Australian Government
Australian Institute of Criminology

Trends & issues in crime and criminal justice

No. 653 July 2022

**Abstract |** This paper reviewed open-source materials including electronic service provider (ESP) transparency reports to provide an overview of the contemporary problem of child sexual abuse material (CSAM) offending on ESP platforms, examine measures currently used by ESPs to detect and prevent CSAM offending, and explore the potential impact of end-to-end encryption on CSAM distribution and detection. The study found that the platforms with the highest user bases are actively detecting and removing CSAM. However, some are less transparent than others about the methods they use to prevent, detect and remove CSAM, omitting key information that is crucial for future best practice in reducing CSAM offending. Further, the adoption of end-to-end encryption by ESPs that detect and remove large amounts of CSAM from their platforms will likely provide a haven for CSAM offenders. Implications for ESPs and international law reform are discussed.

# Child sexual abuse material and end-to-end encryption on social media platforms: An overview

Coen Teunissen and Sarah Napier

Child sexual abuse material (CSAM) is any material (eg images and/or videos) that depicts the sexual abuse of a child. The production, distribution and viewing of CSAM has detrimental impacts on victims. This can include trauma, psychological harm, fear, guilt, shame and betrayal (Gewirtz-Meydan et al. 2018; Salter et al. 2021). Even after the physical abuse ends, victims depicted in CSAM can feel repeatedly victimised each time someone views the abusive material (Salter & Hanson 2021). The Canadian Centre for Child Protection (2017) surveyed 128 victims of CSAM; 70 percent reported constantly worrying about being recognised by someone who had viewed the CSAM showing their abuse and 30 percent reported that they had been recognised by someone online or in person who had seen images of their abuse.

Advances in technology and the growing number of internet sites and platforms have provided offenders unprecedented levels of access to children, CSAM and like-minded individuals with whom to share CSAM (WeProtect Global Alliance 2019).

Consequently, CSAM has proliferated in recent years (Balfe et al. 2015; Bursztein et al. 2019), and the National Center for Missing and Exploited Children (NCMEC) received over 21 million reports of CSAM in 2020 alone (NCMEC 2021), increasing to over 29 million in 2021 (NCMEC 2022). The problem may have been exacerbated by global events such as the COVID-19 pandemic (Interpol 2020). Law enforcement agencies struggle to keep up with the staggering number of CSAM reports to investigate (NCMEC 2021; Netclean 2020). According to Netclean (2020), which surveyed 470 law enforcement officers in 39 countries, police globally were inundated with CSAM cases in 2020. This affected the mental health of some officers and meant they could investigate only the most high-risk cases. Further, use of end-to-end encryption by communication platforms, while designed for user safety, may present challenges for law enforcement in combatting CSAM offending.

Most relevant research conducted thus far has focused on the characteristics of individuals who view, distribute or produce CSAM (eg Brown & Bricknell 2018; Seto & Eke 2015)—for example, the differences between those who commit online offences and those who sexually abuse children in person (Babchishin, Hanson & VanZuylen 2015). Less research has focused on the specific online platforms where CSAM is shared and the methods these platforms use to detect and remove CSAM. Information is publicly available on the number of reports of CSAM detected on specific electronic service provider (ESP) platforms (eg NCMEC 2021); additionally, transparency reports released by each relevant ESP discuss the prevention, detection and removal measures they adopt. However, this information can be inconsistent, lacking in important details and difficult to find because it is scattered across multiple reports and websites. Given the recent dramatic increase in the amount of CSAM detected by ESPs, the adverse consequences for victims and the dynamic nature of this offence type, it is crucial to fully understand the extent and nature of the problem and the current CSAM detection and removal methods used. This will help us to develop best practice methods to prevent and disrupt CSAM offending.

## Aims and method

The purpose of this study is to provide a valuable resource to inform policy and international law reform by consolidating key information on entities that detect CSAM on their platforms. The paper has three aims:

- to provide an overview of the contemporary problem of CSAM offending on mainstream ESP platforms;
- to investigate the measures ESPs currently use to detect, remove and report CSAM on their platforms; and
- to explore the potential impact of end-to-end encryption on CSAM distribution and detection.

The 10 ESPs that sent the largest number of CSAM reports to NCMEC in 2020 were identified via NCMEC's (2021) report, which was the latest available at the time of data collection. Information sought for each identified ESP included: the number of CSAM reports provided to NCMEC; the company protocols for CSAM prevention, detection and removal; and the end-to-end encryption status of the messaging/chat function. This information was obtained directly from ESP websites—for example, from transparency reports, annual reports and official blogs. This material was

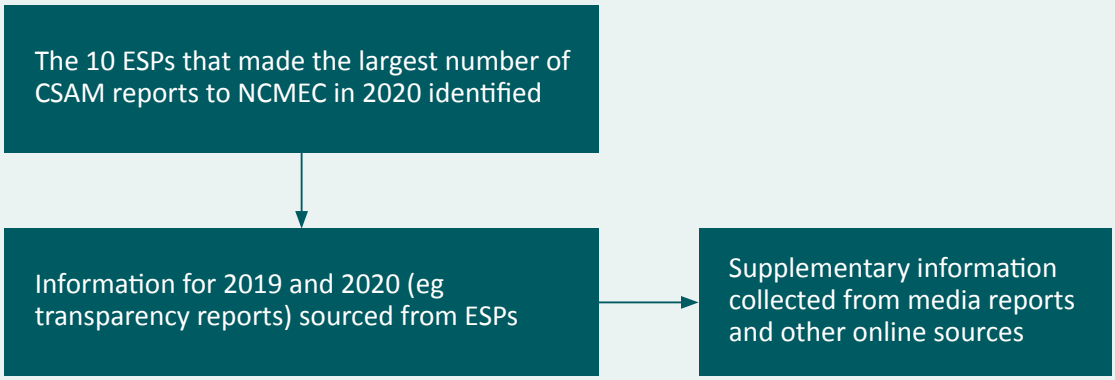
supplemented with media releases and news articles (see Figure 1). The search was limited to data from 2015 onwards and ESP transparency reports from 2019 and 2020 due to data availability limitations, the dynamic nature of online communication platforms and the necessity for this review to be as up to date as possible. However, readers should note that NCMEC (2022) has since released its report on notifications received during 2021.

The paper focuses on open web platforms and does not cover darknet sites, nor describe all available types of websites or encryption or anonymisation technologies which offenders have used to share and access CSAM, such as virtual private networks (VPNs) and peer-to-peer networks. Only ESPs that report CSAM to NCMEC were included. From here on, we use the term 'ESP' to describe both social media platforms and search engines (Table 1).

**Table 1: Acronyms and terminology**

| Acronym or term | Explanation |
|---|---|
| CSAM | Child sexual abuse material (known legally in some countries as child pornography) |
| End-to-end encryption | A system of communication that usually allows only the sender or receiver of a private message or video call to view the content |
| ESP | Electronic service provider (eg Google, Meta) |
| NCMEC | National Center for Missing and Exploited Children |

**Figure 1: Flow diagram of resource selection process**



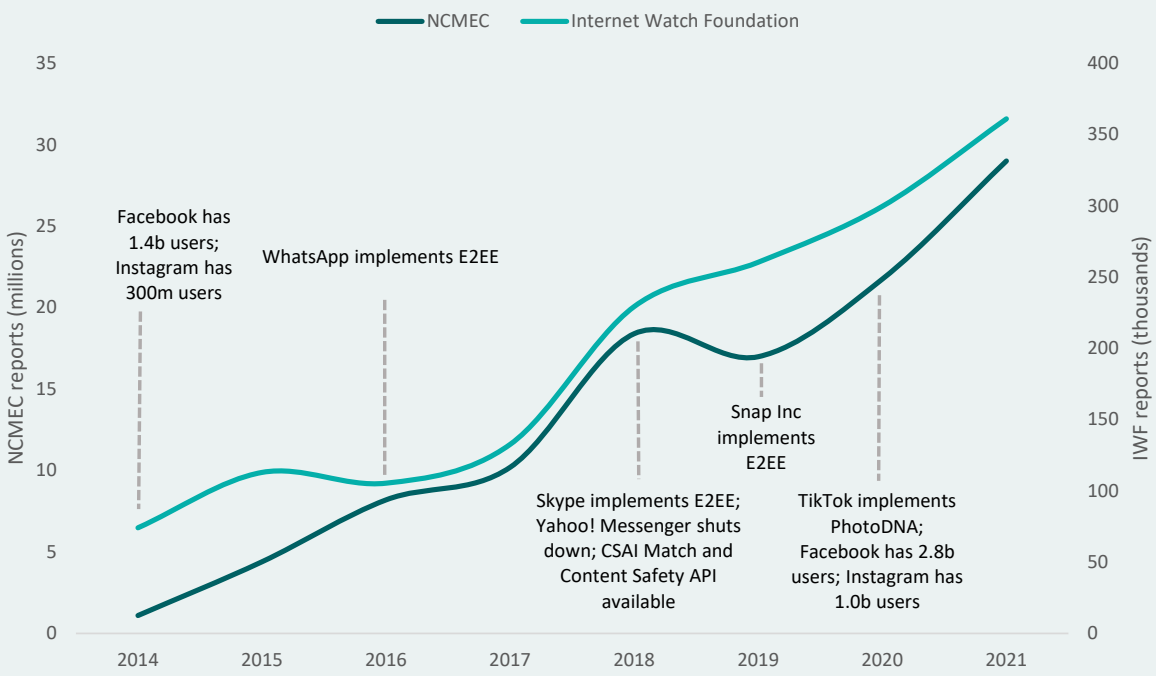
# CSAM detected on online platforms

## Prevalence of CSAM and growth in reports

It is not possible to know the true amount of CSAM available on the internet. However, police and non-government organisations (NGOs) provide insight into the number of reports of CSAM received from businesses and the public. In 2019, the Australian Federal Police received 450 CSAM reports from members of the public and 13,368 from NCMEC relating to Australian victims, offenders or IP addresses. Both of these figures increased in 2020, to 559 and 21,148 reports respectively (Australian Federal Police personal communication 01 September 2021). Unfortunately, these data were not available for Australia before 2019.

Both NCMEC and the Internet Watch Foundation are NGOs to whom members of the public and ESPs can report suspected online sexual exploitation of children, including CSAM. In 2021, 99 percent of such reports received by NCMEC related to CSAM (NCMEC 2022; herein referred to as CSAM reports). The rapid growth in CSAM availability in recent years is evident in the statistics provided by these two organisations. Figure 2 shows a clear upward trend in the number of CSAM reports to both entities across an eight-year period. The number of reports received by NCMEC increased from 1.1m reports (28m CSAM files) in 2014, to 21.7m reports (65.5m CSAM files) in 2020 and 29.3m reports (84.7m CSAM files) in 2021. The number of reports received by the Internet Watch Foundation of URLs containing CSAM increased from 74,000 URLs in 2014 to 299,619 URLs in 2020 and 361,062 URLs in 2021. This increase in CSAM reports coincided with an increase in users of Facebook (from 1.4b to 2.8b) and Instagram (from 300m to 1.0b). Between 2016 and 2017 there was a clear change to the otherwise sharp upward trend in the number of CSAM reports received by NCMEC. This coincided with the year WhatsApp implemented end-to-end encryption (2016). Between 2018 and 2019, there was a clear drop in the number of CSAM reports received by NCMEC (from 18.4m in 2018 to 17m in 2019), coinciding with the year Skype implemented end-to-end encryption and Yahoo Messenger shut down. There was no drop in the number of CSAM reports following Snap Inc implementing end-to-end encryption in 2019. However, due to the small number of data points, it was not possible to conduct an interrupted time series analysis to determine whether the changes in encryption status were associated with the change in trends. It is also possible that other factors influenced these changes in the trends, including an increase in detection and reporting by ESPs.

**Figure 2: Number of CSAM reports to NCMEC and Internet Watch Foundation, 2014–2021**




Note: National Center for Missing and Exploited Children (NCMEC) reports relate to online sexual exploitation of children, most of which comprises CSAM images and/or videos, and each report potentially involves multiple files. Internet Watch Foundation (IWF) reports relate to websites or URLs that contain CSAM. It is possible that there is some double-counting where reports were made to both entities. E2EE=end-to-end encryption

Sources: Internet Watch Foundation 2014–2021; NCMEC 2014–21; Statista 2022a, 2022b; Synstrom 2014

## CSAM on specific platforms

NCMEC publishes the number of CSAM reports made by individual ESPs (NCMEC 2021). At the time of data collection, the most recent report related to 2020. Of the 21.7m CSAM reports received by NCMEC in 2020, Meta, who owns Facebook, Messenger, Instagram and WhatsApp, made 93 percent (*n*=20,307,216; see Table 2). The number of reports from this company dwarfed those of the remaining top five companies reporting the most CSAM: Google (546,704), Snap (144,095), Microsoft (96,776) and Twitter (65,062).

**Table 2: CSAM reports to NCMEC in 2020, detection methods and end-to-end encryption status for top 10 reporting ESPs**

| ESP | CSAM reports 2020 (*n*) | % of reports to NCMEC | CSAM detection and information protection | End-to-end encryption messaging? |
|---|---|---|---|---|
| Meta[a] | 20,307,216 | 93.4 | Uses PhotoDNA, proprietary machine learning and artificial intelligence.<br>Commenced using Google's Content Safety API in 2021 (see below). | WhatsApp—yes (since 2016)<br>Facebook Messenger—encryption can be manually enabled<br>Instagram—no |
| Google[a] | 546,704 | 2.5 | Developed/uses CSAI Match and Content Safety API.<br>Data encrypted at rest and in transit. | Messages—yes (since 2020)<br>Gmail—no |
| Snap Inc | 144,095 | 0.7 | Uses PhotoDNA and CSAI Match. | Snapchat—yes (since 2019) |
| Microsoft[a] | 96,776 | 0.4 | Developed/uses PhotoDNA. | Skype—encryption can be manually enabled (since 2018)<br>Teams—yes<br>Outlook—no |
| Twitter | 65,062 | 0.3 | Uses PhotoDNA. | No |
| INHOPE[b] | 57,170 | 0.3 | N/A—not a platform. | N/A—not a platform |
| Imgur | 31,571 | 0.1 | Messages and images can be monitored and accessed. | No |
| TikTok | 22,692 | 0.1 | Uses PhotoDNA and other technologies.<br>Collects and scans information contained in messages being composed, sent or received. | No |
| Dropbox | 20,928 | 0.1 | Public files only viewable by people who have a link to the file/s (and by Dropbox staff).<br>Data encrypted at rest and in transit. | No |
| Omegle | 20,265 | 0.1 | No information available. | No |

a: Aggregate includes the total for all platforms owned by a company. For example, Facebook, Messenger, Instagram and WhatsApp are included under Meta (previously Facebook Inc); Google, Gmail and YouTube are included under Google; Bing, Xbox, Outlook and others are included under Microsoft

b: INHOPE is a global network of CSAM reporting hotlines partnered with Interpol and Microsoft and funded by the European Union. It is likely that these reports sent to NCMEC relate to reports received by a hotline, rather than CSAM activity on INHOPE's website

Sources: Allen 2011; Arthur 2013; Canegallo 2021; Davis 2021a; Deahl 2018; Dropbox nda, ndb; Google nda, ndb, ndc, ndd, nde; Gruszczyk 2021; Holt 2021; Imgur 2019; Microsoft nda, ndb; NCMEC 2021; Omegle 2022; Signal 2018; Snap Inc 2021a, 2021b; TikTok 2021, 2022; Titcomb 2019; Vincent 2018

# CSAM detection methods used by platforms

As well as examining the extent of CSAM detected on specific platforms, it is important to look at the methods currently used by the platforms to detect and disrupt this offending. The authors could not locate any information on the methods used by Imgur, Dropbox or Omegle to detect or remove CSAM.

## Microsoft

At the time of writing, one of the main technologies industry use to detect CSAM is 'PhotoDNA', which detects the digital fingerprint ('hash') of an image and compares it against a database of previously identified CSAM. This was developed by Microsoft in 2009, and the video-detection capability was implemented in 2018 (Langston 2018). It is provided freely to eligible organisations and businesses, with over 100 companies currently using the tool (Thorn 2016). Microsoft has also developed technology that can flag suspicious conversations for review by human moderators, to assist companies to detect online grooming on their platforms (Gregoire 2020).

According to Microsoft's 2020 transparency reports (Microsoft 2021a, 2021b), of the 1,079,246 abusive images and videos removed, blocked or delisted through Bing in 2020, 99.5 percent were proactively detected by its PhotoDNA technology. Similarly, 99.9 percent of the 177,000 CSAM files removed or blocked from other Microsoft services (OneDrive, Outlook, Skype, Xbox and others) were proactively detected by its technology, resulting in 33,369 accounts being blocked or suspended (Microsoft 2021a, 2021b).

However, empirical evaluations on the effectiveness of this technology are scarce. While it is clearly responsible for detecting a large amount of CSAM (Farid 2019), a New York Times investigation found vulnerabilities in PhotoDNA in the Bing search engine (Keller & Dance 2019). Further, PhotoDNA only detects known CSAM stored in databases such as NCMEC's; CSAM that is new, substantially altered or not yet known to law enforcement is not detected.

## Meta (previously Facebook Inc)

Meta uses PhotoDNA (Allen 2011; Davis 2020) and unspecified artificial intelligence (AI) to proactively detect CSAM, including previously undetected material (Davis 2018). Meta also commenced using Google's Content Safety API in 2021 (Davis 2021a). However, there is little information available on their proprietary AI, including how it detects new CSAM. Meta also uses pop-up messages and other technologies to prevent online grooming and CSAM by deterring users who enter search terms linked with child sexual exploitation (Davis 2021a). Evidence suggests that pop-up warning messages can be effective in deterring individuals from sharing sexual images of young females online (Prichard et al. 2022), but no studies have measured the success of pop-ups currently used by platforms like Facebook. Meta also provides information about the illegality and harms resulting from online sexual exploitation of children to users whose accounts are flagged, disabled and/or removed and reports these users to NCMEC (Davis 2021a).

Meta's transparency report (Facebook nd) provides statistics relating to Facebook, Messenger and Instagram platforms. Nearly all (99%) of the child nudity and sexual exploitation material found on Facebook in 2019 and 2020 was detected and flagged by Meta's technology, compared to one percent being reported by users. In 2019, 99 percent of the 37.4m child nudity and sexual exploitation files removed by Facebook were proactively found by the technology, with only one percent being reported by users. In 2020, this rate remained at 99 percent across the 35.9m files

removed. Similar rates were also observed for Instagram: 97 percent of the 3.3m pieces of abusive content (posts, photos, videos or comments) found in 2020 were detected by Meta's technology (Facebook nd). While this demonstrates that Meta's technology is successful in detecting large amounts of CSAM on Facebook and Instagram, these rates do not compare the content detected versus the content available on the platform. In other words, we do not know how much CSAM was overlooked. This rate has also been questioned by experts in the field who call for more transparency around Meta's definitions of CSAM (Gladstone 2019).

WhatsApp states that unencrypted information such as profile and group photos and metadata (eg dates of contact, IP addresses) on the platform is scanned using technology (not specified) to match photos/videos against known CSAM (WhatsApp 2021). WhatsApp bans over 300,000 accounts per month for sharing CSAM (WhatsApp 2021). Due to its use of end-to-end encryption, WhatsApp is unable to detect child sexual exploitation distributed via private messages unless users report it directly.

## Google

Google developed 'Content Safety API', which uses AI to identify and triage CSAM for review by organisations, providing this tool freely to other companies and NGOs including the Internet Watch Foundation and the Canadian Centre for Child Protection (Google nd). Google and YouTube also developed 'CSAI Match', a proprietary technology to detect CSAM videos. Google makes this technology freely available to NGOs and industry (Canegallo 2021). Google also presents warning messages containing links to reporting hotlines to individuals who enter search terms associated with child sexual exploitation (Google nd).

In 2020, Google detected 4,266,308 CSAM files on its platform (Google 2021). Over this period, 542,621 URLs relating to CSAM were removed from Google's search index, and Google identified 1,449,284 new CSAM hashes and shared them with NCMEC. Additionally, 188,955 CSAM reports were submitted by YouTube, and a total of 175,898 YouTube accounts were disabled for CSAM violations (Google 2021).

Steel (2015) found that Google and Bing's efforts to block searches for terms related to CSAM resulted in a 67 percent drop in CSAM searches during a one-year period over 2013 and 2014 in the United States. Yandex, an ESP located in Russia, did not implement similar blocking efforts nor experience a commensurate drop in CSAM searches during this period. However, the study could not distinguish between the effect of blocking and that of other interventions (eg warning messages) or a user shift from clear web to darknet searches.

## Other companies

### *Snap, TikTok and Twitter*

Snap Inc (Snapchat), TikTok and Twitter use PhotoDNA to detect CSAM images. Snap Inc also uses Google's CSAI Match to detect CSAM videos (Snap Inc 2021a, 2021b). TikTok and Twitter also use other unspecified technologies to detect CSAM. In 2020:

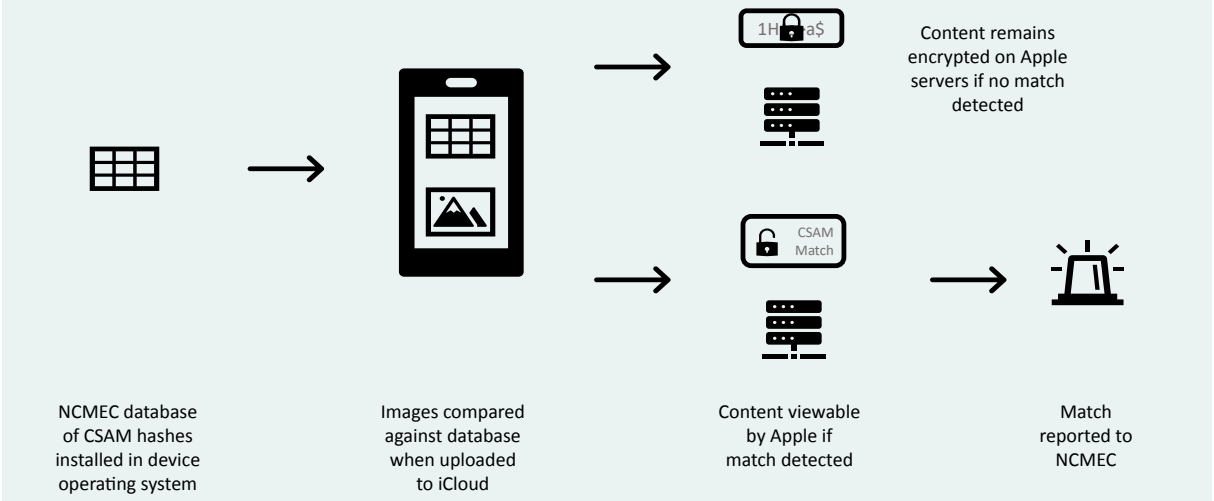
- Twitter suspended 903,613 accounts and removed 19,521 files associated with child sexual exploitation (Twitter 2021a, 2021b);
- Snap Inc deleted 94,686 accounts for suspected child sexual exploitation, including grooming of children, and 1,808 of these accounts belonged to Australian users (Snap Inc 2021a, 2021b); and
- TikTok made 22,692 CSAM reports to NCMEC, compared with 596 in 2019 (TikTok 2021).

### *Apple*

Although Apple was not one of the top 10 ESPs submitting the largest number of CSAM reports, the company developed technology that is important to report in this paper. In August 2021 Apple announced its new technology 'NeuralHash', which scans devices for known CSAM when images are uploaded to iCloud (see Figure 3; Apple 2021a; Nicas 2021). Content that matches a NCMEC hash of known CSAM is flagged for human review and can then be reported to NCMEC and law enforcement for action. This ensures that iCloud content is still encrypted via end-to-end encryption (Apple 2021b), unless it matches known CSAM when it is uploaded (with an error rate of less than one in a trillion per year; Apple 2021a). This technology is innovative in its method of detection; however, it appears to be limited to CSAM images (Apple 2021a), with no indication of being applicable to videos yet. It also has the same limitation as PhotoDNA in that it is unable to detect new or substantially altered CSAM, and is limited to iCloud content.

It remains to be seen how effective this tool will be in detecting and preventing CSAM offending. Other technology company leaders, including the Chief Executive Officer of WhatsApp, have criticised this technology, expressing privacy concerns (Clayton 2021). In September 2021, Apple delayed the implementation of this technology, citing concerns from customers, privacy campaigners and others (Wakefield 2021).

**Figure 3: Explanation of Apple's NeuralHash technology**




Source: Adapted from Apple 2021a

In summary, most ESPs examined are actively detecting and removing CSAM. However, transparency reports relating to CSAM and online sexual exploitation of children are not produced consistently across ESPs. There is also little reliable or detailed information available on definitions of CSAM used and, for some ESPs, the detection and prevention tools used and their effectiveness. Lastly, it is important to note that this study focused on the 10 ESPs who provided the highest number of CSAM reports to NCMEC in 2020. There are many other ESPs for whom the methods of detecting and reporting CSAM were not reviewed in this study.

## End-to-end encryption

End-to-end encryption ensures that information on a platform is visible only to the individual or entity who has the 'key' to decrypt it (Schiemer 2018), and in almost all cases, this is only the sender and recipient. This is designed to protect sensitive and personal information such as messages and transactions (eSafety Commission 2020), and also to protect users from malicious online activity such as cybercrime (Amnesty International 2016). Data show that apps with security features have more active users (Stevens 2020), which demonstrates the appeal of privacy to the public.

Unfortunately, end-to-end encryption presents significant challenges to law enforcement officers who investigate CSAM offending (Netclean 2019), and limits companies' ability to prevent, detect and report CSAM occurring on their platforms. For example, online chat logs are a key form of evidence in CSAM investigations. In one such Australian case, the offender used several popular platforms to distribute CSAM he had produced, which involved severe abuse of babies (*Commonwealth Director of Public Prosecutions v CCQ* [2021] QCA 4 (22 January 2021); warning: contains highly graphic details of abuse). In this case the offender's chat logs were used as evidence to demonstrate the severity of offending that took place. In another case, Meta detected CSAM in a conversation between an Australian man and a Filipino child, leading to the man's arrest when he travelled to the Philippines (Murdoch 2016). If the platforms used by these offenders had implemented end-to-end encryption at that time, this evidence may not have been available for investigations and the offenders may still be at large.

Currently, four of the companies that send the top 10 number of CSAM reports to NCMEC use end-to-end encryption for private messages on some of their platforms: Meta (used on WhatsApp), Google, Snap and Skype (see Table 2). Snapchat introduced end-to-end encryption for private messages including photos in 2019 (Titcomb 2019). WhatsApp has two billion users (WhatsApp 2020)—more than both Facebook Messenger (988 million; Statista 2022c) and Instagram (1 billion; Statista 2022a). However, given its use of end-to-end encryption, CSAM cannot be detected in WhatsApp conversations unless users report it.

Meta plans to implement universal end-to-end encryption on Facebook Messenger and Instagram's private messages in 2023 (Davis 2021b; Kent 2021). These are two of the largest social media platforms in the world. There are concerns that current CSAM detection technologies (eg PhotoDNA, AI) will not work in an end-to-end encryption environment as they will be unable to decrypt the content and scan it for CSAM (NCMEC 2019). To our knowledge, Meta has not publicly proposed a strategy to maintain its ability to detect and report CSAM on these platforms in the end-to-end encryption environment. As most CSAM on Facebook Messenger is detected using PhotoDNA and AI tools (Facebook nd; Farid 2019), NCMEC has estimated that Meta's implementation of end-to-end encryption across all its major platforms would reduce the number of CSAM reports it receives by more than 50 percent (NCMEC 2019). This would mean that more than 18m CSAM files would potentially be undetected and unreported.

# Discussion

In 2020 the number of CSAM reports received by NCMEC reached 21.7m, relating to approximately 65.5m CSAM files. This increased to 29.3m reports in 2021. In 2020 the overwhelming majority of CSAM reports—93 percent—were made by Meta.

The large amount of CSAM detected by Meta may be due to the immense number of users on its platforms—WhatsApp, Messenger and Instagram each have close to or more than a billion users (Statista 2022a, 2022c; WhatsApp 2020)—as well as the methods it uses to proactively detect and remove CSAM (Thorn 2021) and the access it has to platform content. In contrast, ESPs such as Google provide search tools that direct users to third-party websites. While Google can prevent websites from appearing in its search results, it has no control over the content on the sites except for those that it owns (eg Google Drive and YouTube; Google 2021). It has also been noted that large social media platforms such as Facebook have functions that allow like-minded offenders to easily connect with each other and share CSAM (Andrus, Buckley & Williams 2021). Similarly, offenders may use the functions on social media platforms such as Facebook to coerce children to create self-produced CSAM or to engage in live streamed abuse (eg de Santisteban et al. 2018; Napier, Teunissen & Boxall 2021), whereas the functions of services such as Dropbox, OneDrive, Outlook and Gmail may be less conducive to this type of offending.

If Meta introduces end-to-end encryption on Facebook Messenger and Instagram in 2023 as planned, the amount of CSAM currently detected and reported to NCMEC may be reduced by more than half (NCMEC 2019). This means that most CSAM on Meta's platforms will be invisible even to Meta. In 2020 WhatsApp had two billion users (WhatsApp 2020), almost twice as many as Facebook Messenger or Instagram. Yet, due to its use of end-to-end encryption, it remains unclear how much CSAM is circulated on WhatsApp.

## Implications for ESPs

While most ESPs examined in this study are publicly opposed to CSAM and proactively detect and remove it, they are unfortunately still inadvertently facilitating its distribution. There are also concerns that ESPs are deflecting responsibility for preventing CSAM distribution and focusing on reporting CSAM if they find it (Salter & Hanson 2021). The staggering amount of CSAM reported by ESPs places a huge burden on law enforcement, who struggle to keep up with the workload (NCMEC 2021; Netclean 2020). There are several measures that ESPs should adopt to help address the problem.

Firstly, every company should be consistent in their reporting of CSAM and transparent about their definitions of CSAM and the specific measures they use to prevent, detect and report it. Providing this detailed information will help enforce best practice standards and assist companies to improve their tools for preventing harm to children.

Secondly, more responsibility should be placed on ESPs to prevent CSAM from being uploaded in the first instance. These platforms should adopt evidence-based methods such as pop-up warning messages (Prichard et al. 2022), which can deter engagement with CSAM and refer individuals to sources of support to stop their offending. Deterrence messaging campaigns can potentially also

reach large numbers of individuals (Grant et al. 2019). These tools should also be evaluated; Meta currently uses pop-up warning messages to deter child sexual exploitation, yet there is no information publicly available on their impact or effectiveness.

Lastly, ESPs should invest in more innovative technology. Currently, NeuralHash, Apple's proposed technology to scan devices for CSAM, is the only publicly described tool that will detect CSAM in an end-to-end encryption environment. Although Apple has delayed the release of this technology (Wakefield 2021), ESPs should similarly invest in developing technology to prevent CSAM from being uploaded onto their platforms in the first instance. This would supplement their current methods of detecting and removing CSAM from their platforms.

## Implications for international law reform

The beginning of this paper outlined the deleterious effects of CSAM offending on victims. The monumental increase in availability of CSAM on major platforms will result in continuing and increased harm to children. Yet debate continues over the importance of protecting children versus the privacy of individuals (Allen 2021). The adoption of end-to-end encryption by more ESPs will likely provide a haven for CSAM offending, rather than preventing it. Further policy discussions are required about the best way forward in terms of end-to-end encryption which consider the impact it would have on current and future child victims.

Secondly, it would be beneficial for countries globally, including the Five Eyes nations and European nations, to introduce legislation requiring communication platforms to report CSAM consistently and adopt evidence-based detection and prevention measures. Additionally, ESPs should be consistent and transparent in how they report:

- definitions of CSAM;
- the amount of CSAM detected and removed;
- the number of accounts banned, suspended and/or deleted due to child sexual exploitation;
- detailed methods for detecting and removing CSAM;
- detailed methods for preventing CSAM offending (eg messaging campaigns, warning messages); and
- evaluations of the effectiveness of these methods in detecting and preventing CSAM offending.

Adopting such legislation will help reduce the sexual abuse and online sexual exploitation of children globally.

## Limitations and future directions

This study had several limitations which should be considered. Firstly, this paper examined reports to NCMEC, which only companies registered in the United States are required to submit under the US Criminal Code. Hence, this paper did not examine CSAM occurring on platforms owned by companies registered in other countries. It is important to note that only detected CSAM was reported and not all CSAM is detected by ESPs or law enforcement. Finally, research has found that offenders use social media platforms to groom children for sexual abuse and exploitation (de Santisteban et al. 2018) including live streaming of child sexual abuse (Napier, Teunissen & Boxall 2021). However, it was beyond the scope of this paper to examine grooming. Future research focused on how offenders use online platforms to groom children for sexual abuse and exploitation and the methods used by platforms to prevent and detect this offending would be a valuable addition to the knowledge base. Lastly, due to the lack of academic research into this topic, this paper relied heavily on grey literature that had not been peer reviewed.

The problem of CSAM offending on online platforms is growing and changing rapidly, exacerbated by global events such as the COVID-19 pandemic. Therefore, despite its limitations, this paper is valuable and timely in providing an overview of this dynamic and harmful crime to help inform industry, policy and law reform.

# Acknowledgements

The authors would like to acknowledge former Australian Institute of Criminology Research Officer Cameron Long for his assistance with data extraction.

# References

*URLs correct as at March 2022*

Allen E 2021. Defending the privacy of child sexual abuse victims online, in the EU and worldwide. https://www.weprotect.org/blog/defending-the-privacy-of-child-sexual-abuse-victims-online-in-the-eu-and-worldwide/

Allen E 2011. *Facebook to use Microsoft's PhotoDNA technology to combat child exploitation*. https://blogs.microsoft.com/on-the-issues/2011/05/19/facebook-to-use-microsofts-photodna-technology-to-combat-child-exploitation/

Amnesty International 2016. *Easy guide to encryption and why it matters.* London: Amnesty International. https://www.amnesty.org/en/latest/campaigns/2016/10/easy-guide-to-encryption-and-why-it-matters/

Andrus M, Buckley J & Williams C 2021. *Understanding the intentions of child sexual abuse material (CSAM) sharers*. https://research.facebook.com/blog/2021/02/understanding-the-intentions-of-child-sexual-abuse-material-csam-sharers/

Apple 2021a. *Expanded protections for children: Technology summary*. https://web.archive.org/web/20210816220924/https://www.apple.com/child-safety/

Apple 2021b. iCloud security overview. https://support.apple.com/en-us/HT202303

Arthur C 2013. Twitter to introduce PhotoDNA system to block child abuse images. *The Guardian*, 22 July. https://www.theguardian.com/technology/2013/jul/22/twitter-photodna-child-abuse

Babchishin KM, Hanson RK & VanZuylen H 2015. Online child pornography offenders are different: A meta-analysis of the characteristics of online and offline sex offenders against children. *Archives of Sexual Behaviour* 44(1): 45–66. https://doi.org/10.1007/s10508-014-0270-x

Balfe M et al. 2015. Internet child sex offenders' concerns about online security and their use of identity protection technologies: A review. *Child Abuse Review* 24: 427–439. https://doi.org/10.1002/car.2308

Brown R & Bricknell S 2018. What is the profile of child exploitation material offenders? *Trends & issues in crime and criminal justice* no. 564. Canberra: Australian Institute of Criminology. https://www.aic.gov.au/publications/tandi/tandi564

Bursztein E et al. 2019. *Rethinking the detection of child sexual abuse imagery on the internet*. International World Wide Web Conference, San Francisco, 13–17 May: 2601–2607. https://doi.org/10.1145/3308558.3313482

Canadian Centre for Child Protection (C3P) 2017. *Survivors' survey: Full report 2017*. Winnipeg: C3P. https://www.protectchildren.ca/en/resources-research/survivors-survey-results/

Canegallo K 2021. Our efforts to fight child sexual abuse online. https://blog.google/technology/safety-security/our-efforts-fight-child-sexual-abuse-online/

Clayton J 2021. Apple criticised for system that detects child abuse. *BBC News*, 7 August. https://www.bbc.com/news/technology-58124495

Davis A 2021a. Preventing child exploitation on our apps. https://about.fb.com/news/2021/02/preventing-child-exploitation-on-our-apps/

Davis A 2021b. We'll protect privacy and prevent harm, writes Facebook safety boss. *The Telegraph*, 20 November. https://www.telegraph.co.uk/business/2021/11/20/people-shouldnt-have-choose-privacy-safety-says-facebook-safety/

Davis A 2020. Facebook joins industry effort to fight child exploitation online. https://about.fb.com/news/2020/06/fighting-child-exploitation-online/

Davis A 2018. New technology to fight child exploitation. https://about.fb.com/news/2018/10/fighting-child-exploitation/

de Santisteban P, del Hoyo J, Alcázar-Córcoles MÁ & Gámez-Guadix M 2018. Progression, maintenance, and feedback of online child sexual grooming: A qualitative analysis of online predators. *Child Abuse & Neglect* 80: 203–215. https://doi.org/10.1016/j.chiabu.2018.03.026

Deahl D 2018. Skype now offers end-to-end encrypted conversations. *The Verge*, 20 August. https://www.theverge.com/2018/8/20/17725226/skype-private-conversation-end-to-end-encrypted-opt-in

Dropbox nda. *Who can see the stuff in my Dropbox account?* https://help.dropbox.com/accounts-billing/security/file-access

Dropbox ndb. *How Dropbox keeps your files secure*. https://help.dropbox.com/accounts-billing/security/how-security-works

eSafety Commission 2020. *End-to-end encryption trends and challenges – position statement*. Sydney: eSafety Commissioner. https://www.esafety.gov.au/about-us/tech-trends-and-challenges/end-end-encryption

Facebook nd. Community standards enforcement report: Child endangerment: Nudity and physical abuse and sexual exploitation. https://transparency.fb.com/data/community-standards-enforcement/child-nudity-and-sexual-exploitation/facebook

Farid H 2019. *Testimony: House Committee on Energy and Commerce: Fostering a healthier internet to protect consumers*. Washington DC: House Committee on Energy and Commerce. https://energycommerce.house.gov/committee-activity/hearings/hearing-on-fostering-a-healthier-internet-to-protect-consumers

Gewirtz-Meydan A, Walsh W, Wolak J & Finkelhor D 2018. The complex experience of child pornography survivors. *Child Abuse & Neglect* 80: 238–248. https://doi.org/10.1016/j.chiabu.2018.03.031

Gladstone N 2019. Child sexual abuse material spreading 'exponentially' on social media. *Sydney Morning Herald*, 2 July. https://www.smh.com.au/national/child-sexual-abuse-material-spreading-exponentially-on-social-media-20190701-p520nn.html

Google nda. Discover our child safety toolkit. https://protectingchildren.google/tools-for-partners/

Google ndb. Email encryption in transit. https://support.google.com/mail/answer/6330403?hl

Google ndc. Featured policies: Child safety. https://transparencyreport.google.com/youtube-policy/featured-policies/child-safety?hl

Google ndd. Fighting child sexual abuse online. https://protectingchildren.google/intl/en_au/

Google nde. How end-to-end encryption in Messages provides more security. https://support.google.com/messages/answer/10262381?hl

Google 2021. Google's efforts to combat online child sexual abuse material. https://transparencyreport.google.com/child-sexual-abuse-material/reporting

Grant B, Shields B, Tabachnick J & Coleman J 2019. "I didn't know where to go": An examination of Stop It Now!'s sexual abuse prevention helpline. *Journal of Interpersonal Violence* 34(20): 4225–4253. https://doi.org/10.1177/0886260519869237

Gregoire C 2020. Microsoft shares new technique to address online grooming of children for sexual purposes. https://blogs.microsoft.com/on-the-issues/2020/01/09/artemis-online-grooming-detection/

Gruszczyk J 2021. End-to-end encryption for one-to-one Microsoft Teams calls now generally available. https://techcommunity.microsoft.com/t5/microsoft-teams-blog/end-to-end-encryption-for-one-to-one-microsoft-teams-calls-now/ba-p/3037697

Holt K 2021. The Android Messages app now offers end-to-end encryption. *Engadget*, 15 June. https://www.engadget.com/android-messages-end-to-end-encryption-update-191641092.html

Imgur 2019. *Privacy Policy*. https://imgurinc.com/privacy

Internet Watch Foundation 2014–2021. Annual report (various issues). https://www.iwf.org.uk/about-us/who-we-are/annual-report/

Interpol 2020. *Threats and trends: Child sexual exploitation and abuse: COVID-19 impact*. Lyon: Interpol. https://www.interpol.int/News-and-Events/News/2020/INTERPOL-report-highlights-impact-of-COVID-19-on-child-sexual-abuse

Keller MH & Dance GJX 2019. Child abusers run rampant as tech companies look the other way. *New York Times*, 9 November. https://www.nytimes.com/interactive/2019/11/09/us/internet-child-sex-abuse.html

Kent G 2021. Messenger policy workshop: Future of private messaging. https://about.fb.com/news/2021/04/messenger-policy-workshop-future-of-private-messaging/

Langston J 2018. *How PhotoDNA for Video is being used to fight online child exploitation*. https://news.microsoft.com/on-the-issues/2018/09/12/how-photodna-for-video-is-being-used-to-fight-online-child-exploitation/

Microsoft nda. Digital safety content report. https://www.microsoft.com/en-us/corporate-responsibility/digital-safety-content-report

Microsoft ndb. Encryption in Outlook. https://support.microsoft.com/en-us/office/encryption-in-outlook-17149eb8-a82e-405b-af5a-4fb89ce4a418

Microsoft 2021a. *Digital safety content report 2020 H1 – January–June 2020*. https://www.microsoft.com/en-us/corporate-responsibility/digital-safety-content-report

Microsoft 2021b. *Digital safety content report 2020 H2 – Jul–Dec 2020*. https://www.microsoft.com/en-us/corporate-responsibility/digital-safety-content-report

Murdoch L 2016. Australian accused of child sex tourism arrested in the Philippines. *Sydney Morning Herald*, 1 September. https://www.smh.com.au/world/australian-accused-of-child-sex-tourism-arrested-in-the-philippines-20160901-gr6x8x.html

Napier S, Teunissen C & Boxall H 2021. How do child sexual abuse live streaming offenders access victims? *Trends & issues in crime and criminal justice* no. 642. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78474

National Centre for Missing and Exploited Children (NCMEC) 2022. CyberTipline 2021 report. https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata

National Center for Missing and Exploited Children (NCMEC) 2021. 2020 CyberTipline reports by electronic service providers (ESP). https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata#reports

National Center for Missing and Exploited Children (NCMEC) 2019. NCMEC's statement regarding end-to-end encryption. https://www.missingkids.org/blog/2019/post-update/end-to-end-encryption

National Center for Missing and Exploited Children (NCMEC) 2014–2020. Annual report (various issues). https://www.missingkids.org/footer/about/annual-report

Netclean 2020. Netclean report: COVID-19 impact 2020: A report about child sexual abuse crime. https://www.netclean.com/netclean-report-2020/

Netclean 2019. Netclean report 2019: A report about child sexual abuse crime. https://www.netclean.com/netclean-report-2019/

Nicas J 2021. Apple's iPhones will include new tools to flag child sexual abuse. *New York Times*, 5 August. https://www.nytimes.com/2021/08/05/technology/apple-iphones-privacy.html

Omegle 2022. *Omegle Privacy Notice*. https://www.omegle.com/static/privacy.html

Prichard J, Wortley R, Watters P, Spiranovic C, Hunn C & Krone T 2022. Effects of automated messages on internet users attempting to access "barely legal" pornography. *Sexual Abuse* 34(1): 106–124. https://doi.org/10.1177/10790632211013809

Reddit nd. Transparency report 2020. https://www.redditinc.com/policies/transparency-report-2020

Salter M et al. 2021. Production and distribution of child sexual abuse material by parental figures. *Trends & issues in crime and criminal justice* no. 616. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti04916

Salter M & Hanson E 2021. "I need you all to understand how pervasive this issue is": User efforts to regulate child sexual offending on social media. In J Baily, A Flynn & N Henry (eds), *The Emerald International handbook of technology facilitated violence and abuse*. Emerald Publishing: 729–748. https://doi.org/10.1108/978-1-83982-848-520211053

Schiemer J 2018. Strong and responsible: Can encryption be both? *FlagPost*, 3 October. https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/FlagPost/2018/October/Encryption

Seto MC & Eke AW 2015. Predicting recidivism among adult male child pornography offenders: Development of the child pornography offender risk tool (CPORT). *Law and Human Behaviour* 39(4): 416–429. https://doi.org/10.1037/lhb0000128

Signal 2018. Signal partners with Microsoft to bring end-to-end encryption to Skype. *Signal*, 11 January. https://signal.org/blog/skype-partnership/

Snap Inc 2021a. Transparency report: January 1, 2020 – June 30, 2020. https://snap.com/en-US/privacy/transparency/2020-6-31

Snap Inc 2021b. Transparency report: July 1, 2020 – December 31, 2020. https://snap.com/en-US/privacy/transparency/2020-12-31

Statista 2022a. Number of Instagram users worldwide from 2019 to 2023 (in millions). https://www.statista.com/statistics/183585/instagram-number-of-global-users/

Statista 2022b. Number of monthly active Facebook users worldwide as of 4th quarter 2021 (in millions). Hamburg: Statista. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Statista 2022c. Most popular global mobile messenger apps as of January 2022, based on number of monthly active users (in millions). https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/

Steel CMS 2015. Web-based child pornography: The global impact of deterrence efforts and its consumption on mobile platforms. *Child Abuse and Neglect* 44: 150–158. https://doi.org/10.1016/j.chiabu.2014.12.009

Stevens D 2020. Consumers seek out apps with enhanced privacy features to keep in touch in our new normal. https://www.data.ai/en/insights/market-data/consumers-seek-enhanced-privacy-app-features/

Synstrom K 2014. 300 million: Sharing real moments. https://web.archive.org/web/20141210200422/http://blog.instagram.com/post/104847837897/141210-300million

Thorn 2021. Why an increase in reports of CSAM is actually a good thing. https://www.thorn.org/blog/why-an-increase-in-reports-of-csam-is-actually-a-good-thing/

Thorn 2016. Microsoft's PhotoDNA: Leading the fight against child sexual abuse imagery. https://www.thorn.org/blog/photodna-leads-fight-against-child-sex-abuse-imagery/

TikTok 2022. Privacy policy. https://www.tiktok.com/legal/privacy-policy

TikTok 2021. Community guidelines enforcement report: July 1, 2020 – December 31, 2020. https://www.tiktok.com/safety/resources/transparency-report-2020-2

Titcomb J 2019. Snapchat adds end-to-end encryption to protect users' messages. *The Telegraph*, 10 January. https://www.telegraph.co.uk/technology/2019/01/09/snapchat-adds-end-to-end-encryption-protect-users-messages/

Twitter 2021a. Rules enforcement Jan–Jun 2020. https://transparency.twitter.com/en/reports/rules-enforcement.html#2020-jan-jun

Twitter 2021b. Rules enforcement Jul–Dec 2020.
https://transparency.twitter.com/en/reports/rules-enforcement.html#2020-jul-dec

Vincent J 2018. Skype starts testing new 'private conversations' with end-to-end encryption. *The Verge*, 11 January.
https://www.theverge.com/2018/1/11/16878596/microsoft-skype-end-to-end-encryption-private-conversations

Wakefield J 2021. Apple delays plan to scan iPhones for child abuse. *BBC News*, 3 September.
https://www.bbc.com/news/technology-58433647

WeProtect Global Alliance 2019. *Global threat assessment 2019*. London: WeProtect Global Alliance.
https://www.weprotect.org/issue/global-threat-assessment/

WhatsApp 2021. How WhatsApp helps fight child exploitation.
https://faq.whatsapp.com/general/how-whatsapp-helps-fight-child-exploitation/

WhatsApp 2020. *Two billion users – Connecting the world privately.*
https://blog.whatsapp.com/two-billion-users-connecting-the-world-privately

**Coen Teunissen is a Senior Research Analyst at the Australian Institute of Criminology.**

**Sarah Napier is the Research Manager of the Online Sexual Exploitation of Children Research Program at the Australian Institute of Criminology.**

General editor, *Trends & issues in crime and criminal justice* series: Dr Rick Brown, Deputy Director, Australian Institute of Criminology. Note: *Trends & issues in crime and criminal justice* papers are peer reviewed. For a complete list and the full text of the papers in the *Trends & issues in crime and criminal justice* series, visit the AIC website at: aic.gov.au

ISSN 1836-2206 ISBN 978 1 922478 63 4 (Online)
https://doi.org/10.52922/ti78634



GPO Box 1936
Canberra ACT 2601, Australia
Tel: 02 6268 7166



aic.gov.au