

OFFICIAL



Australian Government  
Australian Institute of Criminology

AIC reports  
Submission

---

# Submission to the Parliament of South Australia Select Committee on Artificial Intelligence

Australian Institute of Criminology

OFFICIAL

**Table of contents**

Introduction .....	3
Background .....	3
How will generative AI impact image-based abuse and child sexual exploitation? .....	4
Implications for image-based abuse .....	4
Implications for online sexual exploitation of children .....	5
Other future risks to children from AI technologies .....	5
What actions should be taken to reduce future harm? .....	6
Summary and conclusion .....	7
References .....	8

## Introduction

Thank you for the opportunity to provide a submission to the Select Committee on Artificial Intelligence.

The Australian Institute of Criminology (AIC) is Australia's national research and knowledge centre on crime and justice. We seek to promote justice and reduce crime by undertaking and communicating evidence-based research to inform policy and practice. We have a strong history of producing empirical research into key crime types, including cybercrime, child sexual exploitation and abuse, domestic and family violence and homicide.

This Submission addresses the following point in the Terms of Reference of the inquiry: '*c. issues surrounding the use of AI in the commission of criminal offences*'. Specifically, the Submission will focus on two key crime types that have severe adverse outcomes for victims, and are likely to be significantly impacted by generative AI: image-based abuse, and the online sexual exploitation of children.

## Background

While the internet and enhanced technologies have made our lives easier in so many ways, we also know they have a dark side that fuels criminal activity, ranging from drug dealing (Broadhurst, Ball & Trivedi 2020), money laundering (UNODC nd) and fraud (Emami, Smith & Jorna 2019), to human trafficking (Campana 2022), sexual violence (Wolbers et al. 2022) and child sexual exploitation (Phelan 2022) to name just a few.

Image-based abuse is one such crime that has emerged with the evolution of the internet and other technologies, and occurs when intimate or sexual photos or videos are shared online without consent (Office of the eSafety Commissioner 2017). Madigan et al. (2018) conducted a systematic review and meta-analysis of 39 studies (n = 110,380) that examined the prevalence of 'sexting' (defined as sexually explicit images, videos, or messages) among youth (age range 11.9–17 years), published from 1990–2016. The study found that 12% of respondents had forwarded a sext without consent from the recipient; and 8.4% reported having a sext forwarded without their consent.

This issue has also emerged in the Australian context, among adult and adolescent samples. The Office of the eSafety Commissioner (2017) conducted a survey of 4,122 Australians aged 15-45 years, of whom one in ten (11%) reported they had a nude or sexual photo/video posted online without their consent. Of these, five percent reported that a photo of them was Photoshopped or altered to represent them in a sexual way. Respondents in the study reported negative impacts from their image-based abuse victimisation, including feeling annoyed (65%), humiliated (55%), depressed (40%), and experiencing fear over the discovery of the images by friends (51%), family (48%), an employer (41%), an intimate partner (40%), or children (39%).

In an AIC survey of 9,987 dating app users in Australia, three in four (72%) respondents said they had experienced online sexual violence and harassment by another dating app user, with almost one in five (19%) reporting that they had been subjected to online image-based abuse of a sexual nature (Wolbers et al. 2022).

Another crime type that is facilitated by the internet and advances in technologies is the online sexual exploitation of children. The internet and enhanced technologies have been found to not only make the problem larger but also made the nature of the problem worse, with children abused in more severe ways (Napier, Teunissen & Boxall 2021). The short history of the internet has shown that each time a new technology is introduced to make our lives easier it is quickly adopted for child abuse. In the pre-World Wide Web era, Bulletin Board Systems (a rudimentary type of online forum) introduced in the early 1980s were quickly adopted as a means of sharing child sexual abuse material. Later, websites became an easy way to widely distribute CSAM with paywalls sometimes being used to monetise its use (Kelly & de Castella 2012). Similar early adoption for CSAM

distribution has occurred with peer-to-peer networks (US Department of Justice 2003), instant messaging (Goggin 2022), cloud storage (Worthington 2022), livestreaming (Brown, Napier & Smith 2020) and social media (Teunissen & Napier 2022), creating a very large virtual whack-a-mole environment for law enforcement agencies charged with preventing and investigating child sexual exploitation.

Currently, there is evidence that the number of child sexual abuse material (CSAM) files detected on online platforms is growing. In 2021, the US National Center for Missing and Exploited Children (NCMEC) received over 29 million reports of suspected child sexual exploitation via its CyberTipline, a 35 percent increase on the previous year (NCMEC 2022). In 2022 that number increased again by nine percent to 32 million reports, the highest number ever received, and of which 31.8 million related to CSAM detected on the platforms of electronic service providers (NCMEC 2023). Social media platforms like Facebook, Instagram and Snapchat contributed the largest number of reports.

Sexual extortion is another fast-emerging form of online child sexual exploitation, and involves a child being blackmailed to provide money or other items, with the threat of distributing sexual images of the victim. This crime has become more common in the past two years with perpetrators often engaging with multiple victims simultaneously. The Australian Centre to Counter Child Exploitation said it recorded more than 100 reports of sexual extortion against children on average every month in 2022, which was a 100-fold increase from the previous year (ACCCE 2022).

Victimisation surveys indicate that recording of child sexual abuse and sharing of the imagery causes victim-survivors additional and unique impacts to the contact sexual abuse. This can include trauma, psychological harm, fear, being forever haunted by the images/videos, guilt and shame (Gewirtz-Meydan et al. 2018; Salter et al. 2021). Victims depicted in CSAM can feel repeatedly victimised each time someone views their abusive material (Gewirtz-Meydan et al. 2018), and constantly worried about being recognised by someone who had viewed the CSAM showing their abuse, and in some cases, are actually recognised (Canadian Centre for Child Protection 2017).

Image-based abuse and online sexual exploitation of children are two extremely harmful forms of offending that were already negatively impacting large numbers of victims (Gewirtz-Meydan et al. 2018; Office of the eSafety Commissioner 2017), prior to 2022 when open source generative AI became more widely available (Theil, Stroebel & Portnoff 2023). As generative AI tools are enhanced and adopted by a broader range of users, the implications for image-based abuse and online sexual exploitation of children are particularly concerning, and will be discussed below.

## How will generative AI impact image-based abuse and child sexual exploitation?

### Implications for image-based abuse

In August 2023, Australia's eSafety Commissioner publicly announced the first reports of children using AI to generate sexually explicit images of other children, for the purposes of bullying and humiliation (Long 2023). Speaking to the media, the Commissioner Julie Inman Grant said that while the number of these specific reports are currently small, 'they are concerning, and we know this is just the tip of the iceberg as the technology becomes more sophisticated and widespread' (Long 2023).

Only one month prior, US-based NGO Thorn released a report with Stanford University on the implications of CSAM and generative AI (Theil, Stroebel & Portnoff 2023). They noted that in the short period during the first few months of 2023, there were already several developments to generative AI technologies that augmented end-user control over generation of images and their level of realism, with some 'fake' adult images now very difficult to distinguish from real ones. The report noted that 'near-realistic adult content is currently distributed online in the public and private web and chat forums' (Theil, Stroebel & Portnoff 2023: 2).

In an Australian survey conducted by the AIC in early 2023, 13,887 respondents were asked about cybercrime victimisation, of whom 1.7% (n=238) said that someone had created fake videos or photos of the respondent (i.e. 'deep fakes'; Voce & Morgan 2023). The definition used in the survey was broad and may have captured participant experiences involving manually manipulated images (e.g. Photoshop), as was reported by the eSafety survey (Office of the eSafety Commissioner 2017).

Unfortunately, however, we may see an increase in such reports in the future. Offenders who previously used manual methods to produce this type of abusive content now have increasing access to sophisticated generative AI technologies, which can decrease the effort and time spent offending, allowing them to offend at a larger scale. This would counteract a key element of situational crime prevention, a theory and practice widely adopted by government and other sectors to reduce crime (Wortley & Smallbone 2012). The element of situational crime prevention that is of particular relevance to generative AI, suggests that increasing the effort and risk for offenders will deter them from offending. Thus, if the generation of 'deep fake' imagery becomes increasingly easy and rapid through enhanced AI technologies, while simultaneously lacking in appropriate intervention and safety protocols, image-based abuse involving adult and child victims will likely increase notably in the future. Given the associated harms highlighted in the Background section of this Submission, this could result in burgeoning rates of distress and mental health problems among individuals in the community, in turn augmenting burden on medical and clinical services.

### **Implications for online sexual exploitation of children**

Generative AI offers another opportunity for a new technology to be appropriated for online sexual exploitation of children, including CSAM production and consumption. Early in 2023, the US-based NGO NCMEC received reports of 'fake' CSAM that had been produced by offenders with the assistance of generative AI tools (Murphy 2023). Similarly, in a recent study, Thorn found that less than one percent of CSAM files that were shared in a sample of online networks were AI-generated (Theil, Stroebel & Portnoff 2023). However, they found that this proportion had increased since August 2022.

In a report released with Stanford University, Thorn stated that the use of generative AI in producing realistic CSAM, as well as non-consensual adult sexual content, is growing and likely to continue to do so without intervention by a diverse range of stakeholders across multiple sectors (Theil, Stroebel & Portnoff 2023). The authors suggest future law enforcement investigations and court trials for CSAM cases may be further hampered by AI-generated CSAM being indistinguishable from real CSAM. Furthermore, the authors suggested that on the horizon we are likely to see CSAM offenders using generative AI to produce realistic CSAM in high quality video format.

Worryingly, while there is evidence that offenders are already using generative AI to produce CSAM, there are numerous ways in which AI could in future be used for child sexual exploitation purposes. Some key potential scenarios are outlined below.

### ***Other future risks to children from AI technologies***

Chat bots could be used to groom children for CSAM and sexual extortion, as recently highlighted by the eSafety Commissioner (Butler 2023). Chat-bots could be trained to find young people online, strike up a conversation with them, gain their trust and then request them to provide sexual images or videos. The AI could be trained to take this a step further by then extorting more images from victims under the threat of sharing the original images with friends and family. While reports of sexual extortion have increased notably in recent years (ACCCE 2022), this perpetration could be automated to industrial-scale CSAM production with the use of AI, leaving countless victims in its wake.

Children who are not victims of CSAM can become victims, though using generative AI to produce fake CSAM by altering non-sexual images of real children stolen online. Furthermore, offenders could use this process to expedite and enhance the sexual extortion process. In 2023, the US Federal Bureau of Investigation issued a statement citing an increase in sexual extortion victims reporting

that fake sexual content is being generated from their real photos (Garriss & DeMarco 2023). The report stated that photos that offenders took from victims' social media and other online posts are being altered using AI and used to sexually extort the victims.

Beyond sexual extortion, the number of AI-generated CSAM images distributed online could increase so much so that it overburdens NGO database custodians, moderators, law enforcement, and other parties, resulting in less efficient detection and removal of material and disruption of offending (Theil, Stroebe & Portnoff 2023).

AI could be used to curate existing content. Offenders could command an AI application to create a compilation video of children of a preferred age and gender being abused in certain ways. AI-driven web scraping tools could be used to harvest existing CSAM from the internet and combine them into an edited compilation with video-editing tools.

Child victims depicted in real CSAM are at increased risk of being re-victimised, from offenders using generative AI to alter existing CSAM so that the child is abused in different and more specific ways, including violent and other sadistic forms (Theil, Stroebe & Portnoff 2023).

CSAM could be accidentally created by individuals attempting to create adult content, given CSAM is widely available on the internet (NCMEC 2022) and may be inadvertently used to train AI models.

## What actions should be taken to reduce future harm?

It is with the utmost urgency that all possible effort be made to intervene in the production of fake adult sexual content and CSAM that is generated by AI tools. In collaboration with Stanford University, Thorn (Theil, Stroebe & Portnoff 2023) suggested a number of ways to do this. On 18 August 2023 the Office of the eSafety Commissioner released an AI position statement, which similarly provides a list of actions that should be taken to combat the issue (Office of the eSafety Commissioner 2023). Some key recommendations from both reports are outlined below.

- Biasing AI models against child nudity.
- Watermarking and content provenance (to distinguish fake from real content).
- Passive detection mechanisms (technologies that identify fake content).
- Active monitoring of AI-generated CSAM production networks (material collected and hashes stored in databases, similar to real CSAM).
- Changes to industry CSAM classifications (to include AI-generated content as a separate category).
- Technical collaboration (large tech companies to collaborate on technologies to detect and remove AI-generated content).
- AI ethics and safety by design (incorporate safety by design elements at every stage of the AI development cycle). Safety by design should include:
  - service provider responsibility;
  - user empowerment and autonomy; and
  - transparency and accountability.
- Planning for future advances in generative AI.
- Prevention (education for children, parents, and professionals).
- Protection through legislation (powers to hold tech companies accountable).
- Proactive and systemic change (tech companies to regularly report on actions for reducing harm).

## Summary and conclusion

Key issues arising from this Submission are that:

- Image-based abuse and online sexual exploitation of children has severe negative impacts on victims.
- Generative AI is already being used to produce fake but realistic CSAM, to sexually extort children, and to produce non-consensual fake sexual content of real adults and children.
- The adult content produced by generative AI is currently difficult to distinguish from real imagery, and may be indistinguishable in the near future.
- Without appropriate safety by design protocols, AI-generated CSAM will likely follow a similar pattern to adult content, of increased prevalence and realism.
- Victim-survivors of real CSAM are at great risk of being re-victimised by offenders who re-generate their abusive content available online using generative AI, to change the nature or increase the severity of offending depicted in images.
- AI has the ability to produce CSAM and non-consensual sexual adult content rapidly, and to target multiple children and adults for sexual extortion at a much faster rate than a human can.
- In the future we are likely to see AI generated CSAM in the form of full motion, high quality and realistic videos.
- Without appropriate action from a diverse range of stakeholders across multiple countries and sectors, the sexual crimes committed against children and adults through the use of generative AI will likely grow and become more severe.
- Outlined in this Submission are a number of actions that can be taken to reduce the harm associated with image-based abuse and online sexual exploitation of children, in the context of generative AI.

There are no doubt many other concerning possibilities that involve the use of generative AI in crime perpetration, which have not been covered in this Submission. But those explored here are either already occurring, or would seem only a small step away with the existing tools available. Importantly, it is crucial that technology companies take early and decisive action to prevent further harm to individuals that, without intervention, will arise faster than we are prepared for.



## References

- Australian Centre to Counter Child Exploitation. AFP and AUSTRAC target offshore sextortion syndicates preying on Australian youth. ACCCE: Canberra. <https://www.acce.gov.au/news-and-media/media-release/afp-and-austrac-target-offshore-sextortion-syndicates-preying-australian-youth>.
- Broadhurst R, Ball M & Trivedi H 2020. Fentanyl availability on darknet markets. *Trends & issues in crime and criminal justice* no. 590. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti04244>
- Brown R, Napier S & Smith R 2020. Australians who view live streaming of child sexual abuse: An analysis of financial transactions. *Trends & issues in crime and criminal justice* no. 589. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti04336>
- Butler J 2023. AI tools could be used by predators to 'automate child grooming', eSafety commissioner warns. *The Guardian*, 20 May 2023. <https://www.theguardian.com/technology/2023/may/20/ai-tools-could-be-used-by-predators-to-automate-child-grooming-esafety-commissioner-warns>
- Campana P 2022. *Online and technology-facilitated trafficking in human beings: Full report*. GRETA. <https://rm.coe.int/online-and-technology-facilitated-trafficking-in-human-beings-full-rep/1680a73e49>
- Canadian Centre for Child Protection 2017. *International Survivors' Survey*. Canadian Centre for Child Protection: Winnipeg, Canada. <https://protectchildren.ca/en/resources-research/survivors-survey-results/>
- Emami C, Smith R & Jorna P 2019. Predicting online fraud victimisation in Australia. *Trends & issues in crime and criminal justice* no. 577. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti04015>
- Garriss K & DeMarco N 2023. FBI warns of using AI deepfakes as part of sextortion schemes. *Yahoo! News*, 6 July. [https://news.yahoo.com/fbi-warns-using-ai-deepfakes-212115046.html?guccounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\\_referrer\\_sig=AQAAAJ3MPPPJb7k9RVqnT\\_4XfP6lvFzYx--ZxqNwOcNgwSfemNSHfemSrIDSf9BgTjOZgHch9I-IL4L1ScLM-6Kdh0NaUsshtH0EXeOGAOahZeINjeRgMPy0dhEzvFLLezquO0P7INvo5k3Gc8TLBPU10\\_k8BYYPeoP4k\\_7n6l7U2tKy](https://news.yahoo.com/fbi-warns-using-ai-deepfakes-212115046.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAJ3MPPPJb7k9RVqnT_4XfP6lvFzYx--ZxqNwOcNgwSfemNSHfemSrIDSf9BgTjOZgHch9I-IL4L1ScLM-6Kdh0NaUsshtH0EXeOGAOahZeINjeRgMPy0dhEzvFLLezquO0P7INvo5k3Gc8TLBPU10_k8BYYPeoP4k_7n6l7U2tKy)
- 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>
- Goggin B 2022. Wickr, Amazon's encrypted chat app, has a child sex abuse problem – and little is being done to stop it. *NBC News*, 11 June. <https://www.nbcnews.com/tech/tech-news/wickr-amazon-aws-child-messaging-app-sex-abuse-problem-rcna20674>
- Kelly J & de Castella T 2012. Paedophile net: Did Operation Ore change British society? *BBC News*, 17 December. <https://www.bbc.com/news/magazine-20237564>
- Long C 2023. First reports of children using AI to bully their peers using sexually explicit generated images, eSafety commissioner says. *ABC News*. 16 August. <https://www.abc.net.au/news/2023-08-16/esafety-commissioner-warns-ai-safety-must-improve/102733628>
- Madigan S, Ly A, Rash CL, Van Ouytsel J & Temple JR 2018. Prevalence of Multiple Forms of Sexting Behavior Among Youth: A Systematic Review and Meta-analysis. *JAMA Pediatrics*, 172(4): 327–335. <https://doi.org/10.1001/jamapediatrics.2017.5314>
- Murphy M 2023. Predators exploit AI tools to generate images of child abuse. *Bloomberg*, 23 May. <https://www.bloomberg.com/news/articles/2023-05-23/predators-exploit-ai-tools-to-depict-abuse-prompting-warnings#xj4y7vzkg>
- Napier S, Teunissen C & Boxall H 2021. Live streaming of child sexual abuse: An analysis of offender chat logs. *Trends & issues in crime and criminal justice* no. 639. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti78375>
- NCMEC 2023. *2022 CyberTipline Reports by Electronic Service Providers (ESP)*. National Center for Missing & Exploited Children. <https://www.missingkids.org/content/dam/missingkids/pdfs/2022-reports-by-esp.pdf>
- NCMEC 2022. *CyberTipline 2021 Report*. National Center for Missing & Exploited Children. <https://www.missingkids.org/content/dam/missingkids/pdfs/2021-CyberTipline-Report.pdf>



- Office of the eSafety Commissioner 2023. *Tech Trends Position Statement - Generative AI*. Office of the eSafety Commissioner. <https://www.esafety.gov.au/industry/tech-trends-and-challenges/generative-ai>
- Office of the eSafety Commissioner 2017. *Image-based Abuse National Survey: Summary Report*. Office of the eSafety Commissioner: <https://www.esafety.gov.au/sites/default/files/2019-07/Image-based-abuse-national-survey-summary-report-2017.pdf>
- Phelan M 2022. *Crime & justice research 2022: Online sexual exploitation of children*. Special reports. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/sp78658>
- Salter M, Wong W T, Breckenridge J, Scott S, Cooper S & Peleg N 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>
- Teunissen C & Napier S 2022. Child sexual abuse material and end-to-end encryption on social media platforms: An overview. *Trends & issues in crime and criminal justice* no. 653. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti78634>
- United Nations Office on Drugs and Crime nd. *Money laundering through cryptocurrencies*. UNODC. <https://syntheticdrugs.unodc.org/syntheticdrugs/en/cybercrime/laundryingproceeds/moneylaundering.html>
- United States Department of Justice 2003. File-sharing programs: Peer-to-peer networks provide ready access to child pornography. <https://www.ojp.gov/ncjrs/virtual-library/abstracts/file-sharing-programs-peer-peer-networks-provide-ready-access-child>
- Voce I & Morgan A 2023. *Cybercrime in Australia 2023*. Statistical Report no. 43. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/sr77031>
- Wolbers H, Boxall H, Long C & Gunnoo A 2022. *Sexual harassment, aggression and violence victimisation among mobile dating app and website users in Australia*. Research Report no. 25. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/rr78740>
- Worthington B 2022. Microsoft and Apple among the global companies accused of ‘turning a blind eye’ to child sexual exploitation. ABC News, 15 December. <https://www.abc.net.au/news/2022-12-15/microsoft-apple-child-sexual-exploitation-esafety/101771844>
- Wortley R & Smallbone S 2012. *Internet Child Pornography: Causes, Investigation and Prevention (Vol. 1)*. ABC CLIO.

**Authors:**

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

**Dr Rick Brown is the Deputy Director of the Australian Institute of Criminology**