



Australian Government

Australian Institute of Criminology

# Trends & issues in crime and criminal justice

No. 711 January 2025

**Abstract** | This study examined the intersection of artificial intelligence (AI) and child sexual abuse (CSA), employing a rapid evidence assessment of research on the uses of AI for the prevention and disruption of CSA, and the ways in which AI is used in CSA offending. Research from January 2010 to March 2024 was reviewed, identifying 33 empirical studies.

All studies that met inclusion criteria examined AI for CSA prevention and disruption—specifically, how technology can be used to detect or investigate child sexual abuse material or child sexual offenders. There were no studies examining the uses of AI in CSA offending.

This paper describes the state of current research at the intersection of AI and CSA, and provides a gap map to guide future research.

## Artificial intelligence and child sexual abuse: A rapid evidence assessment

Heather Wolbers, Timothy Cubitt and Michael John Cahill

### Introduction

In recent years, the development of artificial intelligence (AI) has rapidly increased, with availability of this technology significantly expanding. Development of AI technologies has extended to the field of child sexual abuse (CSA), with the scope and scale of the problem—particularly online—becoming too great for manual human-led approaches to manage effectively. However, the use of AI in this field has extended beyond prevention. In early 2023, the US National Center for Missing and Exploited Children received reports of ‘fake’ child sexual abuse materials (CSAM) that offenders had produced with the assistance of generative AI tools (Murphy 2023). Similarly, Australia’s eSafety Commissioner has noted reports of children using AI to generate sexually explicit images of other children, suggesting that it was an indication of a more widespread issue (Long 2023).

This review considers the current state of research literature studying the use of AI in the field of CSA, focusing on studies investigating the use of AI for offending and the prevention and disruption of CSA.

## Artificial intelligence for child sexual abuse offending

According to media reports, surveys and academic reviews, AI technologies are increasingly playing diverse roles in the creation of CSAM, including fabrication (eg CSAM deepfakes), ‘nudifying’ pictures of clothed children, and manipulating images or videos to depict known or unknown children in sexually abusive scenarios (Milmo 2023; Okolie 2023). Existing CSAM has been used to train AI models, meaning offenders are using AI to produce new depictions of previously abused children. Further, reports indicate that offenders are using AI to alter photos from victims’ social media and other online posts and using these altered images to sexually extort the victims (Garriss & DeMarco 2023). According to a survey of 1,040 people aged nine to 17 years in the United States, one in 10 (11%) minors said they knew of cases where their peers had used AI to create sexually explicit images of other minors (Thorn 2024b).

The Internet Watch Foundation found that, in a single month, 20,254 AI-generated images were posted to a CSAM forum on the darknet (Internet Watch Foundation 2023). Concerns have been raised that the ability to generate CSAM using AI could support an increase in CSAM consumption. Growth in AI-generated CSAM creates significant challenges for law enforcement, who work to detect and prevent the distribution of CSAM online. Ultimately, increases in the volume of CSAM online may influence the ability to investigate CSA, as AI-generated can be indistinguishable from real CSAM (Theil, Stroebel & Portnoff 2023). The malicious use of AI technologies for the production of CSAM is growing and is likely to continue to grow without multi-sector intervention (Theil, Stroebel & Portnoff 2023).

## Artificial intelligence for the prevention and disruption of child sexual abuse

As identified in academic research, there are a diverse range of cyber strategies used to combat online CSA (Edwards et al. 2021; Singh & Nambiar 2024). As the field of AI continues to develop, so too does the development of CSA disruption strategies that use AI. For example, published research has shown that AI could be used to identify suspicious financial transactions procuring CSA (eg Cubitt, Napier & Brown 2021; Henseler & de Wolf 2019) or aid in law enforcement investigations by examining CSAM (eg Brewer et al. 2023; Dalins et al. 2018; Westlake et al. 2022). AI technologies may ease the burden on law enforcement by reducing the risk of psychological harms among CSA investigators (Puentes et al. 2023), while increasing the capacity and timeliness of investigations. Further, AI technologies can have a much larger reach across online spaces than traditional methods of prevention and disruption.

While CSAM can be detected and removed across online spaces with hashes (ie unique digital fingerprints), this method is limited to known CSAM on platforms proactively using hashes, meaning there is limited efficacy and it cannot stop the upload and proliferation of undetected, new or edited content. AI has the potential to help address this challenge. For instance, Thorn has developed a machine learning tool to automatically detect, review and report CSAM at scale (Thorn 2024a). This tool is used to screen new content that gets uploaded to Flickr and other online platforms—a task too large for human moderation alone. Beyond detecting CSAM, AI has a range of potential uses for addressing CSA. These include conversation analysis, chatbots, honeypots and web crawlers, all of which show promise in combating CSA, albeit with some significant limitations (eg narrow scope or generalisability, privacy and legal concerns, and lack of robust evaluation; Singh & Nambiar 2024).

## Study aims

AI is the ability of a computer system to simulate human intelligence, such as learning, problem solving, reasoning and decision making—all with some level of autonomy (High-Level Expert Group on Artificial Intelligence 2019). There are several domains of AI, including machine learning (in which an algorithm is trained to learn patterns in existing data), computer vision (interpreting visual information), natural language processing (understanding and generating human language), and generative AI (creating original content). We considered the domains of AI to develop search criteria and identify literature examining AI and its intersection with CSA.

This study aims to establish current AI capabilities in relation to CSA, with the intention of identifying key areas of progress and gaps where further research is required. A rapid evidence assessment was conducted to address the following research questions:

- What are the uses of AI as a part of CSA offending and what are the areas of future risk?
- What are the uses of AI as a part of CSA prevention and disruption, and what are the areas of future potential?
- What are the key gaps in current research that should be addressed?

## Method

### Search strategy

Rapid evidence assessments are accelerated systematic reviews of research undertaken in a restrictive time frame. This study draws on research published in English between January 2010 and March 2024. Studies were excluded if they did not discuss AI in the context of CSA, include primary data (ie reviews or conceptual studies), explain the study's methodology in sufficient detail (eg if they did not detail the data source, the sample or data management for analysis), or have a direct application to CSA. We excluded studies examining the use of AI in medical settings to detect CSA, as a systematic review was recently published on this topic, which identified seven studies that examined the use of AI for predicting child abuse and neglect using medical or protective service data (Lupariello et al. 2023). Of note, our search yielded seven studies from medical settings, all of which were excluded for secondary reasons (ie they did not focus on detection or prevention of CSA).

The Australian Institute of Criminology's JV Barry Library searched 13 databases and relevant websites: the JV Barry Library catalogue, the Australian Institute of Criminology, EBSCO, ProQuest Criminal Justice, DeepDyve, arXiv.org, IEEE Xplore digital library, Office of the eSafety Commissioner, International Centre for Missing and Exploited Children, National Center for Missing and Exploited Children, Australian Centre to Counter Child Exploitation, Thorn and Google Scholar.

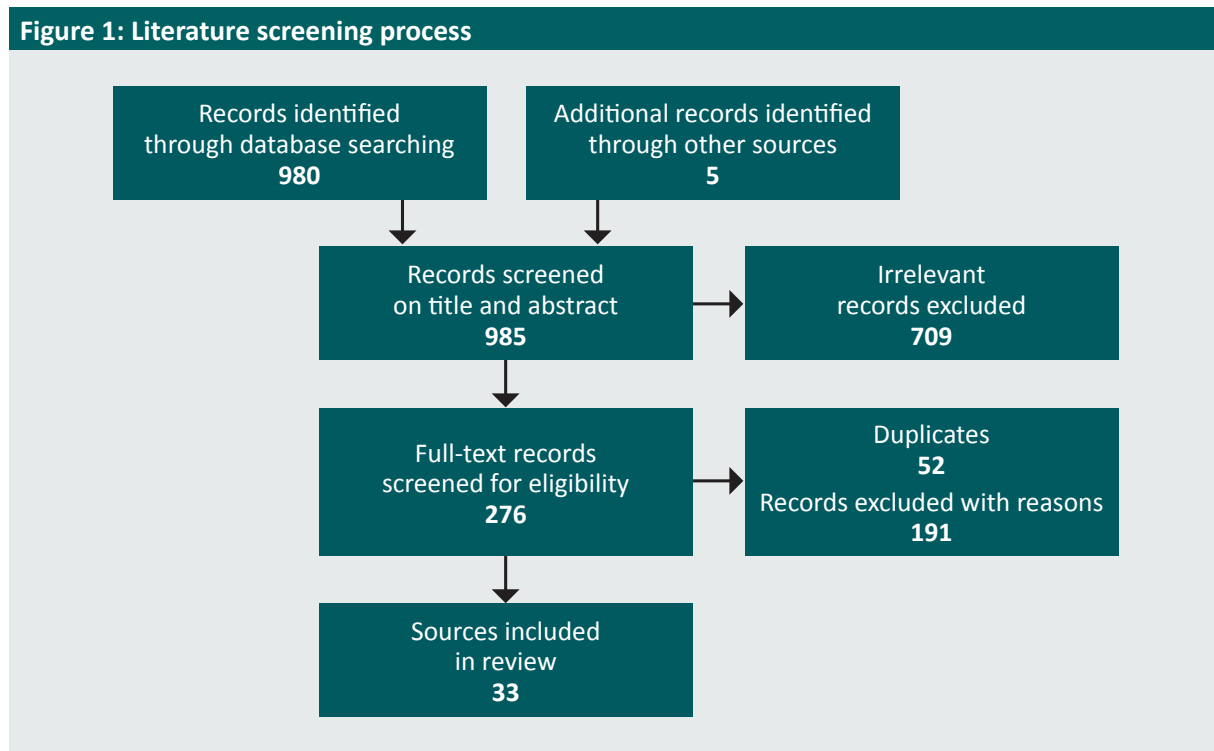
The search terms used combined keywords from three categories capturing AI and its relevant sub-domains, child sexual abuse, and specifying the group of concern to be individuals under 18 years of age:

- *Artificial intelligence* (“machine learning” OR “artificial intelligence” OR algorithm\* OR “deep learning” OR “unsupervised learning” OR “reinforcement learning” OR “generative artificial intelligence” OR “natural language processing” OR “computer vision” OR chatbot OR “image classification” OR “object detection” OR “augmented reality” OR “big data” OR “neural network”);
- *Child sexual abuse* (“child sexual abuse” OR CSAM OR CSEM OR “child abuse material” OR “live streaming” OR “child exploitation” OR “child sexual abuse material” OR “child exploitation material” OR “image-based sexual abuse” OR “image-based abuse” OR “technology-facilitated sexual violence” OR online “sexual exploitation of children” OR “child pornography” OR “indecent images of children” OR grooming); and
- *Child* (child OR children OR “young person” OR “young people” OR adolescent OR teenage\* OR youth OR minor OR “young adult”).

## Screening process

The rapid evidence search yielded 980 records. Five additional studies were identified through wider reading, resulting in identification of an initial 985 records. Titles and abstracts were screened to exclude irrelevant studies ( $n=709$ ). This screening yielded 276 records, of which 52 were identified as duplicates and removed. The remaining 224 sources were assessed for eligibility against the selection criteria with full-text screening, and 191 were excluded because they did not meet the selection criteria. In total, the search yielded 33 sources providing primary information on the role of AI in relation to CSA (see Figure 1).

**Figure 1: Literature screening process**



## Limitations

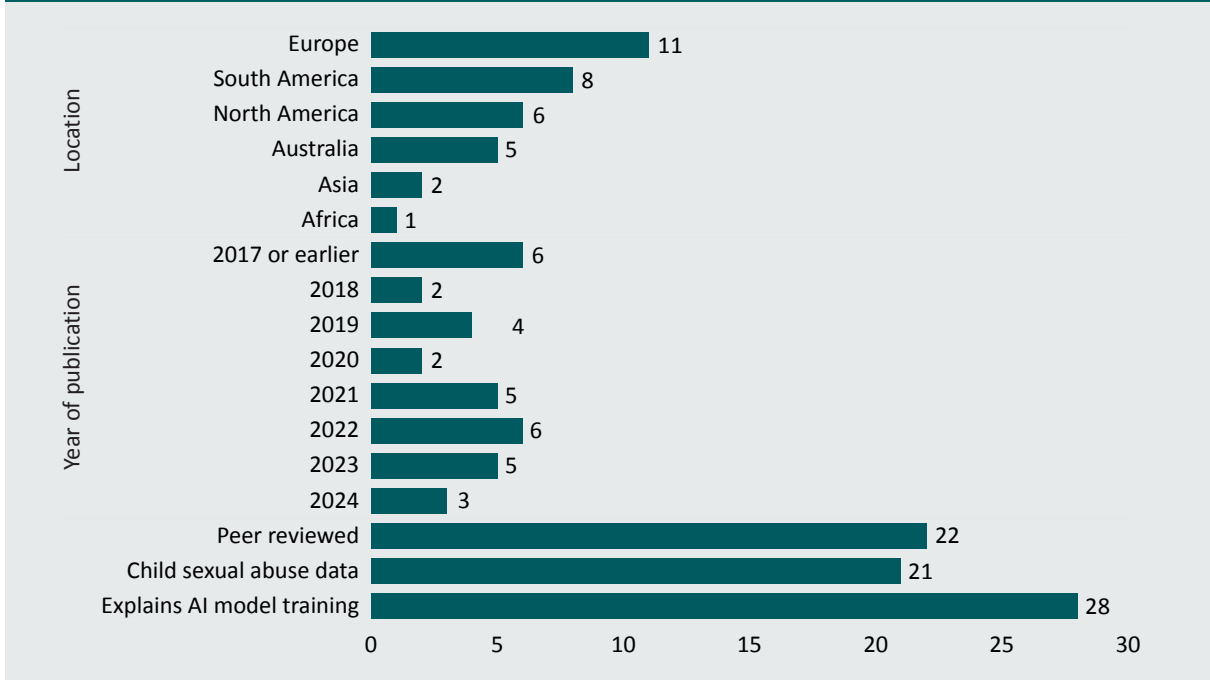
Rapid evidence assessments do not provide the same exhaustive depth or detail as a full systematic review (Ganann, Ciliska & Thomas 2010). Databases that yielded a large number of hits were not searched in full. Rather, the first several hundred items returned by the search were screened, meaning that the most relevant sources were captured. However, search results were not screened exhaustively. Given that search results were presented in order of relevance to the search terms used, the number of sources missed by using this methodology is likely limited.

## Results

### Study characteristics

Research at the intersection of AI and CSA was identified from Europe ( $n=11$ ), South America ( $n=8$ ), North America ( $n=6$ ), Australia ( $n=5$ ), Asia ( $n=2$ ) and Africa ( $n=1$ ). Identified research was published between 2011 to 2024, with a notable increase from 2020 onward. Two-thirds were peer reviewed ( $n=22$ ). Non-peer reviewed sources included conference papers ( $n=9$ ) or pre-prints ( $n=2$ ). While research primarily relied on data from cases of CSA (ie CSAM files, or information on convicted CSA offenders), five studies used data on suspected rather than proven child sexual exploitation (eg reports to a hotline, risky online conversations). Five studies relied on interactions between suspected or known offenders and adults posing as children, and two relied on peripheral datasets using semi-nude non-sexual images of children and a database of faces for age estimation. The majority of included studies explained how their AI model was trained ( $n=28$ ), with five using unsupervised modelling approaches. All identified research examined uses of AI for the prevention, disruption, detection or investigation of CSA, with no studies examining the uses of AI to perpetrate CSA. As reports of malicious use of AI to perpetrate CSA only began to emerge recently (Long 2023; Murphy 2023), it seems unlikely that enough time has passed for empirical research to have been produced on this subject matter.

**Figure 2: Characteristics of included studies (n=33)**



### Quality of evidence regarding efficacy of artificial intelligence in the context of child sexual abuse

Among the 33 reviewed studies, 28 attempted to evaluate the discussed AI tool—typically considering accuracy, precision, recall or another metric specific to the intended goal. When a tool was evaluated, the employed metrics and study aims were diverse, meaning cross-study comparisons of efficacy were not possible. Additionally, the data or sample used for testing—particularly when the data were synthetically produced, or relied on a small or non-generalisable sample—raised questions about how well the model would translate to a real-world setting. Only a minority of studies tested the tool in a real-world or near real-world setting (Brewer et al. 2023; Dalins et al. 2018; Guerra & Westlake 2021; Jin et al. 2024; Ngo et al. 2024; Peersman et al. 2016; Westlake et al. 2022). Due to these limitations, any findings regarding efficacy for each included study should be interpreted with some caution. For example, while a tool may demonstrate a high level of efficacy, this does not necessarily mean the tool would perform well if applied in the real world.

## Artificial intelligence and child sexual abuse

Table 1 presents a summary of the AI tools discussed in published literature. These included tools used for detection, examination or investigation of CSAM or child sexual offenders online.

Table 1: Summary of how artificial intelligence is used in the context of child sexual abuse in the studies reviewed	
Intention of the tool	Approach used
Detect CSAM	Detect CSAM on personal computers using file names and file paths (2, 22, 23)
	Detect CSAM using a combination of tools such as pornography detection, age estimation, skin tone/nudity identifier (8, 9, 16, 21, 22, 24, 25, 29, 32)
Aid in the investigation of CSAM	Separate CSAM into discrete categories by type (8, 15)
	Determine the age of children in CSAM (11)
	Match victims and offenders across CSAM videos using facial and voice recognition (4, 33)
Detect online child sexual offenders	Identify patterns in the locations and folder or file naming practices of websites with CSAM (12)
	Analyse the language used in online conversations to identify threats to children (1, 3, 5, 13, 17, 26, 30)
	Analyse conversations between children and offenders to identify whether offenders intend to contact offend or not (31)
	Distinguish real children from adults pretending to be children in chat rooms (17)
	Using a chatbot to interact with suspects and profile their interest in CSAM (18, 27)
Aid in understanding online child sexual offenders	Scrape hashtags and images from tweets in real time to detect suspected human trafficking of minors (10)
	Crawl the darknet to collect data on the behaviours of child sexual offenders who access and participate on dark websites (14)
	Analyse posts about CSAM and associated metadata to understand the characteristics, behaviours and motivations of CSAM creators (20)
	Understand the characteristics and typologies of offenders who live stream CSA (6, 7)
Other	Detect other files shared online by individuals who have shared known CSAM files (22)
	Analyse text-based reports of child sexual abuse (25)

Note: 1—Agarwal et al. 2022; 2—Al-Nabki et al. 2023; 3—Anderson et al. 2019; 4—Brewer et al. 2023; 5—Cardei & Rebedea 2017; 6—Cubitt, Napier & Brown 2021; 7—Cubitt, Napier & Brown 2023; 8—Dalins et al. 2018; 9—Gangwar et al. 2021; 10—Granizo et al. 2020; 11—Grubl & Lallie 2022; 12—Guerra & Westlake 2021; 13—Isaza et al. 2022; 14—Jin et al. 2024; 15—Laranjeira et al. 2022; 16—Macedo, Costa & dos Santos 2018; 17—Meyer 2015; 18—Murcia Triviño et al. 2019; 19—Ngejane et al. 2021; 20—Ngo et al. 2024; 21—Oronowicz-Jaśkowiak et al. 2024; 22—Peersman et al. 2016; 23—Pereira et al. 2021; 24—Polastro & Eleuterio 2012; 25—Puentes et al. 2023; 26—Razi et al. 2023; 27—Rodriguez et al. 2020; 28—Rondeau et al. 2022; 29—Sae-Bae et al. 2014; 30—Seedall, MacFarlane & Holmes 2019; 31—Seigfried-Spellar et al. 2019; 32—Ulges & Stahl 2011; 33—Westlake et al. 2022



### *Artificial intelligence for detecting and investigating child sexual abuse material*

The most common use of AI in the identified research was to detect CSAM. The primary intention of the AI tools studied was to reduce the burden of manual processing of CSAM by investigators, thereby mitigating mental health impacts, while reducing the time needed to identify CSAM among very large datasets. Tools designed to detect CSAM tended to combine age identifiers (Dalins et al. 2018; Gangwar et al. 2021; Macedo, Costa & dos Santos 2018; Rondeau et al. 2022; Sae-Bae et al. 2014), with skin tone, nudity or pornography detectors (Dalins et al. 2018; Gangwar et al. 2021; Laranjeira et al. 2022; Macedo, Costa & dos Santos 2018; Oronowicz-Jaśkowiak et al. 2024; Peersman et al. 2016; Polastro & Eleuterio 2012; Rondeau et al. 2022; Sae-Bae et al. 2014).

Other models detected CSAM by analysing the words used to describe the picture (Peersman et al. 2016; Ulges & Stahl 2011) or language embedded in audio (Peersman et al. 2016). Three studies discussed tools that analyse file names or file paths to estimate the likelihood of them containing CSAM (Al-Nabki et al. 2023; Peersman et al. 2016; Pereira et al. 2021). Importantly, tools were frequently designed to be implemented in a specific setting, such as on peer-to-peer networking websites (eg Peersman et al. 2016) or on a personal computer or device (eg Polastro & Eleuterio 2012). Many of these studies used authentic CSA data sources and reported that the tool of focus performed well at completing the intended task (Al-Nabki et al. 2023; Gangwar et al. 2021; Oronowicz-Jaśkowiak et al. 2024; Peersman et al. 2016; Pereira et al. 2021; Polastro & Eleuterio 2012). However, others showed limited efficacy (Dalins et al. 2018; Puentes et al. 2023; Ulges & Stahl 2011).

Notably, Peersman and colleagues (2016) described a toolkit that performed multiple functions. While designed primarily to detect CSAM on peer-to-peer networks by analysing filenames, images and audio, the tool also flagged files shared by individuals who have shared known CSAM. This tool was evaluated using real-world CSA case data, demonstrating usefulness in law enforcement settings and considerable accuracy in detecting CSAM when combining filename and image classification.

Beyond the detection of CSAM, several studies discussed uses of AI to assist investigations in other ways. One study aimed to categorise the content of CSAM to aid with triaging (eg solo, non-penetrative, penetrative; Dalins et al. 2018), while another extracted features and labels of CSAM to describe the content without it ever being viewed (Laranjeira et al. 2022). Another important use was to extract and match the biometric features of victims and perpetrators shown in CSAM, allowing for a rapid detection of media associated with an investigation and the identification of links between files (Brewer et al. 2023; Westlake et al. 2022). Finally, a web crawler was designed to find patterns in the locations and CSAM naming conventions of websites with known CSAM (Guerra & Westlake 2021). Three of these studies tested the performance of the AI tool (Dalins et al. 2018; Laranjeira et al. 2022; Westlake et al. 2022), and just one demonstrated strong results. Specifically, Westlake and colleagues (2022) were able to identify and match victims and offenders across a test sample of authentic CSAM files with a high true match rate (between 93.8% and 98.8%) and a low false match rate (between 1.0% and 5.0%).



### *Artificial intelligence for detecting and understanding child sexual abuse offenders*

Two studies described a tool designed to initiate and hold conversations with potential online CSA offenders (ie a chatbot; Murcia Triviño et al. 2019; Rodríguez et al. 2020). This chatbot used generative and rule-based models to produce conversational posts and replies. Based on these interactions, the tool then described each suspect's behaviour, classifying their disposition towards online child sexual offending as indifferent, interested or perverted. The efficacy of this classification was not numerically measured.

Several studies described AI methods designed to understand child sexual offenders through their online behaviours. A tool produced by Granizo and colleagues (2020) scraped posts from X (formerly Twitter) in real time to identify suspected cases of child trafficking. This tool demonstrated some efficacy at recognising the gender and age of individuals depicted in images by analysing their faces or torso.

Two studies discussed tools operating on the darknet. The first study discussed a web crawler that collected data on darknet ecosystems, detecting relevant content and providing information designed to reduce the anonymity of perpetrators, such as links to the surface web that may be used to trace darknet operators (Jin et al. 2024). Similarly, a tool developed by Ngo and colleagues (2024) processed CSAM discussions on the darknet and provided insights into the characteristics, behaviours and motivation of CSAM creators. Both tools performed well in identifying the content of interest and were able to reveal meaningful information about online offending environments and offenders.

Peersman and colleagues (2016) described a model to detect other online files shared by those known to distribute CSAM online, while two studies used machine learning to analyse the characteristics and offending history of individuals who live streamed CSA (Cubitt, Napier & Brown 2023, 2021). In the 2021 study, the model had notable success in identifying individuals who would engage in prolific live streaming of CSA, successfully classifying more than 80 percent of cases (AUROC=0.85; Cubitt, Napier & Brown 2021).

### *Other uses of artificial intelligence in the context of child sexual abuse*

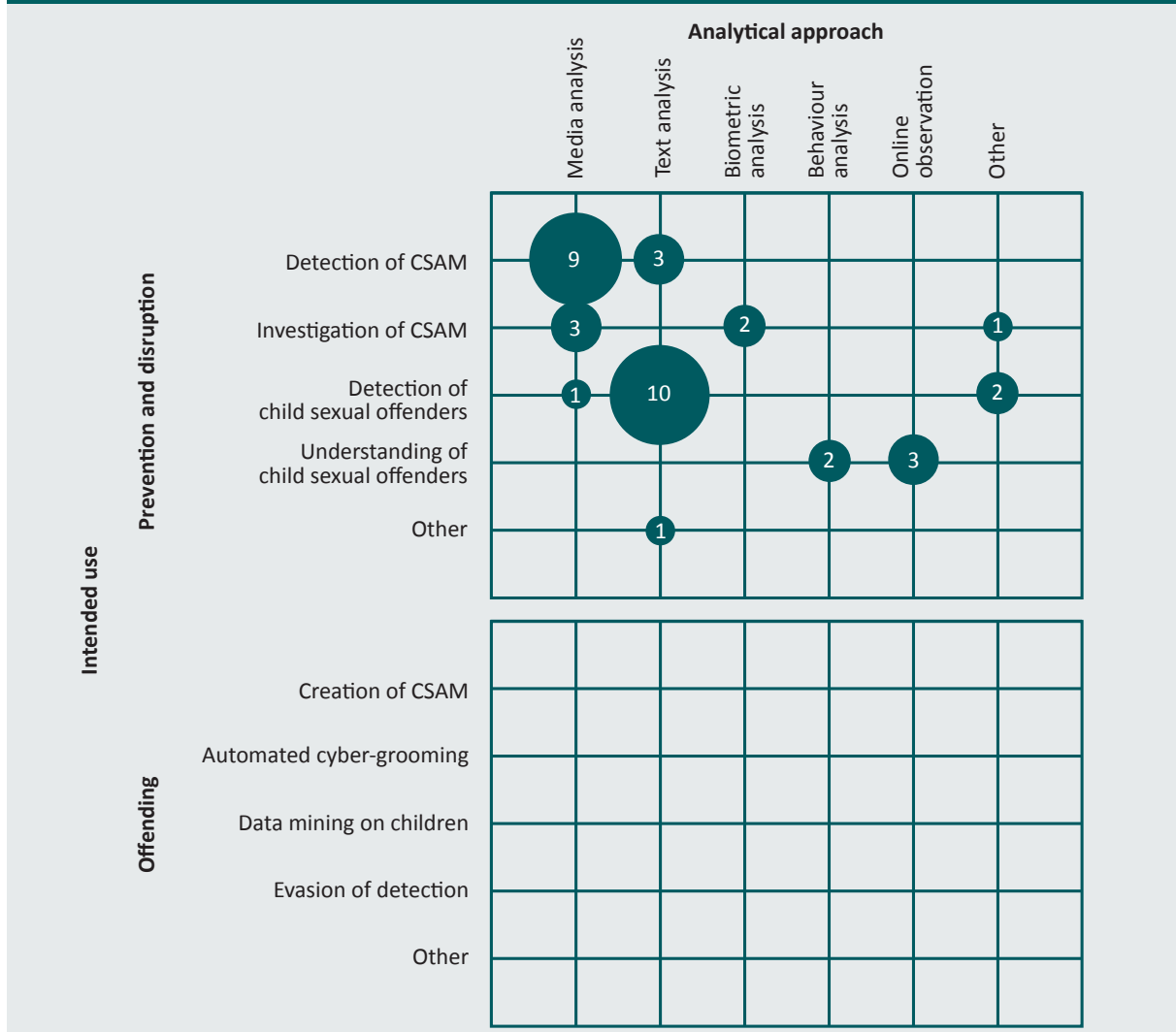
One final tool used a large language model to identify the subject, degree of criminality, and level of impact relating to reports of CSA to a hotline (Puentes et al. 2023). Using this method, the authors aimed to automate the analysis of CSA reports, consequently expediting the process while reducing the exposure of analysts to potentially harmful content. The authors concluded that the approach was an appropriate starting point, but the efficacy was limited.

## Evidence and gap map

We constructed an evidence and gap map plotting the most common analytical domains of AI that were reported, and how they were used in the context of CSA (Figure 3). Importantly, these do not represent all possible AI capabilities and uses in the context of CSA. The analytical procedure and intended use categories broadly capture the most common AI capabilities and uses outlined in the research, but the figure does not critique the quality of evidence or efficacy of the AI technologies discussed. Dots on the graph show where the analytical approach and the intended uses intersect, with the size of the dot reflecting the number of studies. Intersections without dots indicate an absence of evidence, highlighting areas requiring further research.

The evidence and gap map highlighted a significant gap in evidence relating to how AI is used in the perpetration of child sexual offending, with no literature returned in this search. The two most developed areas of research focused on analysis of media (images, audio and videos) to detect CSAM, and text analysis to detect child sexual offenders.

**Figure 3: Evidence and gap map for research on the uses of artificial intelligence in the context of child sexual abuse**



## Discussion







At the time of this review, research has examined AI tools used to detect CSAM or CSA offenders, to aid with investigations, and to improve our understanding of online environments where CSA is produced and shared. The two most common uses of AI were analysing images, audio or video to detect CSAM without requiring humans to view the content, and language processing to detect child sexual offenders through online conversations. Of note, several of the tools described undertook more than one task. For example, some tools were designed to both detect and categorise CSAM (Dalins et al. 2018), to detect CSAM by classifying media content and separately the text of file names (Peersman et al. 2016), or to consider both text and images simultaneously (Granizo et al. 2020).

The introduction of AI technologies in place of human decision-making offers important opportunities (Singh & Nambiar 2024). Benefits include automatically classifying CSAM images and improving the efficiency of detection or classification of images and videos when large volumes of data are obtained. These functions have the potential to reduce the risk of psychological harm to investigators. The automated nature of these tools is particularly important given the demands placed on law enforcement by the recent dramatic growth in CSAM production and sharing. The rate of online sharing and viewing of CSAM is currently beyond manual human detection and intervention. Ultimately, this may mean that a human response alone is not an adequate solution to this increasingly AI driven problem, and opportunities to integrate AI technology into existing CSA prevention strategies should be explored.

## Directions for research

The principal gap in research identified by this review was the use of AI for CSA offending. Further, the studies examined indicated that several of the AI technologies proposed for the prevention or disruption of CSA were not fit for purpose in their current form or did not have sufficient evidence to support their efficacy in a real-world setting. Figure 4 provides a summary of important areas for future research at the intersection of AI and CSA, informed by the evidence and gap map in Figure 3.

**Figure 4: Summary of future research required on artificial intelligence technologies in the context of child sexual abuse**

	Research into the ways that AI is used to commit CSA offences and evade detection, including but not limited to the aggregation and generation of CSAM
	Improving and evaluating existing AI technologies to ensure they are fit for purpose for those who will implement them
	Develop nuanced and appropriate ways to test and implement AI technologies in real-world settings—including the development of ethically sourced datasets that could allow for the training and testing of AI tools
	Continue to build knowledge on CSA and child sexual offender behaviours and characteristics, to inform effective targeting of newly developed AI technologies
	Continued innovation of AI technologies and translation to use for the prevention and disruption of CSA
	Interdisciplinary collaboration in the development of AI technologies with a focus on applied use in regulatory and law enforcement settings

While there is evidence that AI is being used in the process of child sexual offending, particularly CSAM offending, a trend that appears to be expanding (Internet Watch Foundation 2023), there were no identified studies investigating the nature and scope of AI use among offenders. Anecdotal evidence suggests AI is being used to generate deepfake CSAM or could automate cyber-grooming (Butler 2023; Internet Watch Foundation 2023). However, there is limited understanding in the research literature of the scope of this problem or emerging uses of AI for malicious purposes. It is important to implement methods to deter the use of AI for illicit activities. Research should continue to explore the uses of AI among child sexual offenders to better understand the risks posed and ways to address these risks.

While there was a sizable amount of research on the development of AI technologies for CSA prevention and disruption, the research evaluating the efficacy of these models, or how they performed in real-world settings, was limited. It is important to note that tools should only be implemented after their performance has been evaluated. This evaluation should measure the tool's effectiveness (eg accuracy, reliability, specificity), as well as its application (ie whether potential users, such as law enforcement, can employ it and find it helpful). It may also be helpful to measure performance with different data sources or for different tasks, to clarify where models perform well and where they do not.

However, the adoption of AI technology alone for CSA prevention and disruption is unlikely to be a comprehensive solution for either CSA in general or AI produced CSAM. Tech-solutionism, in which technology is implemented as a standalone method of solving a given issue, is often criticised for oversimplifying complex problems, failing to address root causes, and leading to unforeseen negative consequences. Additionally, there are reasonable concerns regarding the use of AI prevention methods without supervision or validation by humans. If AI technology were to be found suitable for use, it should not be adopted at the expense of broader approaches to tackling CSA. Rather, it should be implemented as just one tactic among many to reduce the volume and impact of CSA. Detection and prevention approaches, when featuring AI, should continue to be transparent, should feature a degree of human supervision and should be considered part of a suite of complementary approaches to address CSA, rather than a standalone solution.

Developing and testing AI technologies in the context of CSA prevention may require the use of authentic CSA data sources that reflect real-world settings—for example, CSAM, offender chat logs and police case reports. Of course, there are important ethical and practical implications when building and accessing such data sources. It is appropriate that access to these data are tightly controlled; however, the difficulty accessing data is a significant barrier to assessing whether these AI tools may be useful and implementable. Progress in the field of AI for CSA prevention may therefore require consideration of how researchers can reliably evaluate the efficacy of their tool using appropriate datasets, under agreed upon conditions for access. Additionally, AI models could be advanced by improving knowledge of child sexual offender behaviour (Singh & Nambiar 2024).

AI technology will continue to develop rapidly. While caution should be exercised in implementing these models, innovative approaches to addressing CSA should be encouraged. Each of the directions suggested for future research would substantially benefit from interdisciplinary collaboration, particularly featuring the stakeholders who would ultimately use the technology.

## Conclusion

The research literature has, to date, detailed a range of AI technologies developed with the aim of preventing or disrupting CSA—most commonly, those that detect CSAM or online child sexual offenders. The use of AI to address CSA is an emerging field, and while the evidence for the efficacy of AI technology in this context is limited, the field is moving and developing rapidly. Ultimately, with AI supported CSA offending becoming more widely reported, interest in AI approaches to prevent CSA is likely to grow. This review has emphasised the potential for AI technologies to identify, prevent and disrupt CSA. These technologies offer many advantages but must undergo strict evaluation and safety testing and adhere to ethical protocols before being considered for adoption to complement existing strategies.

## Acknowledgements

This research was conducted as part of the National Office for Child Safety's Child Safety Research Agenda.

## References

URLs correct as at November 2024

\*Included in review

\*Agarwal N, Ünlü T, Wani MA & Bours P 2022. Predatory conversation detection using transfer learning approach. In G Nicosia et al. (eds), *Machine learning, optimization, and data science*. Lecture Notes in Computer Science vol. 13163. Springer International Publishing: 488–499. [https://doi.org/10.1007/978-3-030-95467-3\\_35](https://doi.org/10.1007/978-3-030-95467-3_35)

\*Al-Nabki MW, Fidalgo E, Alegre E & Alaiz-Rodriguez R 2023. Short text classification approach to identify child sexual exploitation material. *Scientific Reports* 13(1): 16108. <https://doi.org/10.1038/s41598-023-42902-8>

\*Anderson P, Zuo Z, Yang L & Qu Y 2019. An intelligent online grooming detection system using AI technologies. *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* New Orleans, LA, USA: IEEE1–6. <https://doi.org/10.1109/FUZZ-IEEE.2019.8858973>

\*Brewer R et al. 2023. Advancing child sexual abuse investigations using biometrics and social network analysis. *Trends & issues in crime and criminal justice* no. 668. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti78948>

Butler J 2023. AI tools could be used by predators to ‘automate child grooming’, eSafety commissioner warns. *The Guardian*, 20 May. <https://www.theguardian.com/technology/2023/may/20/ai-tools-could-be-used-by-predators-to-automate-child-grooming-esafety-commissioner-warns>

\*Cardei C & Rebedea T 2017. Detecting sexual predators in chats using behavioral features and imbalanced learning. *Natural Language Engineering* 23(4): 589–616. <https://doi.org/10.1017/S1351324916000395>

\*Cubitt T, Napier S & Brown R 2023. Understanding the offline criminal behavior of individuals who live stream child sexual abuse. *Journal of Interpersonal Violence* 38(9–10): 6624–6649. <https://doi.org/10.1177/08862605221137712>

\*Cubitt T, Napier S & Brown R 2021. Predicting prolific live streaming of child sexual abuse. *Trends & issues in crime and criminal justice* no. 634. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti78320>

\*Dalins J, Tyshetskiy Y, Wilson C, Carman MJ & Boudry D 2018. Laying foundations for effective machine learning in law enforcement. Majura: A labelling schema for child exploitation materials. *Digital Investigation* 26: 40–54. <https://doi.org/10.1016/j.diin.2018.05.004>

Edwards G, Christensen L, Rayment-McHugh S & Jones C 2021. Cyber strategies used to combat child sexual abuse material. *Trends & issues in crime and criminal justice* no. 636. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti78313>

Ganann R, Ciliska D & Thomas H 2010. Expediting systematic reviews: Methods and implications of rapid reviews. *Implementation Science* 5: 56–66. <https://doi.org/10.1186/1748-5908-5-56>

\*Gangwar A, González-Castro V, Alegre E & Fidalgo E 2021. AttM-CNN: Attention and metric learning based CNN for pornography, age and child sexual abuse (CSA) detection in images. *Neurocomputing* 445: 81–104. <https://doi.org/10.1016/j.neucom.2021.02.056>

Garriss K & DeMarco N 2023. FBI warns of using AI deepfakes as part of sextortion schemes. Yahoo! News, 6 July. <https://www.yahoo.com/news/fbi-warns-using-ai-deepfakes-212047097.html>

\*Granizo SL, Valdivieso Caraguay ÁL, Barona López LI & Hernández-Álvarez M 2020. Detection of possible illicit messages using natural language processing and computer vision on Twitter and linked websites. *IEEE Access* 8: 44534–44546. <https://doi.org/10.1109/ACCESS.2020.2976530>

\*Grubl T & Lallie HS 2022. Applying artificial intelligence for age estimation in digital forensic investigations. <https://doi.org/10.48550/arXiv.2201.03045>

\*Guerra E & Westlake BG 2021. Detecting child sexual abuse images: Traits of child sexual exploitation hosting and displaying websites. *Child Abuse & Neglect* 122: 105336. <https://doi.org/10.1016/j.chiabu.2021.105336>

Henseler H & de Wolf R 2019. Sweetie 2.0 technology: Technical challenges of making the sweetie 2.0 Chatbot. In S van der Hof, I Georgieva, B Schermer & BJ Koops (eds), *Sweetie 2.0: Using artificial intelligence to fight webcam child sex tourism*. Information Technology and Law Series, vol 31. The Hague: TMC Asser Press: 113–134. [https://doi.org/10.1007/978-94-6265-288-0\\_3](https://doi.org/10.1007/978-94-6265-288-0_3)

High-Level Expert Group on Artificial Intelligence 2019. *A definition of AI: Main capabilities and disciplines*. Brussels: European Commission. <https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>

Internet Watch Foundation 2023. How AI is being abused to create child sexual abuse imagery. <https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/>

\*Isaza G, Muñoz F, Castillo L & Buitrago F 2022. Classifying cybergrooming for child online protection using hybrid machine learning model. *Neurocomputing* 484: 250–259. <https://doi.org/10.1016/j.neucom.2021.08.148>

\*Jin P, Kim N, Lee S & Jeong D 2024. Forensic investigation of the dark web on the Tor network: Pathway toward the surface web. *International Journal of Information Security* 23(1): 331–346. <https://doi.org/10.1007/s10207-023-00745-4>

\*Laranjeira C, Macedo J, Avila S & dos Santos JA 2022. Seeing without looking: Analysis pipeline for child sexual abuse datasets. arXiv. Presented at the 5th Conference on Fairness, Accountability and Transparency (FAccT), 2022. <https://doi.org/10.48550/arXiv.2204.14110>

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>

Lupariello F, Sussetto L, Di Trani S & Di Vella G 2023. Artificial intelligence and child abuse and neglect: A systematic review. *Children* 10: 1659. <https://doi.org/10.3390/children10101659>



\*Macedo J, Costa F & dos Santos J 2018. *A benchmark methodology for child pornography detection*. Paper presented at the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images: 455–462. <https://doi.org/10.1109/SIBGRAPI.2018.00065>

\*Meyer M 2015. *Machine learning to detect online grooming* (Master's thesis). <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-260390>

Milmo D 2023. AI-created child sexual abuse images 'threaten to overwhelm internet'. *The Guardian*, 25 October. <https://www.theguardian.com/technology/2023/oct/25/ai-created-child-sexual-abuse-images-threaten-overwhelm-internet>

\*Murcia Triviño J, Moreno Rodríguez S, Díaz López DO & Gómez Mármol F 2019. C3-Sex: A chatbot to chase cyber perverts. *2019 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress*: 50–57. <https://doi.org/10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00024>

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>

\*Ngejane CH, Eloff JHP, Sefara TJ & Marivate VN 2021. Digital forensics supported by machine learning for the detection of online sexual predatory chats. *Forensic Science International: Digital Investigation* 36: 301109. <https://doi.org/10.1016/j.fsidi.2021.301109>

\*Ngo VM, Gajula R, Thorpe C & Mckeever S 2024. Discovering child sexual abuse material creators' behaviors and preferences on the dark web. *Child Abuse & Neglect* 147: 106558. <https://doi.org/10.1016/j.chiabu.2023.106558>

Okolie C 2023. Artificial intelligence-altered videos (deepfakes), image-based sexual abuse, and data privacy concerns. *Journal of International Women's Studies* 25(2): 1–16. <https://vc.bridgew.edu/jiws/vol25/iss2/11>

\*Oronowicz-Jaśkowiak W et al. 2024. Using expert-reviewed CSAM to train CNNs and its anthropological analysis. *Journal of Forensic and Legal Medicine* 101: 102619. <https://doi.org/10.1016/j.jflm.2023.102619>

\*Peersman C, Schulze C, Rashid A, Brennan M & Fischer C 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 & Brown R 2021. Metadata-based detection of child sexual abuse material. <https://doi.org/10.48550/arXiv.2010.02387>

\*Polastro M de C & Eleuterio PM da S 2012. A statistical approach for identifying videos of child pornography at crime scenes. *2012 Seventh International Conference on Availability, Reliability and Security*: 604–612. <https://doi.org/10.1109/ARES.2012.71>

- \*Puentes J et al. 2023. *Guarding the guardians: Automated analysis of online child sexual abuse*. <https://doi.org/10.48550/arXiv.2308.03880>
- \*Razi A et al. 2023. Sliding into my DMs: Detecting uncomfortable or unsafe sexual risk experiences within Instagram direct messages grounded in the perspective of youth. *Proceedings of the ACM on Human-Computer Interaction* 7: article 89. <https://doi.org/10.1145/3579522>
- \*Rodríguez JI, Durán SR, Díaz-López D, Pastor-Galindo J & Mármol FG 2020. C3-Sex: A conversational agent to detect online sex offenders. *Electronics* 9(11): 1779. <https://doi.org/10.3390/electronics9111779>
- \*Rondeau J, Deslauriers D, Howard III T & Alvarez M 2022. A deep learning framework for finding illicit images/videos of children. *Machine Vision and Applications* 33(5): 66. <https://doi.org/10.1007/s00138-022-01318-6>
- \*Sae-Bae N, Sun X, Sencar HT & Memon ND 2014. Towards automatic detection of child pornography. *2014 IEEE International Conference on Image Processing (ICIP)*: 5332–5336. <https://doi.org/10.1109/ICIP.2014.7026079>
- \*Seedall M, MacFarlane K & Holmes V 2019. *SafeChat system with natural language processing and deep neural networks*. <https://sure.sunderland.ac.uk/id/eprint/10968/>
- \*Seigfried-Spellar KC et al. 2019. Chat analysis triage tool: Differentiating contact-driven vs. fantasy-driven child sex offenders. *Forensic Science International* 297: e8–e10. <https://doi.org/10.1016/j.forsciint.2019.02.028>
- Singh S & Nambiar V 2024. Role of artificial intelligence in the prevention of online child sexual abuse: A systematic review of literature. *Journal of Applied Security Research* 1–42. <https://doi.org/10.1080/19361610.2024.2331885>
- Thiel D, Stroebel M & Portnoff R 2023. *Generative ML and CSAM: Implications and mitigations*. Internet Observatory Cyber Policy Center, Stanford. <https://fsi.stanford.edu/publication/generative-ml-and-csam-implications-and-mitigations>
- Thorn 2024a. *Introducing Safer Predict: Using the power of AI to detect child sexual abuse and exploitation online*. <https://www.thorn.org/blog/introducing-safer-predict-using-the-power-of-ai-to-detect-child-sexual-abuse-and-exploitation-online/>
- Thorn 2024b. *Youth perspectives on online safety, 2023: An annual report of youth attitudes and experiences*. <https://www.thorn.org/research/library/2023-youth-perspectives-on-online-safety/>
- \*Ulges A & Stahl A 2011. Automatic detection of child pornography using color visual words. *2011 IEEE International Conference on Multimedia and Expo*: 1–6. <https://doi.org/10.1109/ICME.2011.6011977>
- \*Westlake B et al. 2022. Developing automated methods to detect and match face and voice biometrics in child sexual abuse videos. *Trends & issues in crime and criminal justice* no. 648. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti78566>

**Dr Heather Wolbers is a Senior Research Analyst in the Online Sexual Exploitation of Children Research Program at the Australian Institute of Criminology.**

**Dr Timothy Cubitt is a Principal Research Analyst at the Australian Institute of Criminology.**

**Michael John Cahill is a former Research Analyst in 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: [www.aic.gov.au](http://www.aic.gov.au)

ISSN 1836-2206 (Online) ISBN 978 1 922877 80 2 (Online)

<https://doi.org/10.52922/ti77802>

©Australian Institute of Criminology 2025

GPO Box 1936  
Canberra ACT 2601, Australia

Tel: 02 6268 7166

*Disclaimer: This research paper does not necessarily reflect the policy position of the Australian Government*

[www.aic.gov.au](http://www.aic.gov.au)