 5 Limitations of iCOP/iCAC COP	N=42 28 responses
Quickness	11 (23%)	Only known files can be identified/detected	7 (25%)

Identifying the presence of explicit content featuring children at a location or with a subject and monitoring P2P networks	10 (21%)	Can throw off hash-value comparison with file alteration	3 (11%)
Saving time	7 (15%)	False positives	2 (7%)
Filtering/filtering options to narrow down data	6 (12%)	Validation of data is still necessary	2 (7%)
Searching for known content (via hash values)	5 (10%)	Indiscriminate use of and trust in tools	2 (7%)

Figure 10: Benefits and Limitations of CSAM Detection Tools (Source: Sanchez et al, 2019)

28 suggestions were submitted for improving triage, processing and detection tools, with the top four being: Improvement of filtering mechanisms (images/videos/unrelated artifacts) (n=9), improving filtering options (n=3), automatic age recognition (n=2) and task automatization (n=2). In terms of tool development, the authors conclude that improvements are most needed in skin tone detection, automated child nudity detection, and age estimation (Sanchez et al, 2019).

Table 1: Summary of Included Research 2016 – 2021 focused on CSAM detection using computational tools and their effectiveness in prevention (N=31)

Author	Date	Subject	Model	Data	Evaluation	Application	Region1
Best-Rowden,Hoole & Jain	2016	Face recognition of very young children	Off the shelf face matcher	314 children 0-4 years over 12 months	60.4% reliability in children 3-5 (<3 45% or less)	Database of 0-4 yr olds	India
Chatzis, Panagiotopoulos & Mardiris	2016	Face recognition of very young children	New geometrical feature based on iris geometry	75 images of faces with known age: 14 <6 yrs, 24 6-12 yrs, 22 13-18 yrs, 15 >18.	Adults correct classification rate is 67%, meaning adults can be skipped	N/A	Not stated
Bissias, Levine, Liberatore, Lynn, Moore, Wallach, & Wolak	2016	Detection of known CSAM traffick in P2P	GUID & IP address - The software took hash values identifying known CEM files as input.	3 yrs of logs of public activity on BitTorrent; eDonkey, including Kad; Ares Galaxy; Gnutella; and Gnutella2.	Images validated with LE. Estimate there were about 840,000 unique installations per month, 1 in 10K internet users sharing CEM worldwide.	Not stated	International
Peersman, Schulze, Rashid, Brennan, Fischer	2016	Detection of unknown CSAM traffick in P2P	Filename classification and media classification	10,000 CSA filenames matched with 1,000,000 regular filenames from Gnutella network & 1,000,000 filenames linked to pornography from PicHunter, PornoHub, RedTube and Xvideos	Validated with LE. 92% accuracy for image & 95% accuracy for video detection. Toolkit evaluation with users positive.	iCOP Toolkit	Europe
Westlake & Bouchard	2016	Profiling CSAM networks on open web	Webcrawling & network analysis	10 networks of 300 websites each, (4,831,050 webpages) seeded from a known child sexual exploitation website	Average 'goodness of fit', across ten waves of data collection, ranged from 0.81 to 0.85.	Child Exploitation Network Extractor (CENE)	International

Author	Date	Subject	Model	Data	Evaluation	Application	Region[1]
Zambrano, Sanchez, Torres & Fuertes	2017	Chatbot to attract cyberpaedophiles	Distributed platform includes a module of attraction of paedophile interest, an intelligent engine for question answer analysis, and automatic characterization of paedophile trends.	Theoretical	Theoretical	Bothook	International/ Ecuador
Westlake, Bouchard & Frank	2017	Validity of Automated Webcrawlers as CSAM Data Collection Tools	Webcrawler: Child Exploitation Network Extractor (CENE)	30 networks surrounding seed websites. 10 begun from a CE-seed, 20 comparison networks begun from a sexuality-seed (10) or sports-seed (10) website. Data collected in 10 waves, at an interval of 42 days.	Images from all three of our database's categories were prevalent in CE-seeded networks and minimally in non-CE-seeded networks. CE related search should include a multi-criterion approach, updated regularly.	CENE	International/ Canada
Yiallourou, Demetriou & Lanitis	2017	Addressing problem of lack of datasets of CSAM for tool development	Synthetic datasets + decomposition of tasks	20 synthetic images based on features identified as appropriate/inappropriate by volunteers. Features decomposed into 3 categories (individual characteristics, setting features and interactions)	Synthetic data analysis tool tested on 200 previously unseen images downloaded from the internet -false negative error rate low.	In development	International/ Cyprus
Gangwar, Fidalgo,Alegre, González-Castro	2017	CSAM detection approaches	Evaluation of 5 pornography detection approaches: nudity detection based on skin color, image descriptor, video, deep learning, and oriented to CSA detection.	2.5K CSA files + 2.5K non-CSA	Evaluated with Spanish LE - deep learning based methods can distinguish CSA images with an accuracy of almost 88%	Various	International/ Spanish
Stylianou, Schreier, Souvenir & Pless	2017	Detecting CSAM through location	System based on features extracted from crowd sourcing and a trained neural network	>2.85 million images from >254,000 hotels across the world	Accuracy evaluated on test set of 10K hotel images from 320 randomly selected hotel classes - top-1 accuracy of 19%, top-10 accuracy of 48%, and top-100 accuracy of 80%	TraffickCam	International/ US

Lancaster University

University of BRISTOL

Author	Date	Subject	Model	Data	Evaluation	Application	Region[1]
Dalins, Wilson & Carman	2018	Content and motivations of TOR users to assist LE	Webcrawler	1,155,549 unique URLs accessed from 42,013 unique domains. 408,216 of these were Tor URLs, from 7954 Tor domains (i.e. .onion).	Approx 40% of domains not of interest to LE. child exploitation occurs in 1.75%, 25% more frequently than adult pornography	Tor-use Motivation Model (TMM)	International/ Australia
Horsman	2018	Investigation of live streamed CSA on Periscope	Forensic investigation	Periscope data/artifacts on Android and IOS	Test data (non CSAM)	Artefacts on Periscope available, inc. cached content. Comment based interactions only viewable when accessing the application live.	UK
Macedo, Costa & dos Santos	2018	CSAM database construction and CSAM detection	Stepwise detection a) pornography b) face c) age	2138 images, including 1630 people (1238 child), 508 no people, displaying nudity, sex and different body parts (5111 objects in total)	Evaluated against 2 tools currently used by Brazilian Police (NuDetective and LED (hash based model). Proposed model outperforms both tools on accuracy (79.84%) and recall (64.61%), but lower on precision.	Dataset available to other researchers by request to authors	Brazil
Vitorino, Avila, Perez & Rocha	2018	CSAM detection	Deep convolutional neural networks (CNNs)	CNN Training data set - 160,000 images labelled in 91 common object categories including 20,000 images from Microsoft Common Objects in Context (COCO) dataset.	Experiment and validation dataset - c 58,974 images, with 33,723 depicting CSAM content from Brazilian Federal Police. Accuracy – CSAM image – 86.1% CSAM video – 88% true negative (no positive data)	Not stated	International/ Brazil

Author	Date	Subject	Model	Data	Evaluation	Application	Region1
Triviño, Moreno Rodríguez, Díaz López & Gómez Már	2019	Chatbot to attract cyberpaedophiles	Artificial Conversational Entity (ACE)	ACE executed for 15 hours, gathering 320 chats on Omegle	5 chats classified as cyberpaedophile - chatbot was able to maintain long conversations without the suspect realizing its true nature	C3-Sex	International/ Columbia/ Spain
Sanchez, Grajeda, Baggili & Hall	2019	Forensic value of automated tools	Online survey	Responses of 106 investigators	75% responded to question on CSAM tools. 50% had used iCOP/iCAC COP, 3% had used Yahoo NSFW and 1% had used both	N/A	International/ US
Riesco A., Fidalgo E., Al-Nabki M.W., Jáñez-Martino F., Alegre E.	2019	Detecting CSAM in online notepad services (Pastebin)	TF-IDF encoding with Logistic Regression	17640 text samples crawled from Pastebin	Accuracy of 98.63%	PasteCC 17K	International/ Spain
Hernández-Álvarez	2019	Detecting underage sex traffick in Twitter	Naïve Bayes + SVM	55123 out 1M tweets	0.9 precision 0.88 recall with SVM	N/A	International/ Ecuador
Bursztein, Clarke, DeLaune … Thakur	2019	Longitudinal measurement of CSAM	Source, hash values, file format, distribution vector	23,494,983 NCMEC reports 1998-2017	CSAM growing – 40% occurred in 2017. CSAM is global 68% reports relate to Asia, 19% Americas, 6% Europe & 7% Africa. Solutions beyond P2P critical	N/A	International
Tariq	2019	Nudity and skin detection for combating adolescent texting	Review	45 studies retrieved since 2008	Researchers should consider addressing nudity detection at the hardware level to prevent digitization of these images before they cause harm.	N/A	International
Islam, NurMahmood, Watters & Alazab	2019	Age detection in CSAM	Convolutional Neural Network (CNN)	2000 images from NCMEC, LAG and FG NET – 1K Adult, 1K child augmented (flipping horizontally,	Validated on test set of 1200 images (97% accuracy)	Not stated	International/ Australia

				vertically and scaling) 8000 images			

Author	Date	Subject	Model	Data	Evaluation	Application	Region[1]
Lee, Ermakova, Ververis, Fabian	2020	CSAM detection	Review	21 papers included from Google Scholar, ScienceDirect, and IEEE Xplore Digital Library search	CSAM detection applications yield the best results if multiple approaches are used in combination, such as deep-learning algorithms with multi-modal image or video descriptors merged together.	N/A	International/ German
Al Nabki, Fidalgo, Alegre & Alaíz-Rodríguez	2020	CSAM detection on devices seized by LE	ML classifiers using Logistic Regression and Support Vector Machine & adapting two popular CNN models	65,351 samples (64,857 Non-CSEM and 9,330 CSEM file names)	Average class recall of 0:86 and a recall rate of 0:78	4NSEEK	International/ Spanish
Chaves, Fidalgo, Alegre, Jánez - Martino & Biswas	2020	Age detection in CSAM	Deep Expectation model	IMBD and MORPH datasets - 130000 minor and young adult images —5000 images by age	Reduced the Mean Absolute Error (MAE) from 7:26, 6:81 and 6:5 respectively, to 4:07	Part of 4NSEEK	International/ Spanish
Blanco-Medina, Fidalgo, Alegre, Alaiz-Rodríguez, Jáñez-Martino & Bonnici	2020	Forensic Text Recognition in CSAM	Combined text recognizers with rectification networks and super-resolution algorithms	648 text crops from 232 CSA images	Improved text recognition by 0.93% using only resolution-based techniques	Part of 4NSEEK	International/ Spanish

Author	Date	Subject	Model	Data	Evaluation	Application	Region[1]
Spalazzi, Paolanti, Frontoni	2021		A multi-CPU master-slave architecture designed to extract images and videos from file	N/A	Tested on 3 CSAM cases with LE. 94% time reduction, 22% false positive rate.	Digital Inspection Bag	International/ Italy
Gangwar, González-Castro, Alegre, Fidalgo	2021	CSAM detection	CNN architecture with a novel attention mechanism and metric learning	2M pornographic images and 80K child images	Test dataset comprising 1M adult porn,1M non-porn images, and 5,000 real CSA images - 92.7% accuracy	AttM-CNN	International/ Spanish
Leclerc, Drew, Holt, Cale & Singh	2021	CSAM in Darknet	Crime Script Analysis	Interviews with 29 online LE investigators	Present 3 phases of the crime script for producing and distributing CSAM on the darknet (Set Up e.g. searching for CSAM in clear web, crime completion e.g. consuming CSAM and crime continuation e.g. joining other for a).	N/A	International/ Australia
Van der Bruggen &Blokland	2021	CSAM in Darknet fora	Crime Script Analysis	Random samples of 499,110 posts and thread titles from four English-language CSEM fora (up to 100 posts per subforum)	Step-by-step description of the crime process, starting with the preparations necessary to access Darkweb CSEM fora and ending with the post activity behaviors Triangulation with police case	N/A	International/ Netherlands
Zubaidi	2021	Monitoring CSAM in Malaysia	Survey	Desk research and interviews with LE and judicial personnel		N/A	Malaysia
Pereira, Dodhia & Brown	2021	CSAM detection online	Filepath classification	1,010,000 file paths from 55,312 unique storage systems provided by Project VIC	Adversarial Examples with Random Modifications in Targeted Keywords in training data set 97% accuracy, 94% recall	Recommend model used alongside other digital forensic tools	International

What tools for the disruption of child sexual abuse media in digital environments are effective in prevention?

Evaluation of effectiveness is challenging given the range of different models and approaches. Within study evaluations of computational tools are generally reported in terms of accuracy, precision and recall. Rates vary depending on the model and focus (Figure 11).

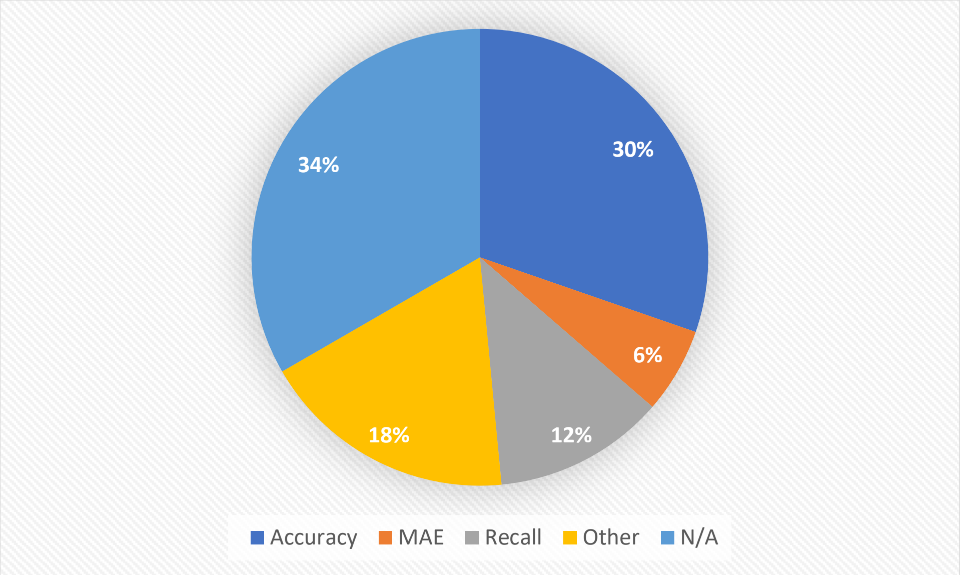


Figure 11: Evaluations of CSAM Detection Models

For over a third of the studies retrieved evaluation was not applicable, for a variety of reasons. This group includes surveys, forensic and crime script analyses, reviews, social media analyses and chatbot development where the model leads to a report of findings produced by or about computational tools, rather than an assessment of their validity. Where tools were developed specifically for the detection of CSAM, either directly or indirectly through age and nudity detection tools, accuracy was the most commonly reported assessment criterion. This ranged from 60% for the detection of young children (Best-Rowden, Hoole and Jainto, 2016) to 97% for filepath analysis (Pereira, Dodhia, and Brown, 2021). Recall was also reported in some studies ranging from 65% for a stepwise law enforcement processing model (Macedo, Costa, and.dos Santos, 2018) to 94% for filepath analysis (Pereira, Dodhia, and Brown, 2021). Other methods included calculating Mean Average Error (for age estimation and CSAM severity detection), Goodness of Fit (network analysis) and estimation of reduction in law enforcement processing time (94% compared with other models (Spalazzi, Paolanti and Frontoni, 2021)).

From a wider view, there is little evidence that computational tools aimed at preventing CSAM have been effective. Despite concerted research activity, several reports indicate that CSAM has increased in volume and severity over time (IWF, NCMEC). Given the difficulties of accessing content for research purposes, only one study has conducted a longitudinal analysis outside clearing house reports. Fortin and Proul (2019) analysed CSAM content on 112 hard drives of 40 convicted CSAM offenders across 8 years (2004-2012) using the Forensic Tool Kit (FTK). Raters were trained on age estimation and the COPINE scale. A sample of over 61K images extracted from over 5M CSAM were coded, with a mean of 1,038 per offender. They identify 4 patterns over time, only one of which depicted a reduction in severity. 37.5% of the sample were categorised as having a ‘Degenerating Spiral Pattern’ where COPINE levels increased from 4 (suggestive images) to 6 (explicit sexual images) and age of interest reduced from a mean of 12.2 to a mean of 6.9 years. The ‘Sexualized Adolescent Pattern’ (20%) depicted a similar increase in COPINE score (from 4 to 6) and an increase in age of subjects (from 13 to 14). A pattern described as ‘Boy/Girl-Love’ shows the COPINE level dropping from 7-6 and the age of subject reducing from 17 to 7. A ‘De-Escalation’ pattern, where COPINE levels reduced from 7 to 5-6 and ages increased from 11-14 years was found in 22.5% of the sample. Overall, the most frequent age of subject was 6-12 years and the mean COPINE score was 5-6 (68%).

This apparent failure in CSAM prevention “stems from the fact that the development of detection and processing technologies substantially lags behind the development of content creation and sharing capabilities” (Bursztein et al, 2019, p2605). The crime script analysis reveals that motivations to increase CSAM production may be strengthened by its online nature where it is easier to find reinforcement for paedophile tendencies, and where the structure of forums encourages the sharing of more and more extreme material. Given the inherent limitations of law enforcement resource and capacity it would suggest that computational efforts should prioritise reliable and accurate processing technologies alongside improving detection methods.

LE receive reports from various sources and are generally overwhelmed with the volume of CSAM material that must be processed. Content on seized devices must be assessed manually supported by automated tools. The task is threefold; age verification of the person(s) detected in the image, identifying the child sexual abuse depicted and matching with legal requirements, including the amount of time allowed. Any sexual acts depicted must be evaluated against a range of complex legal provisions. In the UK, for example, analysts must identify whether the image or video depicts a child under 18 and whether the activity is of an ‘indecent’ nature, meeting the requirements of UK sentencing law. A pilot study by Kloess, Woodhams and Hamilton-Gilchrist (2021) finds that this is less than straightforward with inter-rater agreement varying, particularly with respect to age verification in older adolescence and offence level. This work is time consuming and demanding across all jurisdictions. Previous research has focused on psychological effects of investigating CSAM (see for example, Bourke & Craun, 2013), identifying a third of investigators (n=600) experienced high levels of secondary trauma and half reported lower levels (Powell et al, 2015) .

A key issue for LE in the adoption and use of automated CSAM tools is their impact on evidence for prosecution. In Europe such evidence must be explainable and processes should be documented step by step. Following implementation of GDPR in 2018 all private subjects are entitled to 'human produced explanation of any AI based algorithms' (Raaijmakers, 2019). It is also clear that, notwithstanding tool transparency, laws in all countries have failed to effectively adapt to the use of computational tools as valid evidential tools. Experiences and learning from the use of DNA evidence offer a potential way forward but it is clear that accuracy still needs to improve before the room for reasonable doubt is reduced to a level that minimizes the need for manual inspection. Rapid improvement in explainability and transparency will possibly assist with this, although no study was found that addresses this in relation to CSAM.

Part II: CSAM in ASEAN countries

Website Search on CSAM in ASEAN

Owing to the lack of primary research over the last five years in the region a webpage search was delivered to gather more background information on ASEAN countries. Using country-specific google search domains, for example, https://www.google.com.ph for Philippines, we searched the key words ‘‘online child sexual abuse’’ to collect relevant reports, news and blogs dated after 2016, from 10 ASEAN countries’ government websites and 17 organizations, including UNICEF, ECPAT International, WeProtect Global Alliance and Interpol [see Appendix *]. This background search resulted in 197 outputs retrieved, which were stored as .pdf files.

A preliminary analysis of these results checked their relevance against our search strings using *AntConc*, a corpus analysis toolkit [Anthony, 2020]. Content was first filtered for context within a 100-word range of the 8 search strings separately to discover the most relevant information. A collocate search for each search string was followed to explore the frequencies of those words that appeared more than once and within a 10-word range (5 words left from the key words and 5 words right) of the search strings. The ‘Concordance Plot’ function was applied to map the locations of these key contexts in files that identified one or more of the search strings more than 20 times. Both the concordance and collocate searches and plots helped to highlight the most relevant outputs for full text review (N=24). Findings from the web site review were then analysed manually for themes.

Although the several research papers retrieved for abstract and title screening did not meet inclusion criteria, primarily because their focus was not specifically on CSAM and digital forensics tools, they do provide additional knowledge about CSAM in the region and so are also reported under the themes.

Key CSAM Themes in ASEAN

Theme 1: CSAM is Increasing in the Region

Clearing House data finds that three ASEAN account for over a quarter of all CSAM reports worldwide; Indonesia (11%, 1.7M), Thailand (11%, 1.7M) and Vietnam (4%, 0.7M) (Bursztein et al, 2019, p2605). A survey with 5,302 18-20 year olds across 54 countries, conducted in 2021, found 52% of boys and girls under 18 had experienced at least one online sexual harm in Southeast Asia (ref?).

Several sources note the increase in CSAM during the global Covid-19 pandemic (IOCTA, WeProtect Global Alliance, 2021). Data published by NCMEC for 2020 clearly shows a 47.8% rise across ASEAN (Table 2), exceeding 100% in Cambodia and Vietnam, and a two thirds increase in the Philippines, continuing the increase reported by Peersman (2020).

Country	*2019*	*2020*	*% increase*
Brunei	2,070	2,119	2.4
Cambodia	91,458	188,328	106
Indonesia	840,221	986,648	17.4
Lao PDR	23,599	24,404	3.4
Malaysia	183,407	204,506	11.5
Myanmar	233,681	319,617	36.7
Philippines	801,272	1,339,597	67.1
Singapore	18,426	22,769	23.5
Thailand	355,396	397,743	11.9
Vietnam	379,554	843,963	122.4
Total	2,929,084	4,329,694	47.8
NCMEC Total	16,987,361	21,751,085	28.0

TABLE 2: NCMEC'S CYBER TIPLINE REPORTS SENT TO ASEAN LAW ENFORCEMENT AGENCIES FOR REVIEW BETWEEN 2019 AND 2020. SOURCE: NCMEC.

Live streaming and 'capping' are reported to have increased in the region during lockdown from March to May 2020, with further evidence that demand is coming from Europe and the global North. In the majority of cases (69%) it is financially motivated, facilitated by adult female relatives close to the victim (IJM, 2020 – cited in WeProtect, 2021). Self-generated images are an increasing source of CSAM across the world. These images may be coerced or produced without the victim's knowledge or understanding. They may also be used for commercial ends by children in ASEAN. For example, WeProtect (2021) report a study in Cambodia where girls use their own sexual media to sell beauty products online.

Theme 2: There is a partial understanding of the nature and extent of CSAM in the region

Data on ASEAN has to be extracted from several sources. Some sources are country specific, such as the ECPAT country reports, others refer to South Asia or Asia and the Pacific which may include part, or all, ASEAN. Coverage of CSAM in the Philippines, Thailand, Vietnam and Cambodia is stronger than in other ASEAN. This is particularly relevant when seeking data on law enforcement and responses to CSAM.

Data on seven ASEAN countries extracted from a 60-country benchmarking exercise (Economist Intelligences Unit, 2019) reveals varying levels of engagement with CSAM across the region (Table 3). With the exception of Cambodia and the Philippines, levels of data collection relating to prevalence were assessed as non-existent. Victim support initiatives exist in six countries, but offender support which might considered key to prevention, was only observed in Thailand. Tech engagement ranged from none to 100% (Malaysia).

Country	Victim Support	Offender Support	Data Collection	Tech Engagement
Cambodia	80.0	0.0	76.9	0.0
Indonesia	12.0	0.0	0.0	0.0
Malaysia	20.0	0.0	0.0	100
Myanmar	0.0	0.0	0.0	0.0
Philippines	40.0	0.0	38.5	66.7
Thailand	32.0	50.0	0.0	33.3
Vietnam	20.0	0.0	0.0	33.3

Table 3: 2019 Benchmark Index on CSAE Response for ASEAN. Source: Economist Intelligences Unit.

An overview of Sexual Exploitation of Children (SEC) in ASEAN is provided by Davy (2017). This shows how offline SEC facilitated by the trafficking of women and girls for the purposes of sexual exploitation and sex tourism often has an online dimension, and thus connects with CSAM. Girls and boys are recruited and sold online, creating enduring evidence of their sexual exploitation.

A rich understanding of CSAM production in the Philippines is offered by Ramiro et al (2019) who conducted an ethnographic case study in two barangays (administrative districts) in Manila identified as ‘hot spots’ for online CSA and CSE. Data was collected from 144 systematically sampled key informants, 12 from each of 3 age groups from each area (13-17, 18-24, 25-50) plus 6 community members (community leaders, LE and money exchange personnel). CSAM is reportedly produced as part of a process of sexual exploitation. Children are often introduced into online sex work through their peer network. Nude and sexually explicit photos are created to both entice buyers of sex or in response to requests from buyers. This then leads to ‘consumation’ either by a ‘foreigner’ or local men who meet the child for sex. Images and videos may be taken by the perpetrator and shared with others.

Although not presented as such, processes similar to stages in crime script analysis are described, depicting how children become involved, patterns of perpetrator contact, and intermediary roles. Payment or ‘sponsorship’ features in all five processes (Figure 12).

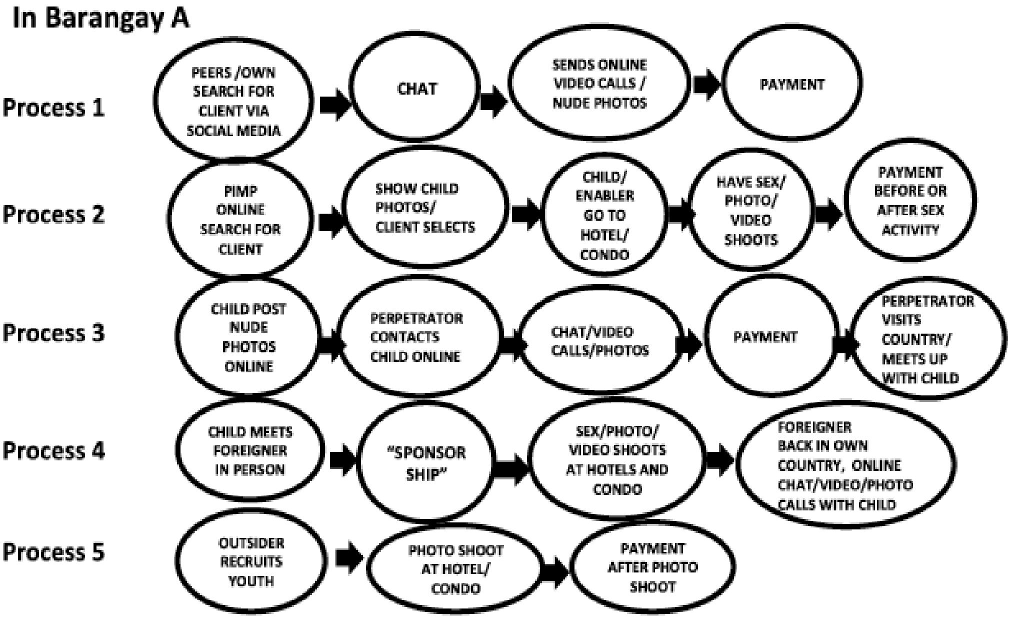

In Barangay A
Process 1
PEERS /OWN SEARCH FOR CLIENT VIA SOCIAL MEDIA
CHAT
SENDS ONLINE VIDEO CALLS / NUDE PHOTOS
PAYMENT
Process 2
PIMP ONLINE SEARCH FOR CLIENT
SHOW CHILD PHOTOS/ CLIENT SELECTS
CHILD/ ENABLER GO TO HOTEL/ CONDO
HAVE SEX/ PHOTO/ VIDEO SHOOTS
PAYMENT BEFORE OR AFTER SEX ACTIVITY
Process 3
CHILD POST NUDE PHOTOS ONLINE
PERPETRATOR CONTACTS CHILD ONLINE
CHAT/VIDEO CALLS/PHOTOS
PAYMENT
PERPETRATOR VISITS COUNTRY/ MEETS UP WITH CHILD
Process 4
CHILD MEETS FOREIGNER IN PERSON
"SPONSOR SHIP"
SEX/PHOTO/ VIDEO SHOOTS AT HOTELS AND CONDO
FOREIGNER BACK IN OWN COUNTRY, ONLINE CHAT/VIDEO/PHOTO CALLS WITH CHILD
Process 5
OUTSIDER RECRUITS YOUTH
PHOTO SHOOT AT HOTEL/ CONDO
PAYMENT AFTER PHOTO SHOOT

Figure 12: The Online CSE/A process as reported in one barangay in Manila (Source: Ramiro et al, 2019)

Drawing on social norm theory, the researchers claim the children are led by ‘custom’ norms; those that meet needs and desires for consumer goods, to give money to their parents, or to get a better life by marrying a foreigner. All informants agreed that CSAM and CSE involving children aged 12 or under was ‘wrong’, but for those aged 13-17 it was considered that at that age they had a better understanding of what they were doing. The main psychological consequences noted were loss of self-esteem and shame, both for the child and their family. However, community members rarely report to the police seeing it as a bad thing to do:

‘[When you report], you are the bad person to them. So just keep still. Like just be blind. It is not your life that gets destroyed anyway. It is their life. It may just be your opinion that what they are doing is wrong’. (Male, 18–24, Brgy. B) ibid p10.

The researchers note that reporting is a challenge in the Philippines due to a ‘culture of tolerance or silence’…the social stigma associated with being a sexual victim, and a lack of trust in authorities’ (p11) along with the Philippines’ legal age of consent being 12 years.

Theme 3: Data on responses to CSAM are limited

From a global perspective, We Protect (2021) maintain that deterrence measures operated by tech companies are helpful. Recent initiatives include warning messages, operated by approximately 60% of platform providers, which have reduced search activity on these platforms. Some initiatives direct CSAM searchers to help seeking initiatives but it is not clear how many (if any) are tailored towards ASEAN offenders given the lack of offender support in the region (Table 3).

The Indonesian government has addressed access negative content by publishing a list called TRUST+™ Positif which includes websites with radicalisation, gambling, drugs and pornography content, including CSAM. Pornography websites are the largest category. All internet service providers are required to block these sites. The list is populated through external reports and back engine crawling on these reported sites. Aristofany, Saptawti and Asnar (2018) note the limitations of this approach, specifically that negative content websites remain undetected. They test a data mining approach using an association class algorithm (Apriori) to overcome this limitation but found this algorithm yielded a high number of false positive results.

Merton (2020) provides a brief report focused on CSE and CSAM production in the Philippines, where it is claimed that up to 95% of victims do not receive support. The article references the CURE Foundation, which provides support for victims in Cebu. Rehabilitation is noted as difficult because "of the co-offenders, 82% are parents or immediate family members. Some 18% are neighbours or members of the community from which the girls come" p747. Victims have signs of trauma; headaches, anxiety attacks, self-harm, memory lapses., low self-worth, depression, and sexualised behaviour. Following 12-20 therapy sessions UNICEF and other NGOs are working to find foster placements for victims and support reintegration. Training for law enforcement and prosecution personnel, and school-based education programmes are seen as a way forward. These findings stand in contrast to those of Ramiro et al which suggested facilitators were more often peers or 'go-betweens' rather than family members and that psychological or physical trauma responses in victims were not always evident. This may be due to the different sample of victims being studied; young people (male and female) from Barangays in Manila some of whom were victims of CSE and some who were not, and girls who were victims of CSE/A from Cebu.

In related work, finding that 14% of adolescent girls in Asia and the Pacific report forced sexual debut, Yount et al (2020) report on a protocol for adapting and applying a sexual consent awareness programme (Global Consent), currently the subject of a Randomised Control Trial in Vietnam. This, and other studies (see Lovelle et al, 2017), point to the fact that most of the research on offending behaviour, including theories of such, derives from the global north and may not straightforwardly apply to the very different cultural contexts of ASEAN countries.

Theme 3: The Legal Context in ASEAN is Challenging

Current legislation, and the way it is implemented may undermine the gravity of the offence and fails to be a sufficient deterrent in ASEAN countries. Analyses of offender ethnicity, such as in the ICSE database (ECPAT and INTERPOL, 2018), which find that only 3.2% are of Asian ethnicity, is likely to be a reflection of levels of awareness, reporting and prosecution, rather than the true distribution. For example, in the Manila Baringays, Ramiro et al (2019) found perpetrators were both foreign and Filipino, and those not from the Philippines were said to include Japanese, Indians, South Koreans and ‘Arabs’ along with those with Caucasian features. Furthermore, an ECPAT review of the region notes that the ‘vast majority of of child sex offenders in Southeast Asia are nationals of the countries of the region’ (Davy, 2017, p16) and that girls are the primary victims. It also identifies inefficient criminal justice and child protections systems, along with reports of police corruption as impediments to prosecution. Improvements to law enforcement responses were reported in 2017, mainly through the creation of cybercrime units (Philippines, Thailand, Malaysia, Singapore, Vietnam and Cambodia). These efforts have improved information exchange both across ASEAN, through ASEANAPOL and internationally via INTERPOL.

Considerable advances have been made in the region over the last two years at the policy level. In 2019 the ASEAN Secretariat, published a mid-term review of the ASEAN Regional Plan of Action on the Elimination of Violence against Children 2016-2025 (ASEAN, 2019). This reports that the ASEAN Regional Dialogue on Child Online Sexual Exploitation in 2018 enabled significant progress to be made on establishing minimum legal standards and along with production of an ASEAN Declaration on Child Sexual Abuse and Exploitation Online. Cambodia, Indonesia, the Philippines, Thailand and Viet Nam are members of WeProtect Global Alliance and have committed to implementing the WeProtect Model National Response (WeProtect, 2016)

These regional advances represent a major step forward in prosecuting and thus preventing CSAM. However, differences in legislation and cultural attitudes influence the application of law and policy. WeProtect Global Alliance emphasised in 2019 that 'stark differences between the Global North and South are creating a worrying global discord' p34. This includes inconsistencies in legislation, which are apparent across the ASEAN region. Despite most countries ratifying the UN CRC which sets the age of a child as up to 18, laws regarding consent and sexual relations vary widely (Kokolaki et al., 2020). Across ASEAN, the age of consent for both sexes ranges from 12 to 18 (see Appendix B), and any sexual acts depicted must be evaluated against a range of complex legal provisions. As in the rest of the world, this is time consuming and demanding.

An illustration of the legal challenges is provided by Zubaidi (2021) who reviews legislation and practices in combatting CSAM in Malaysia. Historically, responses to CSAM and CSA were characterised by a 'lack of urgency' by both government and LE. She reports that this changed in 2018 with the introduction of Malaysia Internet Crime Against Children Investigation Unit (MICAC) and the adoption of the Internet Crime Against Children: Child Online Protective Services (ICACCOPS) monitoring software. In 2017 when the interviews were conducted, there were reports of limitations in the use of the software due to Data Protection legislation (concerning crawling the open web and image identification), dependence on IP addresses which may have multiple users, technical expertise of investigators and lack of resources. As one police officer noted:

'Under Section 40 (1) of the Data Protection Act 2010, a data user shall not process "any sensitive personal data". Sensitive personal data includes information on matters such as health or any other personal data as determined by the relevant Minister. This would include an individual's private communications held on his personal devices' (Zubaidi, 2021, p199).

In this context, practices vary between regions and prioritisation of crimes that are perceived to be more serious (such as murder), with more chance of prosecution;

"Law enforcement agencies may be more willing to pursue cases if clear digital evidence of the crime confirms its severity, and this increases the likelihood of a successful prosecution. In determining the severity of an image, the authorities are required to determine: whether actual children are depicted; the identity of the children; and their location in anticipation of future testimony" (Zubaidi, 2021, p195).

These evidential requirements are challenging, particularly proving whether a child is depicted in the images and LE may not follow up on cases where age verification is problematic, such as images depicting children in late adolescence. Zubaidi notes that

MCMC and CSM (Cyber Security Malaysia) have developed software (Prototype A) to identify suspects in surveillance videos which could also be applied to identify victims in seized images but its status is not clear.

Discussion

Outcome 1: Assessment of how tools to prevent the production and dissemination of CSAM can be expanded.

The continued growth in CSAM, markedly during the pandemic, suggests that despite all efforts, technical and social, the prevention of CSAM is at best hypothetical. The expansion of CSAM over the last five years is facilitated by two trends that can be concluded from this review: firstly, the separate trajectories of research, child protection knowledge and approaches (largely covered by NGO and policy reports, such as those by Ecpat, WeProtect and so forth) and technical research directed at detecting CSAM (reported in the academic international peer review literature) and secondly a disconnection between academic and private sector research in this field.

With the exception of the Bracket Foundation (2020), NGO and policy reports focus on the problem of online sexual abuse of children, highlighting its characteristics and prevalence as an industry, an organised crime enterprise, a product of sex tourism, the role of self-generated and social media, the trauma and impact on children, and so forth. These are the contexts in which CSAM is produced with often serious consequences for the children involved. The call is generally for integrated policies and responses, at a national and regional level, in the hope that these will deter perpetrators and assist victims. The technical literature reports on a range of different models to address CSAM detection, including webcrawlers, filename and filepath analyses, chatbots and various age and skin detection techniques, often paying little attention to the application of tools by law enforcement and the real-world challenges of evidence gathering and prosecution. As accuracy, precision and recall in the detection of CSAM are greatly improved by deep learning models, the massive scale of retrieval presents law enforcement and clearing houses such as NCMEC and IWF with a significant challenge; how to manage the volume of CSAM? The Bracket Foundation (2020), Burszstein et al (2019), and Sanchez et al (2019) recommend an emphasis on improving computational approaches to the law enforcement processing task. These technologies must address the enduring problem that processing must also fit with evidential requirements, which vary across the world. In this context, the recent focus (particularly in Europe following implementation of GDPR) on improving transparency and explainability, presents a well-known dilemma in the field: how to make the tools effective and at the same time not report sufficient detail that perpetrators find ways to circumvent them. What may be required, and seems to be lacking, is an agreed international standard for CSAM tool evaluation for evidential purposes that is shared by child protection organisations, the private sector and researchers.

Secondly, almost all the tools in use (such as those identified by the Bracket Foundation, 2020) are commercialised. Whilst they may or may not be effective, evaluations of effectiveness are a) not provided and b) effective for what is not clear. If it is the detection of CSAM, this in itself will not effectively prevent the production of CSAM or further victimisation; partly for resource reasons (as noted above) and partly for displacement reasons. Internet service providers and platforms may manage to reduce CSAM on their sites by using these tools, but the sheer volume reported to law enforcement means investigation cannot keep up. Further, as Lee et al (2020) observe, even where platforms such as Google disrupt and deter CSAM, perpetrator activity is merely displaced to sites and jurisdictions where no such deterrence is in place. Dark web fora, along with P2P and bulletin boards, provide an unprecedented nurturing and supportive environment for creating, sharing and disseminating CSAM.

There is no straightforward solution to this seemingly intransigent problem but a greater awareness in the child protection field of computational tool development, and a reciprocal growth in awareness in the cybercrime/computational tool development field of the practicalities of child protection would undoubtably help.

Outcome 2: What learning can be obtained from 1) in respect of ASEAN countries?

Despite a number of efforts by international organisations such as INTERPOL, UNICEF, Ecpat and EVAC working alongside state parties in ASEAN, CSAM originating from the region continues to grow. The general conclusions from the technical review above apply to ASEAN, but with particular challenges specific to the region.

CSAM in ASEAN is more of a commercial activity than it is in some other regions, enhanced by significant poverty, trafficking and sex tourism. This is known to be connected to live streaming, and few computational tools have been developed to date to disrupt or otherwise prevent live streaming. The review found live streams can be capped and are shared in P2P networks and Tor forums. In addition, forensic analysis of devices can detect live stream artifacts, dependent on platform, although these only provide meta-data that could corroborate other sources of evidence. Evidence can be captured in real-time, but this would require victims of CSA collaborating with law enforcement in the region.

Whilst the main language of CSAM in the region appears to be English, this may be because it is the most commonly detected language through existing computational tools. The ASEAN region hosts many different language speakers, including those in residence. Limited research on child sexual abuse and CSAM in the region finds that perpetrators of child sexual abuse can be both foreign and local, and that many countries and languages are represented. This will be incorporated as far as is possible into the ICOP project offering the potential to detect unknown CSAM originating in ASEAN and CSAM user countries, such as Japan, China, India, South Korea and Russia, along with more familiar language speakers from North America, Australia, the UK and Europe.

With the exception of Zubaidi (2021) no NGO reports or research were found to offer data on the development or application of computational tools to address CSAM in the region (either for detection or processing).

In the absence of reliable ASEAN data, the following key questions have emerged from the review:

1. How can a consensus be established amongst law enforcement in ASEAN to improve detection of CSAM?
2. What tools are currently in use by law enforcement in ASEAN and how effective are they in terms of the law enforcement task?
3. Given the estimated prevalence of CSAM emerging from the region, how can the negative consequences of scaling up victim identification be mitigated?

In terms of a legal consensus, whilst controversial, one proposal by INTERPOL (although not specific to ASEAN) is that it could be formed around a baseline CSAM definition that includes:

• the victim is a real child;

• the victim is pre-pubescent, or in the very first signs of puberty (typically under 13 years old);

• the imagery conveys either:

> – sexual activity of the child, with the child, in the presence of a child, between children;
>
> or

– a focus on the vagina, penis or anal region of the child;
and

- the image is verified by several specialists from different countries (WeProtect, 2019).

Such a definition provides a framework for prioritising features that could be identified with the latest generation of computational tools that have the potential to encourage broader uptake by LE in ASEAN. Further research is required to identify tools in use in the region and further information of the applicability of tools for processing and evidential purposes. This will form the next stage of the study.

References

Kemal Veli Acar. 2017. Webcam Child Prostitution: An Exploration of Current and Futuristic Methods of Detection". International Journal of Cyber Criminology 11.1, 98–109.

Mhd Wesam Al Nabki, Eduardo Fidalgo, Enrique Alegre, Rocio Alaíz-Rodríguez. 2020. File Name Classification Approach to Identify Child Sexual Abuse. In *ICPRAM*, 228-234.

Laurence Anthony. 2020. AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Army Aristofany, Gusti Ayu Putri Saptawati, Yudistira Asnar. 2018. Internet browsing history data analysis for automatic negative content website identification (Case Study: TRUST+™ Positif). In *2018 5th International Conference on Data and Software Engineering (ICoDSE)* (1-6). IEEE.

Ali Al Zahrani, Mohamad Ahtisham Wani, Wasim Ahmad Bhat. 2021. Forensic Analysis of Twitch video streaming activities on Android. *Journal of Forensic Sciences*, 66, 1721-1741.

Edo Barfian, Bambang Heru Iswanto, Sani Muhamad Isa. 2017. Twitter Pornography Multilingual Content Identification Based on Machine Learning, *Procedia Computer Science*, 116, 129-136.

Guru Swaroop Bennabhaktula, Enrique Alegre, Dimka Karastoyanova, George Azzopardi. 2020. Device-based Image Matching with Similarity Learning by Convolutional Neural Networks that Exploit the Underlying Camera Sensor Pattern Noise. *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods – Vol 1: ICPRAM*, 578-584. DOI: 10.5220/0009155505780584

L. Best-Rowden, Y. Hoole and A. Jain. 2016. Automatic Face Recognition of Newborns, Infants, and Toddlers: A Longitudinal Evaluation, *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 1-8. DOI: 10.1109/BIOSIG.2016.7736912.

Bissias, George Dean et al. 2016. Characterization of contact offenders and child exploitation material trafficking on five peer-to-peer networks. *Child Abuse & Neglect*, 52, 185-99.

Blanco-Medina, P., Fidalgo, E., Alegre, E., Alaiz-Rodríguez, R., Jáñez-Martino, F. and Bonnici, A. 2020. Rectification and Super-Resolution Enhancements for Forensic Text Recognition. *Sensors*, *20*, 5850.

Bourke, M.L., Craun, S.W. 2013. Secondary traumatic stress among internet crimes against children task force personnel: impact, risk factors, and coping strategies. *Sexual Abuse J. Res. Treat*. 26, 586-609

Bromley S.T., Sheppard J., Scanlon M., Le-Khac NA. 2021. Retracing the Flow of the Stream: Investigating Kodi Streaming Services. In: Goel S., Gladyshev P., Johnson D., Pourzandi M., Majumdar S. (eds) *Digital Forensics and Cyber Crime. ICDF2C 2020.*

Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 351. Springer, Cham. https://doi.org/10.1007/978-3-030-68734-2_13

Elie Bursztein, Einat Clarke, Michelle DeLaune, David M. Elifff, Nick Hsu, Lindsey Olson, John Shehan, Madhukar Thakur, Kurt Thomas, and Travis Bright. 2019. Rethinking the Detection of Child Sexual Abuse Imagery on the Internet. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 2601–2607. DOI:https://doi.org/10.1145/3308558.3313482

V. Chatzis, F. Panagiotopoulos and V. Mardiris. 2016. Face to Iris Area Ratio as a feature for children detection in digital forensics applications, *Digital Media Industry & Academic Forum (DMIAF)*, 121-124, DOI: 10.1109/DMIAF.2016.7574915.

Deisy Chaves, Eduardo Fidalgo, Enrique Alegre, Rocío Alaiz-Rodríguez, Francisco Jáñez-Martino, George Azzopardi. 2020. Assessment and Estimation of Face Detection Performance Based on Deep Learning for Forensic Applications. *Sensors*, 20, 4491; doi:10.3390/s20164491

Deisy Chaves, Eduardo Fidalgo, Enrique Alegre, Francisco Jánez -Martino, Rubel Biswas. 2020. Improving age estimation in minors and young adults with occluded faces to fight against child sexual exploitation. In *VISIGRAPP (5: VISAPP)*, 721-729.

Janis Dalins, Campbell Wilson, Mark Carman. 2018. Criminal motivation on the dark web: A categorisation model for law enforcement, *Digital Investigation*, 24, 62-71. https://doi.org/10.1016/j.diin.2017.12.003.

Deanna Davy. 2017. *Regional Overview: The Sexual Exploitation of Children in South East Asia*. Bangkok, Thailand: Ecpat International.

Economist Intelligence Unit. 2020. *Out of the Shadows: Shining the Light on the Response to Child Sexual Abuse and Exploitation.* The Economist Intelligence Unit Limited.

ECPAT International. 2017. Online child sexual exploitation: An analysis of emerging and selected issues. *ECPAT International Journal* 12: 1–63

Europol. 2019. *Internet organised crime threat assessment 2019*. The Hague: Europol

Francis Fortin, Sarah Paquette, Benoit Dupont. 2018. From online to offline sexual offending: Episodes and obstacles, *Aggression and Violent Behavior*, 39, 33-41. https://doi.org/10.1016/j.avb.2018.01.003.

Francis Fortin and Jean Proul. 2019. Sexual Interests of Child Sexual Exploitation Material (CSEM) Consumers: Four Patterns of Severity Over Time, *International Journal of Offender Therapy and Comparative Criminology*, 63(1) 55–76.

Virginia Franqueria, Joanne Bryce, Noora Al Mutawa, Andrew Marrington. 2018. Investigation Indecent Images of Children cases: Challenges and suggestions collected from the trenches, *Digital Investigation*, 24, 95-105.

Abhishek Gangwar, Eduardo Fidalgo, Enrique Alegre and Víctor González-Castro. 2017. "Pornography and child sexual abuse detection in image and video: A comparative evaluation," *8th International Conference on Imaging for Crime Detection and Prevention (ICDP 2017)*, 2017, 37-42, doi: 10.1049/ic.2017.0046.

Abhishek Gangwar, Víctor González-Castro, Enrique Alegre, Eduardo Fidalgo. 2021. AttM-CNN: Attention and metric learning based CNN for pornography, age and Child Sexual Abuse (CSA) Detection in images, *Neurocomputing* 445 (2021) 81–104.

Global Alliance against Child Sexual Abuse Online (GACSAO). 2016. *2015 Threat assessment report.* Global Alliance against Child Sexual Abuse Online.

Henseler Hans, Rens de Wolf. 2019. Sweetie 2.0 Technology: Technical Challenges of Making the Sweetie 2.0 Chatbot. In: van der Hof S., Georgieva I., Schermer B., Koops BJ. (eds) *Sweetie 2.0. Information Technology and Law Series*, vol 31. T.M.C. Asser Press, The Hague. https://doi.org/10.1007/978-94-6265-288-0_3.

M. Hernández-Álvarez. 2019. "Detection of Possible Human Trafficking in Twitter", *International Conference on Information Systems and Software Technologies (ICI2ST)*, 2019, pp. 187-191, doi: 10.1109/ICI2ST.2019.00034.

Graeme Horsman. 2018. A forensic examination of the technical and legal challenges surrounding the investigation of child abuse on live streaming platforms: A case study on Periscope, *Journal of Information Security and Applications*, 42, pp 107-117. https://doi.org/10.1016/j.jisa.2018.07.009.

International Centre for Missing & Exploited Children (ICMEC). 2018. *Child sexual abuse material: model legislation & global review*. https://www.icmec.org/wpcontent/uploads/2018/12/CSAM-Model-Law-9th-Ed-FINAL-12-3-18.pdf.

IJM, 2020 Technical and Financial Sector Indicators of Livestreaming. Shared by IJM, 11/03/2021 In We Protect Global Alliance. 2021. *Global Threat Assessment 2021*. https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2021.pdf

Internet Watch Foundation (IWF). 2018. *Trends in child sexual exploitation: Examining the distribution of captures of live-streamed child sexual abuse*. Cambridge, UK: Internet Watch Foundation.

Islam M., Mahmood A.N., Watters P., Alazab M. 2019. Forensic Detection of Child Exploitation Material Using Deep Learning. In: Alazab M., Tang M. (eds) *Deep Learning Applications for Cyber Security. Advanced Sciences and Technologies for Security Applications*. Springer, Cham. https://doi.org/10.1007/978-3-030-13057-2_10.

Emmanouela Kokolaki, Evangelia Daskalaki, Katerina Psaroudaki, Meltini Christodoulaki, Paraskevi Fragopoulou. 2020. Investigating the dynamics of illegal online activity: The power of reporting, dark web, and related legislation. *Computer Law & Security Review* 38 (2020) https://doi.org/10.1016/j.clsr.2020.105440.

Juliane Kloess, Jessica Woodhams, Cathering Hamilton-Giachritsis. 2021. The challenges of identifying and classifying child sexual exploitation material: Moving towards a more ecologically valid pilot study with digital forensics analysts. *Child Abuse and Neglect*, 118 (2021) https://doi.org/10.1016/j.chiabu.2021.105166.

Leclerc, B., Drew, J., Holt, T., Cale, J. and Singh, S. 2021. Child sexual abuse material on the darknet: a script analysis of how offenders operate, *Trends and issues in crime and criminal justice*, no. 627, Australian Institute of Criminology.

Hee-Eun Lee, Tatiana Ermakova, Vasilis Ververis, Benjamin Fabian. 2020. Detecting child sexual abuse material: A comprehensive survey. *Forensic Science International: Digital Investigation*, 34, 301022.

María Inés Lovelle; Montserrat Yepes-Baldó; Marina Romeo; Miguel Ángel Soria. 2017. Online Child Pornography: A Cultural Focus, *Acción Psicológica*, 14, 2, 99-112. https://doi.org/10.5944/ap.14.2.20766

Jo˜ao Macedo, Filipe Costa, Jefersson A. dos Santos. 2018. "A Benchmark Methodology for Child Pornography Detection," *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 455-462, doi: 10.1109/SIBGRAPI.2018.00065.

Sheri Madigan, Anh Ly, Christina Rash, Joris Van Ouystel, Jeff Temple. 2018. Prevalence of Multiple Forms of Sexting Behavior Among Youth: A Systematic Review and Meta-analysis, *Jama Pediatrics*, 172 (4), 327-335.

Felix Mayer and Martin Steinebach. 2017. Forensic Image Inspection Assisted by Deep Learning. In *Proceedings of the 12th International Conference on Availability, Reliability and Security (ARES '17).* Association for Computing Machinery, New York, Article 53, 1-9. DOI: https://doi.org/10.1145/3098954.3104051.

Martina Merten. 2020. Tackling online child sexual abuse in the Philippines. *The Lancet*, 396, 747-8.

Mourad Ouzzani, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. 2016. Rayyan — a web and mobile app for systematic reviews. *Systematic Reviews*, 5:210, DOI: 10.1186/s13643-016-0384-4.

Allison Parks, Charlotte Sparre, Elin Soderquist, Stefan Arver, Gerhard Andersson, Viktor Kaldo, Katarina Gorts-Oberg, Christoffer Rahm. 2020. Illegal Online Sexual Behavior During the COVID-19 Pandemic: A Call for Action Based on Experiences from the Ongoing Prevent It Research Study. *Archives of Sexual Behavior*, 49, 1433-1435.

Claudia Peersman, Christian Schulze, Awais Rashid, Margaret Brennan, Carl Fischer. 2016. iCOP: Live forensics to reveal previously unknown criminal media on P2P networks, *Digital Investigation*, 18, 50-64. https://doi.org/10.1016/j.diin.2016.07.002.

Pereira, M., Dodhia, R., Anderson, H. and Brown, R. 2020. Metadata-based detection of child sexual abuse material. *arXiv preprint arXiv:2010.02387*.

Powell, M., Cassematis, P., Benson, M., Smallbone, S., Wortley, R. 2015. 'Police officers' perceptions of their reactions to viewing internet child exploitation material'. *J. Police Crim. Psychol*. 30, 103-111.
Ethel Quayle and Nikolaos Koukopoulos. 2019. Deterrence of Online Child Sexual Abuse and Exploitation, *Policing*, 13, 3, 345-362.

Stephan Raaijmakers. 2019. Artificial Intelligence for Law Enforcement: Challenges and Opportunities, Cybercrime and Forensics, *IEEE Security & Privacy*, 17, 5, 74-77.

Afsaheh Razi, Seunghyun Kim, Ashwaq Alsoubai, Gianluca Stringhini, Thamar Solorio, Munmun De Choudhury, Pamela Wisniewski. 2021. A Human-Centered Systematic Review of the Computational Approaches for Online Sexual Risk Detection. *Proc. ACM Human-Computer Interaction*, 5, CSCW2, Article 465. https://doi.org/10.1145/3479609.

Laurie S. Ramiro, Andrea B. Martinez, Janelle Rose D. Tan, Kachela Mariano, Gaea Marelle J. Miranda, Greggy Bautista. 2019. Online child sexual exploitation and abuse: A community diagnosis using the social norms theory. *Child Abuse and Neglect*, 96 (2019) 104080.

Riesco A., Fidalgo E., Al-Nabki M.W., Jáñez-Martino F., Alegre E. 2019. Classifying Pastebin Content Through the Generation of PasteCC Labeled Dataset. In: Pérez García H., Sánchez González L., In Castejón Limas M., Quintián Pardo H., Corchado Rodríguez E. (eds) *Hybrid Artificial Intelligent Systems*. *HAIS 2019. Lecture Notes in Computer Science*, vol 11734. Springer, Cham.

Laura Sanchez, Cinthya Grajeda, Ibrahim Baggili, Cory Hall. 2019. A Practitioner Survey Exploring the Value of Forensic Tools, AI, Filtering, & Safer Presentation for Investigating Child Sexual Abuse Material (CSAM), *Digital Investigation*, 29, S124-S142.

Luca Spalazzi, Marina Paolanti, Emanuele Frontoni. 2021. An offline parallel architecture for forensic multimedia classification, *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-021-10819-x.

Chad M.S. Steel, Emily Newman, Suzanne O'Rourke, Ethel Quayle. 2020. An integrative review of historical technology and countermeasure usage trends in online child sexual exploitation material offenders. *Forensic Science International: Digital Investigation*, 33, 300971.

Stylianou, J. Schreier, R. Souvenir and R. Pless. 2017. "TraffickCam: Crowdsourced and Computer Vision Based Approaches to Fighting Sex Trafficking," *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 1-8, doi: 10.1109/AIPR.2017.8457947.

J. Murcia Triviño, S. Moreno Rodríguez, D. O. Díaz López and F. Gómez Mármol. 2019. "C3-Sex: a Chatbot to Chase Cyber Perverts," *IEEE DASC/PiCom/CBDCom/CyberSciTech*, 50-57, doi: 10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00024.

UNODC (United Nations Office on Drugs and Crime). 2017. *Trafficking in Persons from Cambodia, Lao PDR and Myanmar to Thailand*. Bangkok: UNODC and Thailand Institute of Justice.

Van der Bruggen, M. and Blokland, A. 2021. A crime script analysis of child sexual exploitation material fora on the Darkweb. *Sexual Abuse*, *33*(8), pp.950-974.

Paulo Vitorino, Sandra Avila, Mauricio Perez, Anderson Rocha. 2018. Leveraging deep neural networks to fight child pornography in the age of social media, *Journal of Visual Communication and Image Representation*, 50, Pp 303-313, https://doi.org/10.1016/j.jvcir.2017.12.005.

WePROTECT Global Alliance. 2016. *Preventing and tackling child sexual exploitation and abuse – A model national response*.

We Protect Global Alliance. 2021. *Global Threat Assessment 2021*. https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2021.pdf

Bryce Westlake and Martin Bouchard. 2016. Liking and hyperlinking: Community detection in online child sexual exploitation networks. *Social science research*, 59, 23-36.

Bryce Westlake, Martin Bouchard and Frank, R. 2017. Assessing the Validity of Automated Webcrawlers as Data Collection Tools to Investigate Online Child Sexual Exploitation, *Sexual Abuse*, 29(7), pp. 685–708. doi: 10.1177/1079063215616818.

E. Yiallourou, R. Demetriou and A. Lanitis. 2017. On the detection of images containing child-pornographic material, *24th International Conference on Telecommunications (ICT)*, pp. 1-5, DOI: 10.1109/ICT.2017.7998260.

Kathryn Yount, Tran Hung Minh, Quach Thu Trang, Yuk Fai Cheong, Irina Bergenfel, Jessica Sales. 2020. Preventing Violence in College Men: A Randomized-controlled Trial of GlobalConsent, *BMC Journal of Public Health*, 20,1, 1331-1331.

P. Zambrano, M. Sanchez, J. Torres and W. Fuertes. 2017. BotHook: An option against Cyberpedophilia, *1st Cyber Security in Networking Conference (CSNet)*, pp. 1-3, doi: 10.1109/CSNET.2017.8241994.

Nabilah Hani Ahmad Zubaidi. 2021. Monitoring Internet Pornography (ICP) in Malaysia. Pertanika, *Journal Social Science and Humanities*, 29, 185-203.
DOI: https://doi.org/10.47836/pjssh.29.s2.13

Appendix

Title and Abstract Sift Protocol

	QUESTION	ANSWER	QUALIFIER	ACTION
Q1	Does the title or abstract include online child sexual abuse material?	No		Exclude: code as “Not CSAM”
		Yes		Go to Q2a.
Q2a	Is the document a book / book chapter?	No		Go to Q2b.
		Yes		Exclude: code as “Book”
Q2b	Is the document a conference proceedings?	Yes	Abstract only lists titles of talks / lectures included.	Sift through the titles, assess individual papers exclude or include based on: Exclude: not CSAM Include: Code as PROC
		No		Go to Q2c
Q2c	Is the abstract field blank or says “NA”	Yes		Leave decision and key column blank
		No		Go to Q3
Q3	Is the document an editorial / opinion piece?	No		Go to Q4.
		Yes		Exclude: code as “Opinion”
Q4	Does the title or abstract mention online CSA ***and*** some form of technical intervention?	No		Exclude: code as “CSAM”
		Yes – answer q	Does it relate to law enforcement?	Exclude: code as “LE”
		Yes – answer q	Does it relate to deterrence of CSAM in policy?	Exclude: code as “POL”
		Unclear	Example: could be about use of websites	Include: code as “FTR”
		Yes Answer q	Is it a review of CSAM-TECH?	Include: code as “Review”
		Yes		Go to Q5
Q5	Does the title or abstract mention any of the ASEAN countries*	No		Include: code as “TECH”
		Yes		Include: ASEAN