SOSC3202 Report

The social, policy and legislative considerations of AI-generated Child Sexual Abuse Material

Authors:

Ramida Suttipron *48311922*
Suparada Pongpisan *48307840*
Kassandra Buck *47075003*
Ana Lorita *32003208*
Aariyana Hussain *47314270*

Acknowledgements

We would like to acknowledge and express gratitude towards all the people who contributed to this research project. To our academic advisors and lecturers, Miss Elizabeth-Rose Ahearn, Professor Paul Henman, Dr Awais Hameed Khan and Associate Professor Jenny Povey, thank you for all your guidance and advice. To our industry partner Ashley Carvalho and the broader ICMEC Australia team, thank you for your trust and assistance throughout the entire research process. Lastly, thank you to all our participants for sharing your time, expertise and experiences, without which this project would have not been possible.

Who are ICMEC Australia

The International Centre for Missing and Exploited Children (ICMEC) Australia is a not-for-profit organisation working in Australia to support efficient detection, reporting, and prosecution of child sexual exploitation. They envision a world where children are safe from exploitation and abuse because child sexual exploitation crimes are prevented, offenders convicted, and victims relieved.

Executive Summary

Rapid advancements in Artificial Intelligence (AI) technology have led to a new and increasing concern that of AI-generated Child Sexual Abuse Material (CSAM) (Dennehy et al., 2020). As it is a rapidly evolving threat, this study aims to expand on existing research to explore public understanding, service provisioning and regulatory responses in Australia to improve safety for minors (Flynn et al., 2022).

Research questions:

1. What are the public perceptions of AI-generated CSAM in Australia?
2. How is AI-generated CSAM introducing new risks to society?
3. What are the inadequacies of the current Australian system regarding AI-generated CSAM regulation?

Our mixed-methods methodology involved an online survey that received 128 responses from the public, and four interviews with domain experts from NGOs and academia. Through thematic analysis, this study's findings have been separated into four broad themes regarding AI-generated CSAM: *Psychological and Social Risks to Children*, *Public Awareness*, *Challenges in Regulation,* and *Systemic Gaps*. Key insights from this study involved a high level of concern from the public and experts regarding AI-generated CSAM due to its increasing ease of generation and the psycho-social impacts on children, but there was limited knowledge regarding its support services, reporting pathways and detection. On a larger scale, the findings also revealed the need for various technological, legislative and regulatory reforms.

Key recommendations from the study:

- Awareness efforts must be targeted towards specific populations such as teachers and avoid unnecessary fearmongering from the public.
- Tech companies must be held accountable for implementing efforts such as" Safety by Design".
- Legislative and regulatory reforms must involve better-defined legal thresholds to clarify what constitutes AI-generated CSAM, and align them with international legislation.
- Reporting pathways and support services must be improved in terms of awareness, accessibility and quality.
- Future research must involve further insights from parents and caregivers.

Introduction

The rapid development of generative Artificial Intelligence (AI) models has led to new concerns and risks across online spaces, which require a holistic approach from social, policy and legislative perspectives. The particular concern this research seeks to explore is that of AI-generated Child Sexual Abuse Material (CSAM). Generative AI can be used in a variety of ways that can contribute to CSAM, including creating false images and videos of real or non-existent children who have or have not been physically abused (ICMEC Australia, 2023). Overall, Generative AI systems pose significant concerns as malicious actors can utilise this technology to induce physical, psychological, political and economic harm (Blauth et al., 2022).

Although AI-generated material may impact people of all ages, children are the focus of this study as they are particularly vulnerable, often having a significant online presence (Dennehy et al., 2020). Exposure to AI-generated CSAM has lasting impacts on children's social-emotional development, along with the overall need to protect them from physical and psychological harm (Adams et al., 2023). The rapid development of AI-generated CSAM requires an urgent response to address the gaps in current research and practice. It is of particular importance as victims of digital abuse often express that their experiences have been minimised (Flynn et al., 2022). As such, pervasive and malicious crimes have serious implications for children's well-being and safety. Current practices across social, policy and legislative spaces need to evolve to address these gaps.

Our research is conducted in collaboration with the International Centre for Missing & Exploited Children (ICMEC) Australia. ICMEC Australia has supported this research to gain a greater understanding of the risks introduced by AI-generated CSAM in Australia and assisted us in engaging with different stakeholders to account for various perspectives regarding this issue. ICMEC Australia is supporting this research as they have identified systemic knowledge gaps regarding AI-generated CSAM in Australia across legislative, social, and political spaces, which can be better addressed through this study's insights from the public and industry experts.

Our overall research topic involves exploring the social, policy and legislative considerations of AI-generated Child Sexual Abuse Material (CSAM) in Australia. The research aims to understand the Australian public's awareness of AI-generated CSAM, its risks and vulnerabilities for children, and the inadequacies of the current Australian system's regulation of the issue. These aims were addressed through a public online survey and interviews with industry experts. The research questions are as follows:

1. What are the public perceptions of AI-generated CSAM in Australia?
2. How is AI-generated CSAM introducing new risks to society?
3. What are the inadequacies of the current Australian system regarding AI-generated CSAM regulation?

Literature Review

The literature review involves the key definitions used in this study followed by the psychological and social impacts of online sexual abuse on victim-survivors. The review then elaborates on the rapid development of generative AI models and the challenges introduced as a result, across social, policy, and legislative domains. Finally, the review explores the role of the technology industry.

Definitions

Child Sexual Abuse Material: "images, pictures, films, videos, or computer-generated visual material that depict a child in a sexually explicit manner" (Oxford Research Encyclopedia of Criminology, 2024). Under Australian Commonwealth legislation, it is an offence to possess, produce, supply, or obtain such material (ICMEC Australia, 2024).

Deepfake: "A video of a person in which their face or body has been digitally altered so that they appear to be someone else, typically used maliciously or to spread false information." (Oxford English Dictionary, 2023)

Image Based Sexual Abuse (IBSA): "When someone shares, or threatens to share, an intimate image or video of a person without their consent" (ESafety Commissioner, 2024).

Generative AI: "Deep-learning models that can generate high-quality text, images, and other content based on the data they were trained on" (Martineau, 2023).

AI-generated CSAM (AI-CSAM): Is the use of generative AI to produce child sexual abuse material. Resulting in a product that appears to involve children in a sexual manner, this can include images, videos and/or audio (ICMEC Australia, 2024.)

Sextortion/ Sexual extortion: This cybercrime refers to online blackmail from a perpetrator threatening to reveal explicit images of a person unless they give in to their demands, which are sometimes for financial gain (ICMEC Australia, 2024).

Victim-blaming: Questioning people who have experienced violent behaviour (such as a survivor of sexual violence) as if they were somehow responsible for it, instead of placing the responsibility where it belongs: on the person who harmed them (SACE, 2022).

Psychological and Social Impacts on Victim-Survivors

Limited public awareness, and the inefficacy of victim support services can lead to severe psychological and social impacts (Flynn et al., 2022; Martin, 2021; McGlynn et al., 2021). The existing research on the experiences of victim-survivors of online sexual abuse highlights that insufficient public awareness regarding the severity of the harms experienced by victims often results in a hesitancy to report incidents of online sexual abuse (Flynn et al., 2022; Martin, 2021; McGlynn et al., 2021). Victim-survivors frequently express concerns about not being believed, facing victim-blaming, and feeling dismissed by authorities (Flynn et al., 2022; Martin, 2021; McGlynn et al., 2021). These issues undermine policing and prevention efforts (Flynn et al., 2022; Martin, 2021; McGlynn et al., 2021).

Moreover, McGlynn and colleagues (2021) note that the digital nature of online sexual abuse, which allows content to be continuously shared, viewed, and rediscovered, perpetuates harm for victim-survivors and reinforces their traumatic experiences. Many victim-survivors express ongoing fear, anxiety, feelings of unsafety, and the constant concern that the abuse may reemerge at any time (McGlynn et al., 2021). Lack of support from families, community networks, and society at large further exacerbates their isolation and fear, causing them to recoil from society (McGlynn et al., 2021; Okolie, 2023). Therefore, there is a significant imperative for the establishment of practical and robust support systems to facilitate the recovery and reintegration of victim-survivors of online sexual abuse into society (Okolie, 2023).

Given the limited research on public perceptions regarding AI-generated CSAM, findings from an online survey on this topic will be critical for informing public awareness levels and misconceptions. These insights will aid in producing effective policies, practices, and support systems to address these challenges.

Risks and Implications of Generative AI for CSAM

Generative AI can accelerate the production of CSAM thus impacting more victims. Studies specifically related to AI-generated CSAM have indicated that, within a year, AI algorithms are anticipated to advance to a level of sophistication where they can generate highly realistic CSAM that is indistinguishable from real imagery depicting actual instances of child abuse (Janjeva et al., 2023; Thiel et al., 2023). This issue is compounded by the availability of open-source generative AI models, which can be trained extensively on adult content. These models are often equipped with or without content filters that can be bypassed and rendered freely accessible to users with limited technical proficiency (Janjeva et al., 2023; Thiel et al., 2023). Furthermore, Janjeva and colleagues (2023) underscore that the proliferation of AI technologies has augmented pre-existing vulnerabilities among children by accelerating the speed and scale of non-consensual creation, display, and distribution of CSAM, causing harm to a larger proportion of the population than before.

Considering this, the proliferation of AI-generated CSAM should be considered an urgent and pernicious problem. It poses new vulnerabilities to children and presents a substantial challenge to various stakeholders, including all levels of government, law enforcement agencies, non-governmental aid and support organisations, academia, and civil society (Blauth et al., 2022; Janjeva et al., 2023). Thus, insights from industry experts on the new vulnerabilities and risks posed by AI-generated CSAM to children, as well as the complex challenges faced by various stakeholders, will be crucial for informing the creation of robust strategies to combat online child abuse and enhance societal awareness (Dennehy et al., 2020; McGlynn et al., 2021).

Challenges in Social, Policy, and Legal Domains

Given the rapidly evolving and complex nature of AI-generative technologies, there is a critical concern that laws and platform regulations may struggle to keep pace with deepfake and digitally altered imagery abuse (Flynn et al., 2022; Okolie, 2023). The literature on deepfakes and IBSA highlights how the swift progression of generative AI models facilitates CSAM proliferation and poses significant challenges in social and legal domains (Flynn et al., 2021; Flynn et al., 2022; Okolie, 2023). Flynn and colleagues (2022) contend that existing Australian laws criminalising IBSA fail to encompass deepfakes, thereby diminishing their applicability, and severity, disregarding the harm inflicted on victim-survivors. The authors further stress that deepfake and digitally altered imagery abuse should not be dismissed as a distant concern or merely seen as a component of IBSA (Flynn et al., 2022). Instead, they argue that deepfake and digitally altered imagery abuse should be treated as a distinct category of abuse that requires dedicated legal and support frameworks (Flynn et al., 2022).

The Role of the Technology Industry

Okolie (2023) argues that the rapid pace of technological advancements and delays in legal reforms render sole reliance on legislation inadequate to address the issue of malicious deepfake use. Collaborative efforts among various stakeholders, including the tech industry, government, and support service providers, are crucial for strengthening safety measures and support (Okolie, 2023). For instance, Okolie (2023) emphasises the importance of adopting digital education as a societal norm in the digital era.

As illuminated in the existing scholarly discourse, the rapid advancement of AI has established new pathways for exploitation, leading to unprecedented production and dissemination of sexual abuse material. The ensuing challenges posed include limited societal awareness, inadequate legislative frameworks, and the need for new policy and practice formulation. The research also emphasises the necessity to address these escalating issues in online sexual abuse regulation and prevention efforts. However, scholarly exploration focused on AI-generated CSAM within the Australian context remains scarce. Therefore, our research will build on existing knowledge from pertinent studies and contribute to fill the current gaps in the literature on online sexual abuse. The study will encompass public perceptions regarding AI-generated CSAM, the new societal risks associated with it, and the inadequacies of the current Australian system for regulating AI-generated CSAM.

Methodology

The Pragmatic Approach

This research adopts a mixed-method approach to assess public awareness of AI-generated CSAM and examine professional perspectives on the risk of AI-generated CSAM. Through integrating quantitative and qualitative findings, we attained a nuanced understanding of AI-generated CSAM and were able to derive actionable insights (Dawadi et al., 2021).

We adopt the pragmatic paradigm, facilitating a comprehensive examination of the subject from varied viewpoints, utilising diverse methodologies, and combining and comparing quantitative and qualitative data results to address research questions. Pragmatism is defined as "shared beliefs among members of a specialty area" (Brierley, 2017, p.140). We have chosen to use the pragmatic approach as our research paradigm due to its flexibility and suitability for mixed-methods research projects (Brierley, 2017). Furthermore, the flexibility of the pragmatic approach prevents the researchers from being restricted by ontological and epistemological issues when addressing different research questions (Brierley, 234017), thus allowing the researchers to construct a diverse in-depth perspective of AI-generated CSAM from multiple viewpoints.

Online public survey

To explore the public's awareness of AI-generated CSAM, and thus answer the first research question, an online survey featuring open-ended and close-ended questions was conducted. With the focus on objectivity and obtaining quantifiable data, a positivist epistemology was used to examine respondents' perceptions of the outlined issue. This enables social research surveys to acquire empirical ontology by providing descriptive information (Hasan, 2014). Consequently, this epistemological framework enables the research project to use the collected data as empirical evidence on public awareness levels and merge it with qualitative interview data in the analysis (Ghiara, 2019). A total of 128 survey respondents were recruited. These participants were collected through convenience sampling, as they were recruited based on accessibility and convenience. Participants accessed the research survey via links or QR codes shared on social media platforms including Facebook, LinkedIn, and the ICMEC Australia website, as well as through physical flyers distributed across the University of Queensland campus. This sampling methodology enabled the researchers to collect a diverse range of respondents, thus better representing the majority public perception.

The online survey took approximately 15 minutes and was accessed via the Qualtrics platform. It comprised multiple choice and short-answer open-ended questions. It was designed to be as short as possible while collecting the required information to improve the completion rate. If survey participants withdrew before completing the questionnaire, their incomplete responses were included in the dataset.

Sample characteristics

The most prevalent age bracket of participants was 18-24 (Figure 2), with a strong gender skew towards women, with zero respondents selecting ‘non-binary’ or ‘other’ (Figure 1). While the convenience sampling of the survey most likely contributed to the skewing of the 18-24 age bracket, this aligns with existing studies in Australia which show that participants in this age range tend to be more aware of AI and may be more likely to respond to surveys regarding this topic (Selwyn & Cordoba, 2022). The younger demographic skew of this study’s participants probably explains the low numbers of parent, grandparent, and caregiver respondents.

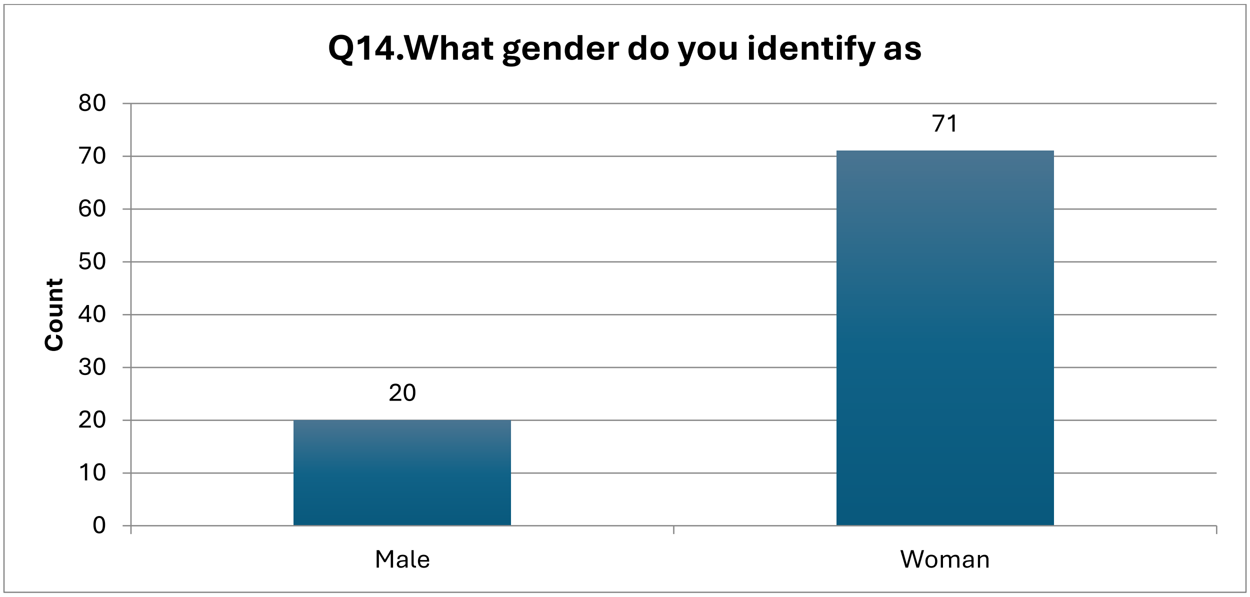


Figure 1. Reported Gender of Survey Participants

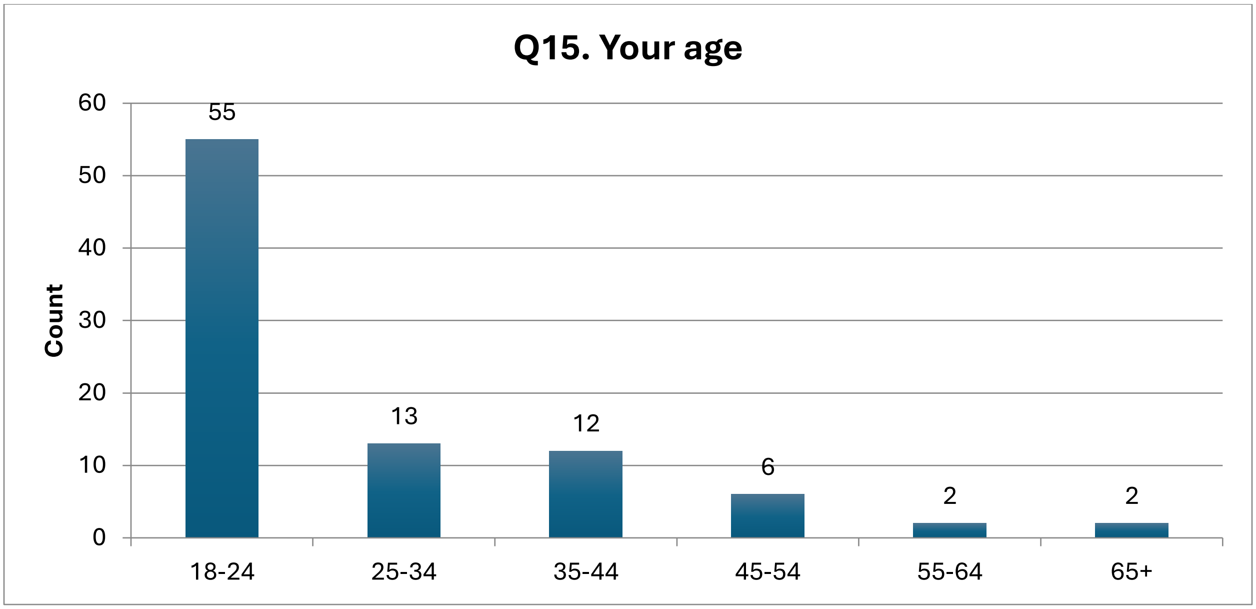


Figure 2. Age of Participants

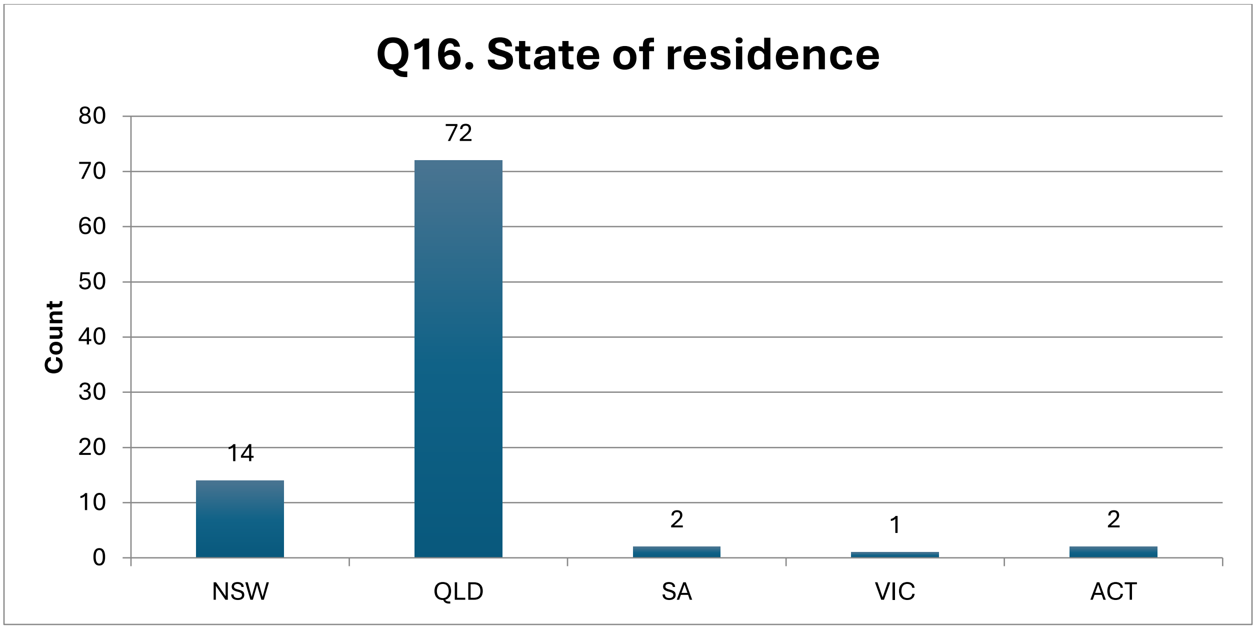


Figure 3. Participants' State of Residence

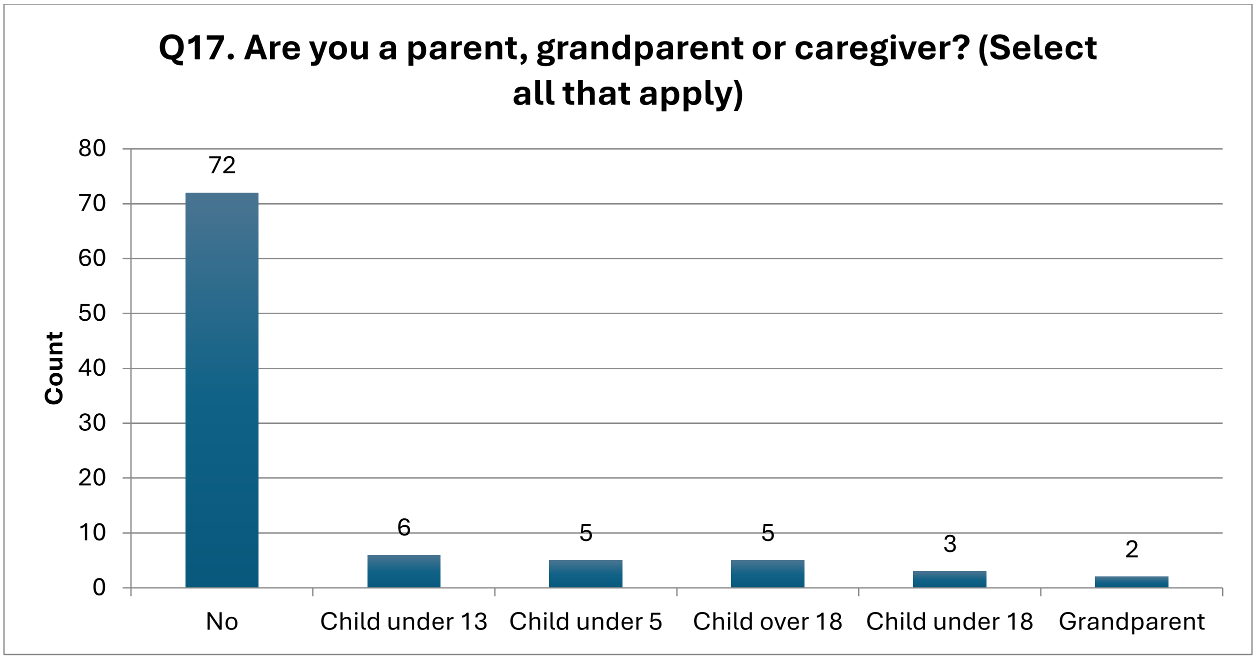


Figure 4. Participants' Parent/Caregiver Status

Semi-structured online interviews

To examine the risks of AI-generated CSAM and the challenges of the current system in dealing with cases, thus answering the second and third research questions, a qualitative method was applied through semi-structured interviews with experts working in this field. To understand the subjective and contextual interpretation of this issue, interpretivist epistemology was used instead of positivism. Interpretivism assisted in understanding the complexity and variability of socio-behavioural phenomena surrounding AI-generated CSAM from the professionals who work involved in this issue (Ghiara, 2019). The interview theme includes questions on the societal impact of AI-generated CSAM, the associated risks and challenges, the effectiveness of current policy and legal responses, and potential technological solutions to mitigate these issues (see Appendix B). The interview data underwent analytical scrutiny to compare the current system's challenges with mainstream perspectives, illuminating critical risks and issues requiring attention.

With our aim to gather the perspectives of professionals regarding the impact of AI-generated CSAM children, their families and the wider community, our research conducted semi-structured interviews with four selected interviewees. Participants were selected through purposive sampling by ICMEC Australia recruited from their working group. These included industry experts from academia, NGOs, and law enforcement who can provide valuable insights into the new risks introduced by AI-generated CSAM. Purposive sampling was employed to ensure that the interviewees had the expertise to answer our research question, specifically examining the ongoing risks of AI-generated CSAM from the perspectives of professionals who work closely with this issue. They were informed of the project's aims and consent was obtained before the interviews. Upon agreeing to the outlined conditions, participants were asked to sign the consent form and return it before the interview. The online interviews were conducted through Zoom and Microsoft Teams, with participants committing to a one-time session lasting approximately 30 minutes. During the interview, data was collected through digital audio recordings, to be transcribed for analysis. If a participant chose to withdraw before, during or after the interview session, their data would have been excluded from the analysis, and any recordings or transcripts were deleted to maintain confidentiality.

Data analysis

Merging findings from both datasets a triangulation strategy was used to provide a more comprehensive understanding of the research problem by cross-analysing evidence from qualitative and quantitative methods (Dawadi et al., 2021). This involved overlapping the qualitative responses regarding the risk and vulnerability of AI-generated CSAM with public perceptions toward this issue from the survey.

Quantitative data was analysed using Qualtrics and Excel. Enabling the research team to analyse, manage, and produce graphical visualisations of data. Two main analyses were conducted. Descriptive statistics was used to summarise participant demographics, including age, gender, and location. Crosstab descriptive statistics yielded a view of subgroups' perceptions regarding AI-generated CSAM.

To analyse the qualitative data, thematic analysis was conducted using Braun & Clarke's (2006) approach. Responses were transcribed using transcription software and then reviewed for accuracy following Flick's (2018) transcription conventions approach. After transcribing the data, it was coded into categories, followed by an examination to identify key themes. Inductive coding was used to identify themes.

Integrating quantitative and qualitative survey data, a triangulation strategy was used to converge on a more comprehensive understanding of the topic (Dawadi et al., 2021). Within this strategy, qualitative data significantly contributed to explaining some of the quantitative findings regarding Australian perceptions toward AI-generated CSAM (Dawadi et al., 2021). This involved overlapping the qualitative responses regarding the risk and vulnerability of AI-generated CSAM with the numerical results.

Ethical implications of the survey

We chose an online survey as it was the most cost-effective tool to gather a wide range of opinions from the public in a short time frame while preserving anonymity (Beam, 2023). The subjects were required to be Australian residents over 18; we excluded people under 18 due to concerns over psychological distress and general concerns related to conducting research with minors as well as an inability to give consent.

We were seeking to understand what the public knew and what they would do if they encountered CSAM content. We countered the potential for psychological distress by adequately disclosing the nature of the survey's content, not asking questions directly related to experience of sexual abuse or CSAM and providing links to support services. We assured anonymity by using non-identifiable data collection methods.

Ethical implications of the interviews

We used our partnership with ICMEC Australia to access members of their working group. The ethical concerns of working with experts are social impacts such as their comments impacting their work life or employer. This was mitigated through anonymisation. The psychological risks were mitigated as those selected were industry experts in this field and thus this research did not expose them to content beyond their daily work.

Limitations of the study

The survey was accessed online and through physical flyers distributed across the University of Queensland Saint Lucia campus. As stated above/below, the sample involves a disproportionate skewing towards young people and a strong skewing towards women, with them accounting for 77% of the total survey responses. With the survey being online and distributed on campus this likely contributed to the young person bias, as the age distribution on UQ's campus mostly skews young. Some older demographics may be unfamiliar with online surveys and, as a result, less likely to participate. Therefore, additional research should be conducted to ensure a more representative sample. Due to the limited sample size especially concerning parents and caregivers, we were unable to produce statistically significant results in some areas limiting our ability to perform demographic comparisons.

Results and Discussion

The online survey received 128 responses with participants sharing their insights, concerns and recommendations regarding AI-generated CSAM. The four expert interviews involved three NGO representatives and one academic representative, who have been identified as "Experts 1, 2, 3 and 4" to protect their identities. These experts shared their insights on how AI has exacerbated risks for children, the adequacy of national and international legislation and regulation, and potential solutions for addressing AI-generated CSAM on individual and societal levels.

By utilising thematic analysis, the insights from the public and industry experts have been separated into four categories: Psychological and Social Risks to Children, Public Awareness, Challenges in Regulation, and Systemic Gaps.

Psychological & Social Risks

When asked about their concern regarding AI-generated CSAM, both the survey participants and industry experts highlighted various psychological and social risks posed by this material for children. Open-ended survey answers and interview insights further underscore the significant mental health impacts on child victims of AI-generated CSAM and the reoccurring trauma due to the circulation of abuse material across social media. As one expert noted:

> "*I think that with generated AI images, there's gonna be very similar impacts for those victims, knowing that photos of them that are sexually explicit are being shared and are out there on the Internet, even if they are just generated through AI. It will be really concerning for many of them and may actually have some real impacts in terms of their mental health and well-being.*" – Expert 2

Such insights align with existing studies, whereby authors have noted that the accessibility of online CSAM perpetuates harm for victim-survivors and reinforces their traumatic experiences through feelings of continuous fear, anxiety, and unsafety regarding their abuse material resurfacing at any time (McGlynn et al., 2021). Moreover, one expert raised concerns about the disruption of neurological and behavioural development of child victims of AI-generated CSAM. The long-lasting psychological impacts on these victims are further emphasised in numerous studies examining the experiences of survivors of online sexual abuse, where victims reported facing various mental health challenges, including post-traumatic stress disorder (PTSD), low self-esteem, self-harming behaviours, and persistent low moods (Schmidt et al., 2023; Yasmine, 2024).

The concern regarding impacts on social and interpersonal relationships was also highlighted in survey participants' open-ended answers when they were asked to list any concerns regarding AI-generated CSAM, whereby one participant stated, "Negative social impacts for victims through damaging relationships with their friends and potential future employability difficulties." Studies have elaborated on such social risks faced by victims of online child sexual abuse, indicating that the disclosure or discovery of CSAM may result in victim-blaming and stigmatisation from peers, family members, and the broader community (Schmidt et al., 2023; Yasmine, 2024). Consequently, victims may experience social exclusion and marginalisation (Yasmine, 2024). Additionally, some victims isolate themselves from friends and family due to feelings of embarrassment (Schmidt et al., 2023). This loss of contact can further diminish their motivation to engage in schoolwork, leading to significant educational disruptions (Schmidt et al., 2023; Yasmine, 2024).

Beyond the impacts on children, CSAM also impacts adults. Two-thirds (67%) of survey participants believed they would be quite to extremely affected emotionally by AI-generated CSAM if they encountered it online, with 30% of the responses responding 'very' and 37% 'extremely affected' (see Figure 5).

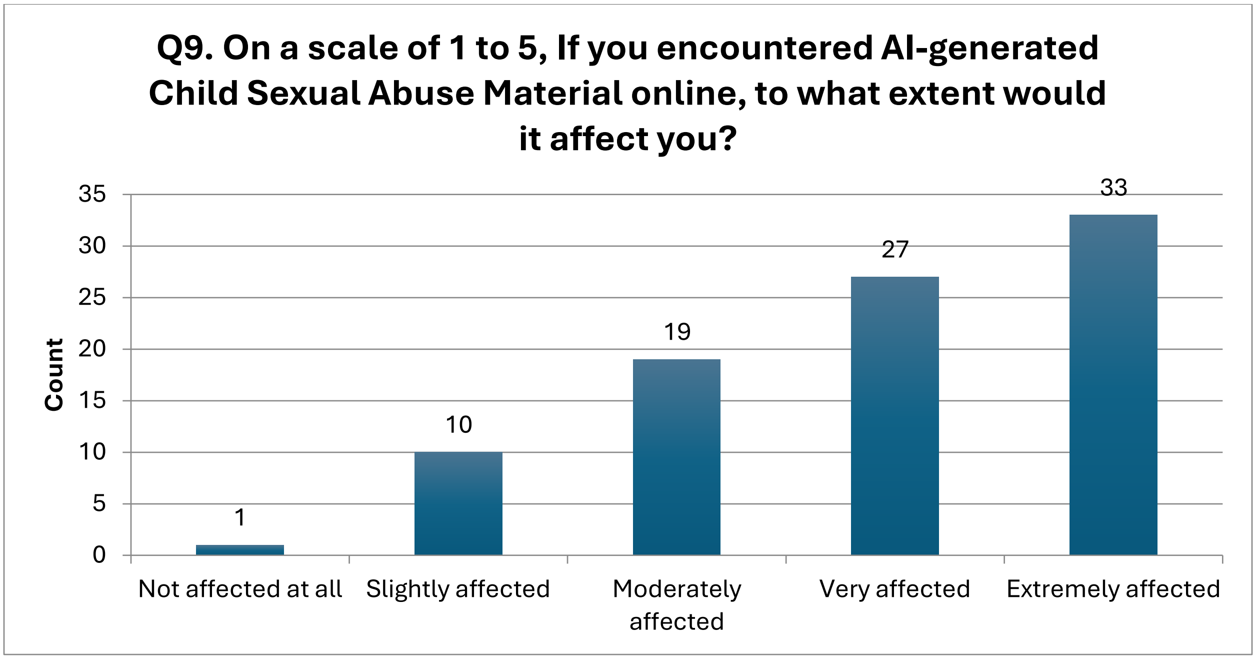


Figure 5. Participant reported impact of AI-generated CSAM

Supporting this, one expert highlighted that the impacts of AI-generated CSAM extend beyond the victim-survivors, to affecting their families, peers, and communities. Such insights are echoed throughout numerous studies regarding the extent of victimisation, whereby children are not the only ones directly affected by online abuse material but also their caregivers and surrounding families (Martin, 2014; Fong et al., 2020). The discovery of CSAM leads to significant emotional and psychological distress for caregivers and families, stemming from concerns about their child’s physical health, mental well-being, and future functioning (Fong et al., 2020). It also reinforces negative beliefs and self-blame regarding their perceived inability to protect their children (Fong et al., 2020).

Public Awareness of AI-generated CSAM

The findings from the online survey and industry experts regarding the public's awareness towards AI-generated CSAM relates to three main areas: Familiarity with the Issue, Where to Go for Help, and Improving Awareness Efforts.

Familiarity with the Issue

Overall, the public has limited familiarity with the topic. According to the survey, 84% of respondents answered that they have "little" to "no" familiarity with AI-generated CSAM (see Figure 6). Parents or caregivers reported having an even lower familiarity, of the 21 parents or caregivers surveyed 19 (90.5%), (see Figure 8.) reported that they have "little" to "no" familiarity with AI-generated CSAM. But we can also see that the participants had a reasonable idea of the definition of AI-generated CSAM with 62% saying the given definition matches their preconceived idea (see Figure 7).

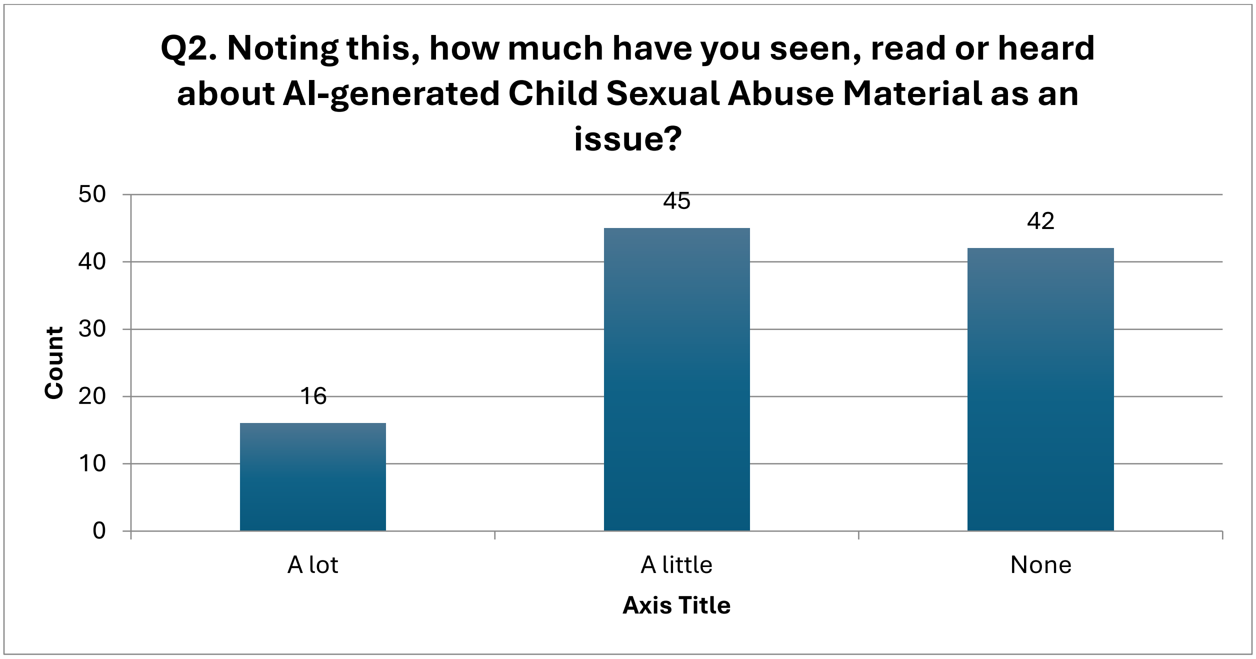


Figure 6. Participant reported awareness of AI-generated CSAM

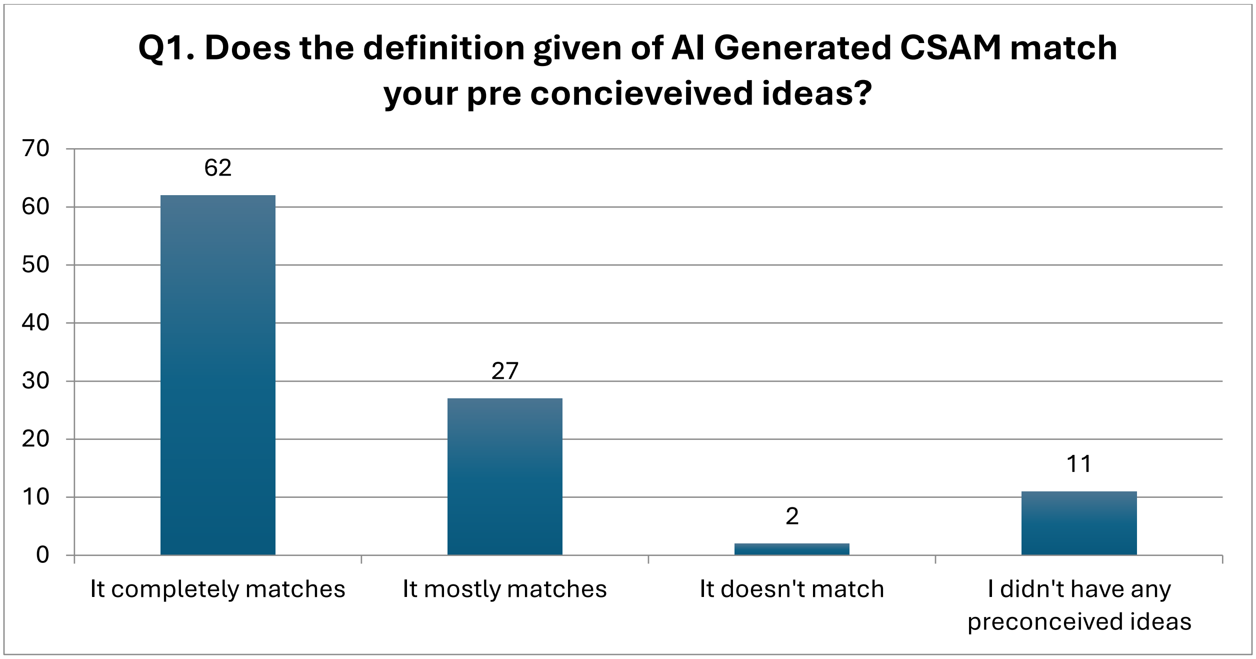
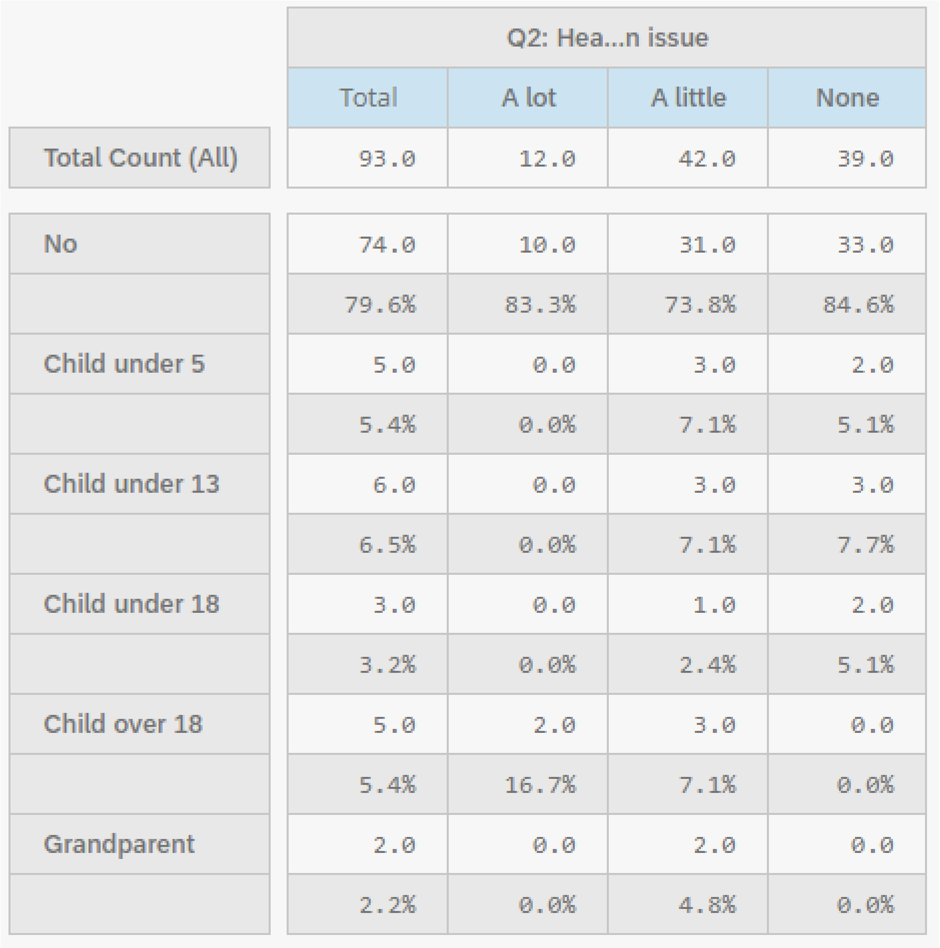


Figure 7. The definition given: AI-generated Child Sexual Abuse Material (CSAM) can be defined as the use of Artificial Intelligence to create sexualised images or videos of children that are either: Digitally altered to depict a real child (deepfakes), or Completely generated depicting a child that doesn't exist.



	Q2: Hea...n issue			
	Total	A lot	A little	None
Total Count (All)	93.0	12.0	42.0	39.0
No	74.0	10.0	31.0	33.0
	79.6%	83.3%	73.8%	84.6%
Child under 5	5.0	0.0	3.0	2.0
	5.4%	0.0%	7.1%	5.1%
Child under 13	6.0	0.0	3.0	3.0
	6.5%	0.0%	7.1%	7.7%
Child under 18	3.0	0.0	1.0	2.0
	3.2%	0.0%	2.4%	5.1%
Child over 18	5.0	2.0	3.0	0.0
	5.4%	16.7%	7.1%	0.0%
Grandparent	2.0	0.0	2.0	0.0
	2.2%	0.0%	4.8%	0.0%

Figure 8. Crosstabulation of question 2 and parental status

Similarly, most of the experts stated that they had not experienced a case of AI-generated CSAM directly. However, they were familiar with the issue as they have been in contact with law enforcement and agencies that work closely on this issue. From their past conversations about this topic, one expert noted that the issue has not been seriously introduced to the public.

> *"We saw the articles and news around schools in Australia, where peers were generating images of their peers and distributing them, yet there's been no real discussion in the general space. More broadly, there's no one talking about this."* – Expert 3

> *"While we've seen it and its impact here in Australia, we haven't actually seen some positive conversations that it should have."* – Expert 3

They further noted that

> *"There's still a lot of confusion by the general public around what the risks of AI are."* – Expert 3

This underscored the necessity of enhancing public education and awareness regarding AI-generated sexual abuse material, as highlighted by McGlynn and colleagues (2021) and Martin (2021). The data shows that the public is largely unfamiliar with AI-generated CSAM, with few having seen, read, or heard about it, indicating that the media and institutions have not effectively addressed this issue (Martin, 2021). This lack of attention has led to limited public understanding regarding the impact of AI-generated CSAM and the knowledge of how to respond to it. This is corroborated by our findings suggesting that the public has limited knowledge of AI-generated CSAM and its related support services. Therefore, improving public education and initiatives on this issue would lead to a deeper understanding of victim-survivors' harms and experiences, helping foster a more supportive environment and encourage appropriate responses from both the public and support services (McGlynn et al., 2021).

Where to Go for Help

Findings on public awareness regarding this issue revealed a significant gap in relevant support should people encounter CSAM or be victims of it. About 4 in 5 respondents (81.8%) were unfamiliar with support services related to AI-generated CSAM. Parents and caregivers demonstrated slightly higher awareness, with 73.7% reporting unfamiliarity however this is within our error margins (see Figures 9 and 10).

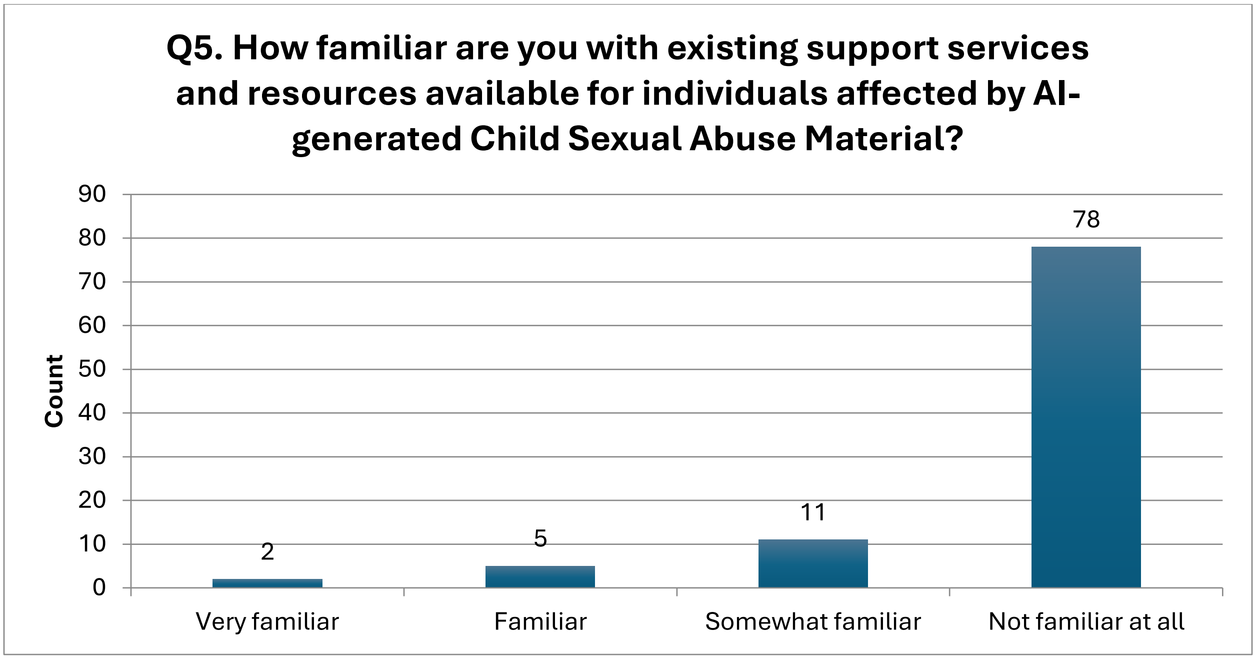
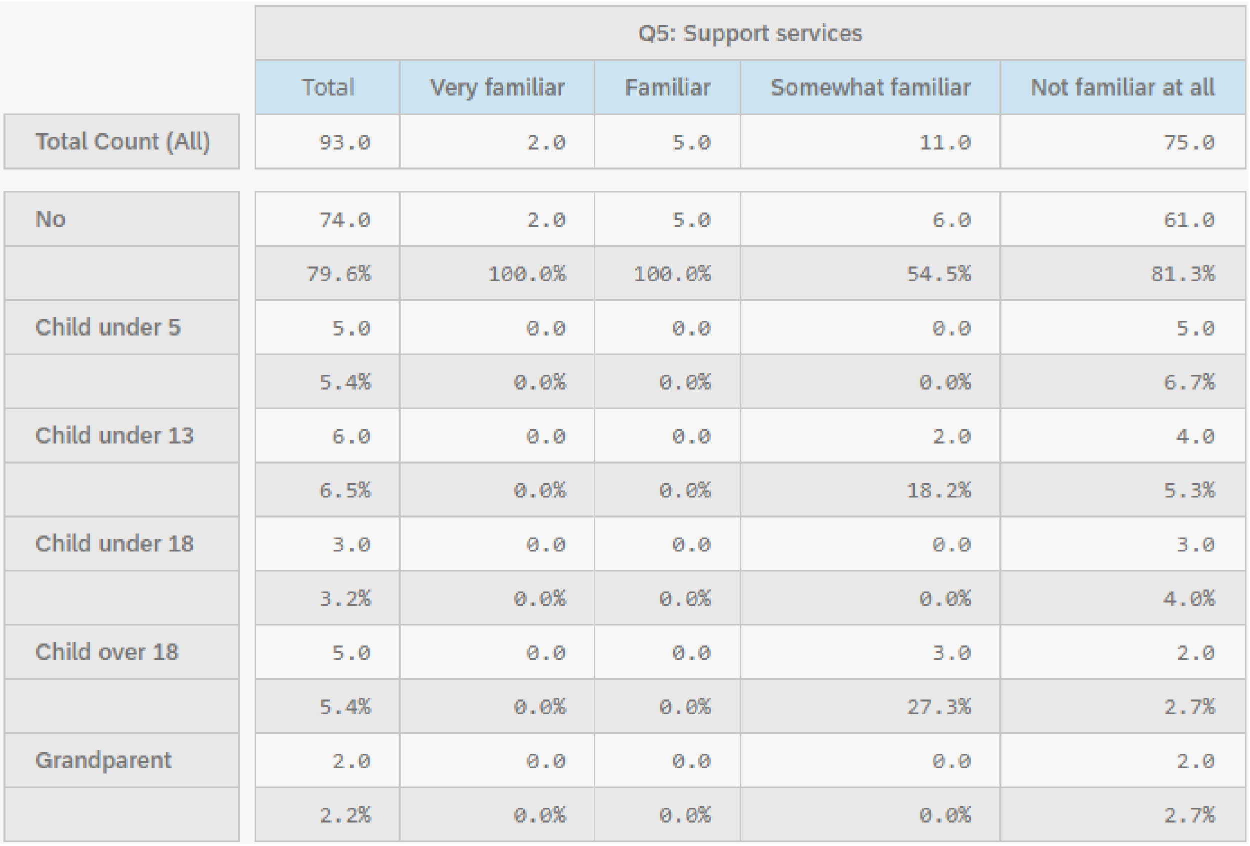


Figure 9. Participant Reported Familiarity with AI-generated CSAM



	Q5: Support services				
	Total	Very familiar	Familiar	Somewhat familiar	Not familiar at all
Total Count (All)	93.0	2.0	5.0	11.0	75.0
No	74.0	2.0	5.0	6.0	61.0
	79.6%	100.0%	100.0%	54.5%	81.3%
Child under 5	5.0	0.0	0.0	0.0	5.0
	5.4%	0.0%	0.0%	0.0%	6.7%
Child under 13	6.0	0.0	0.0	2.0	4.0
	6.5%	0.0%	0.0%	18.2%	5.3%
Child under 18	3.0	0.0	0.0	0.0	3.0
	3.2%	0.0%	0.0%	0.0%	4.0%
Child over 18	5.0	0.0	0.0	3.0	2.0
	5.4%	0.0%	0.0%	27.3%	2.7%
Grandparent	2.0	0.0	0.0	0.0	2.0
	2.2%	0.0%	0.0%	0.0%	2.7%

Figure 10. Crosstabulation of question 5 and Parental status

Experts also reinforced these findings, pointing out that gaps in legal definitions and the absence of a standardised framework for AI-generated CSAM in Australia contribute to insufficient public acknowledgment of its impact and a low level of awareness regarding the issue. One expert highlighted:

> *"From a community level, there probably isn't the level of awareness that we think is needed."* – Expert 1

These findings suggest significant inadequacies in public understanding and knowledge about available support services for victim-survivors of AI-generated CSAM. These services have not been adequately introduced or promoted, leaving individuals uncertain about whom to contact and where to seek help when such incidents occur.

Improving Awareness Efforts

Despite the public's lack of familiarity with the issue, many survey participants wanted to know more about AI-generated CSAM, thus suggesting improvements in the implementation of awareness efforts. In open-ended comments, 13 survey respondents emphasised the need for greater public awareness efforts regarding the (prevalence of) malicious use of AI to generate CSAM and its associated risks and harms. One of these respondents underscored the importance of implementing better education on the risks of Internet usage for children, parents, and schools. Furthermore, two respondents highlighted the importance of awareness efforts and education on appropriate actions to take if individuals encounter such material or become victims themselves, including guidance on available reporting pathways.

While the experts also reinforced the importance of awareness efforts, they focused more on a targeted approach aimed at raising awareness among specific groups, such as school staff, technology companies, and therapists. For example, one expert indicated that while enhancing public awareness of AI-generated CSAM can generate interest and encourage policy change, it may also lead to unnecessary concern among large groups of people who lack the leverage to control this issue. They stated:

> *"I think increasing the degree of concern about child sexual abuse is really important and awareness and all of that within the community. I can also say how increasing*

> *unnecessary concern among a large group of people [who] were kind of not within their control to do anything about it is also problematic."* – Expert 2

The representative further remarked:

> *"So, I think that this is probably more productive, to make targeted populations aware of it for prevention rather than kind of a scattergun approach, or everyone needs to know what is going on."* – Expert 2

Supporting this notion, Flynn and colleagues (2022) similarly proposed a targeted approach to awareness efforts, suggesting that education, training, and prevention campaigns on the harms of AI-generated CSAM and deepfakes should be implemented for police, criminal justice agencies, and public health and victim support workers. These efforts should specifically address the nature and impact of AI-generated CSAM, existing response options, and strategies for supporting victims (Flynn et al., 2022).

Referring to current awareness initiatives, experts further stated:

> *"There's been amazing strides by, you know, the government and what the eSafety Commissioner is doing is, is incredible in that space. And so, there's, there's always room for growth, always room for improvement."* – Expert 4

Such insights shed light on how organisations such as the eSafety Commissioner have adopted different strategies to promote online safety, but there is also room for improvements as AI continues to evolve and introduce new risks. For instance, established by the eSafety Commission in 2020, the *School Community Engagement Plan* provides strategies to help Australian schools engage their communities in addressing and preventing online safety issues (Sarma, 2022). A key feature of this plan is the *eSafety Toolkit for Schools,* which aims to raise awareness among students, parents, and educators about online risks, including online CSAM, and offers mechanisms for reporting online crimes against children (Sarma, 2022). However, there is still a need to update existing resources to include explicit comprehensive information on AI-generated CSAM, given the recent development of generative AI.

Challenges in Regulation

The findings from the survey and interviews also highlighted various challenges in regulating AI-generated CSAM. These findings have been separated into three main areas: Ease of Creation and Access, Law Enforcement Challenges, and Reporting Pathway Challenges.

Ease of Creation and Access

Another theme regarding the risks of AI-generated CSAM that arose from the participants was its ease of creation and accessibility. The interviewed experts explained how the ease of creating AI-generated CSAM increases its use as a tool for school bullying, which highlights concerns regarding the misuse of AI by children in educational settings (Chen & Lin, 2024). Experts also highlighted concerns regarding AI-generated CSAM being easily created through photos of children available online, with the borderless nature of the internet allowing perpetrators to exploit regions with weaker law enforcement. One expert highlighted:

> *"Online anonymity leads to perpetrators feeling invisible to the law."* – Expert 2

Such expert insights regarding borderless crimes highlight a significant concern regarding the ease of creation and access of AI-generated CSAM. These insights link to studies regarding the ubiquitous nature of the internet and highlight concerns regarding the difficulties of regulating and addressing digital crimes (Maier, 2010). Survey participants echoed experts' concern that the ease of generation is a key aspect that allows for widespread proliferation if left unchecked. Thus, the pervasive nature of online abuse presents various risks in regulating AI-generated CSAM.

The ease of creating AI-generated CSAM leads to concerns related to growth in its creation. As such 36% of survey participants believed that AI-generated CSAM has been occurring "often" in Australia (see Figure 11).

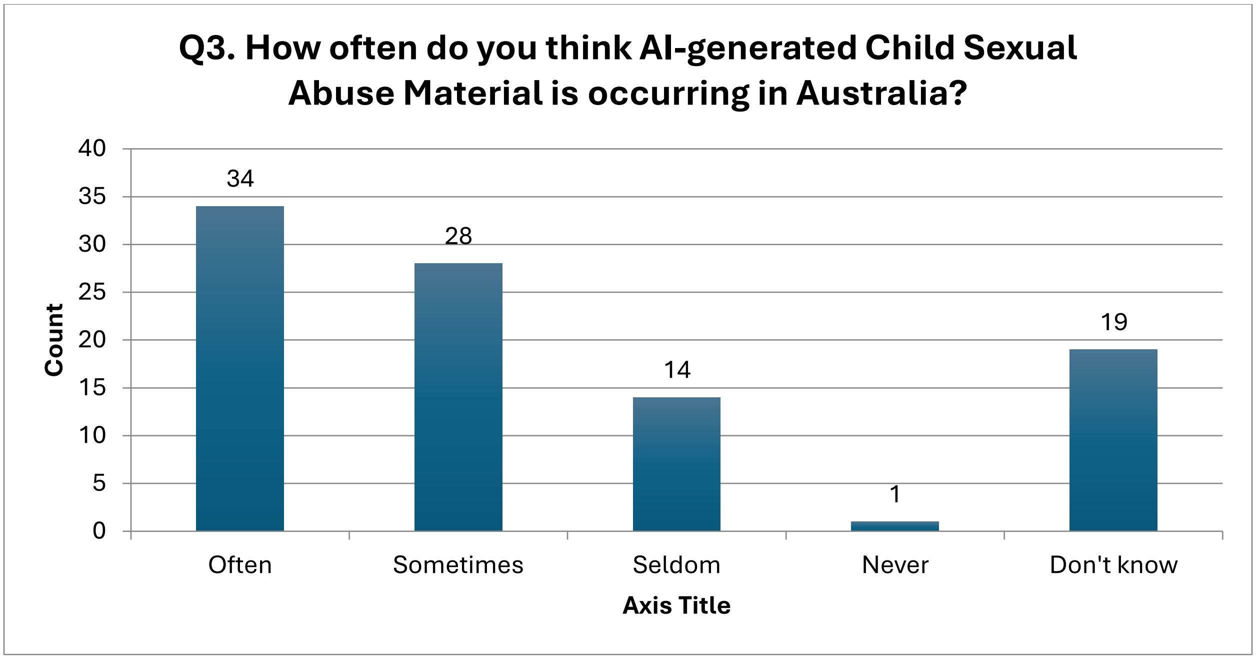


Figure 11. Participant reported occurrence of AI-generated CSAM

It is difficult to determine the actual amount of AI-generated CSAM that is present, however in 2024, the Australian Federal Police arrested two people: one for possession (AFP, 2024a) and one for production of AI-generated CSAM material (AFP, 2024b).

Additionally, when asked about the likelihood of a minor's photo being used by AI to generate abusive material, 54% of respondents answered that it was "very likely" or "likely" to happen (see Figure 12). The only group that answered the likelihood of minor’s images being used to generate AI-generated CSAM as “unlikely” were respondents with no children, accounting for 19.4% of the childless demographic group (figure 13). This data highlighted varying assessments of the occurrence of AI-generated CSAM across population and parental status.

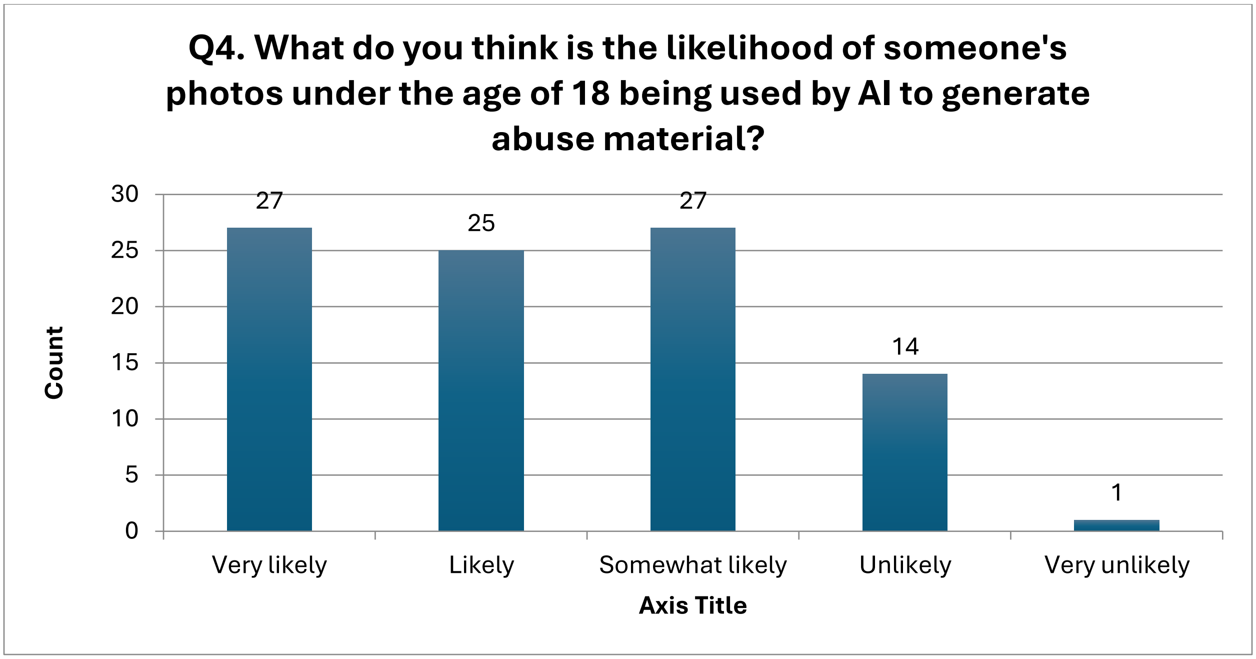
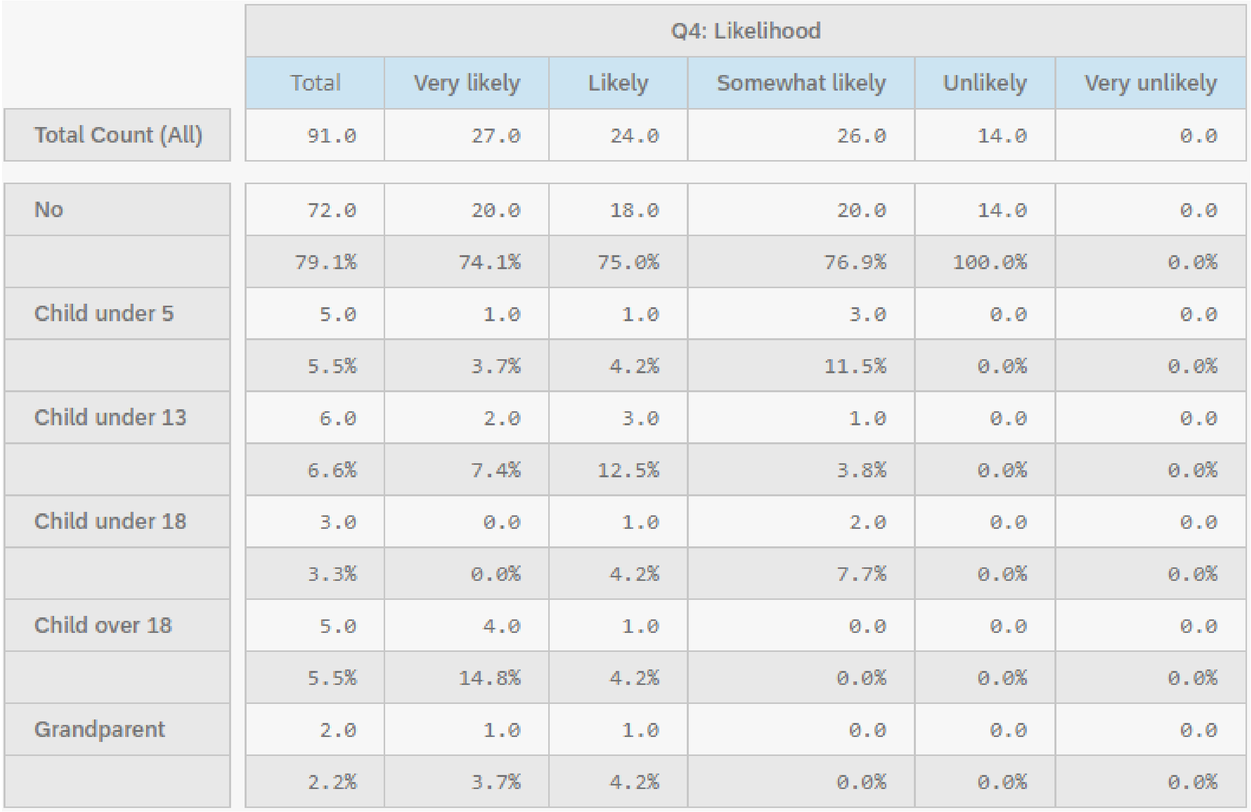


Figure 12. Participant reported likelihood of a minor's photo being used to produce AI-generated CSAM



	Q4: Likelihood					
	Total	Very likely	Likely	Somewhat likely	Unlikely	Very unlikely
Total Count (All)	91.0	27.0	24.0	26.0	14.0	0.0
No	72.0	20.0	18.0	20.0	14.0	0.0
	79.1%	74.1%	75.0%	76.9%	100.0%	0.0%
Child under 5	5.0	1.0	1.0	3.0	0.0	0.0
	5.5%	3.7%	4.2%	11.5%	0.0%	0.0%
Child under 13	6.0	2.0	3.0	1.0	0.0	0.0
	6.6%	7.4%	12.5%	3.8%	0.0%	0.0%
Child under 18	3.0	0.0	1.0	2.0	0.0	0.0
	3.3%	0.0%	4.2%	7.7%	0.0%	0.0%
Child over 18	5.0	4.0	1.0	0.0	0.0	0.0
	5.5%	14.8%	4.2%	0.0%	0.0%	0.0%
Grandparent	2.0	1.0	1.0	0.0	0.0	0.0
	2.2%	3.7%	4.2%	0.0%	0.0%	0.0%

Figure 13. Crosstabulation of question 4 and parental status

This is another area where it is difficult to determine the likelihood that photos would be used in these models however the AI image dataset LAION-5B has been found to contain child abuse images (Thiel, 2023) along with photos of Australian children scraped from the internet (Kemp, 2024). As a result, experts advise parents to avoid posting pictures of their children online especially in public places to avoid these photos from being included in these datasets.

Law Enforcement Challenges

The findings, particularly from the experts, also revealed significant challenges that law enforcement and other regulatory bodies face when regulating AI-generated CSAM. Experts explained how law enforcement agencies currently struggle to detect and manage the vast scale of AI-generated CSAM. Further developments of generative AI image models will in turn make AI-generated CSAM more realistic leading to increased difficulties for law enforcement in determining if CSAM content involves physical abuse or is AI-generated. One expert highlighted:

> *"Lifelike AI-generated CSAM complicates real victim identification, as it becomes increasingly difficult for law enforcement to distinguish between real people and AI composites."* – Expert 2

Experts also highlighted ethical concerns about informing victims and families, especially if they were unaware of the images, informing victims must be done by weighing up the psychological harms of informing them. Such concerns highlight the complex challenges in regulating AI-generated CSAM, and how these challenges are also rapidly evolving as technologies continue to develop (Flynn et al., 2022; Okolie, 2023). Not only did experts refer to the difficulties in regulating online abuse when it occurs on a large scale, but it was also stated that local police often lack the expertise in dealing with cases of online abuse.

Although studies have reported the need for regulation to improve as AI becomes more pervasive (Flynn et al., 2022), the experts highlighted the complexities in doing so as such reforms involve stakeholders across various local, state, and federal levels. Despite these complexities, studies have highlighted the responsibility that social service providers and law enforcement have in protecting youth populations from online abuse, through more proactive rather than reactive approaches (Caddle et al., 2023).

Reporting Pathway Challenges

In the context of a lack of public knowledge about support services, some survey respondents identified the need for an accessible and publicly recognised reporting system to be established. From our survey, we can see that the most common response was to report AI-generated CSAM to the platform with 63 people selecting it followed by 45 people selecting searching online for an answer, this indicates that the public has a limited understanding of what to do with the number of people selecting the "search online" option. With reporting to the platform being the most selected option this places the burden on platforms to adequately address instances of AI-generated CSAM. The experts interviewed state that there is a need for a clear and defined pathway for public reporting of AI-generated CSAM to ensure effective handling of reports rather than relying on the platforms themselves. The experts further highlighted that without such a pathway, current reporting mechanisms often conflate these cases with broader cybercrime issues. As a result, reports may be directed to inappropriate channels, where they are likely to be lost or overlooked. As they explained:

> *"Unfortunately, I think that there is probably a lot of conflation of this issue with the broader umbrella of cybercrime as well. And so, you might have a lot of reports that find their way to e-safety, or they find their way to report cyber or something like this, like other less appropriate reporting areas, and probably disappear into the ether."* – Expert 3

Supporting this view, Christensen and colleagues (2021) emphasised the need to improve CSAM reporting systems to address challenges in handling cross-border flows, especially between countries with differing legislation and enforcement. This is to avoid AI-generated CSAM production moving to countries with weak enforcement. They suggested that these systems should provide a non-intimidating and anonymous interface between the public and law enforcement to encourage public reporting (Christensen et al., 2021). Additionally, they argued that law enforcement agencies should offer regular feedback to show how public contributions have positively impacted efforts to combat the issue, thus fostering ongoing public engagement (Christensen et al., 2021). The International Association of Internet Hotlines (INHOPE) further identified key elements necessary for effective reporting systems, such as ensuring adequate infrastructure and staff, managing and analysing reports efficiently, and raising public awareness to encourage more frequent reporting (Christensen et al., 2021).

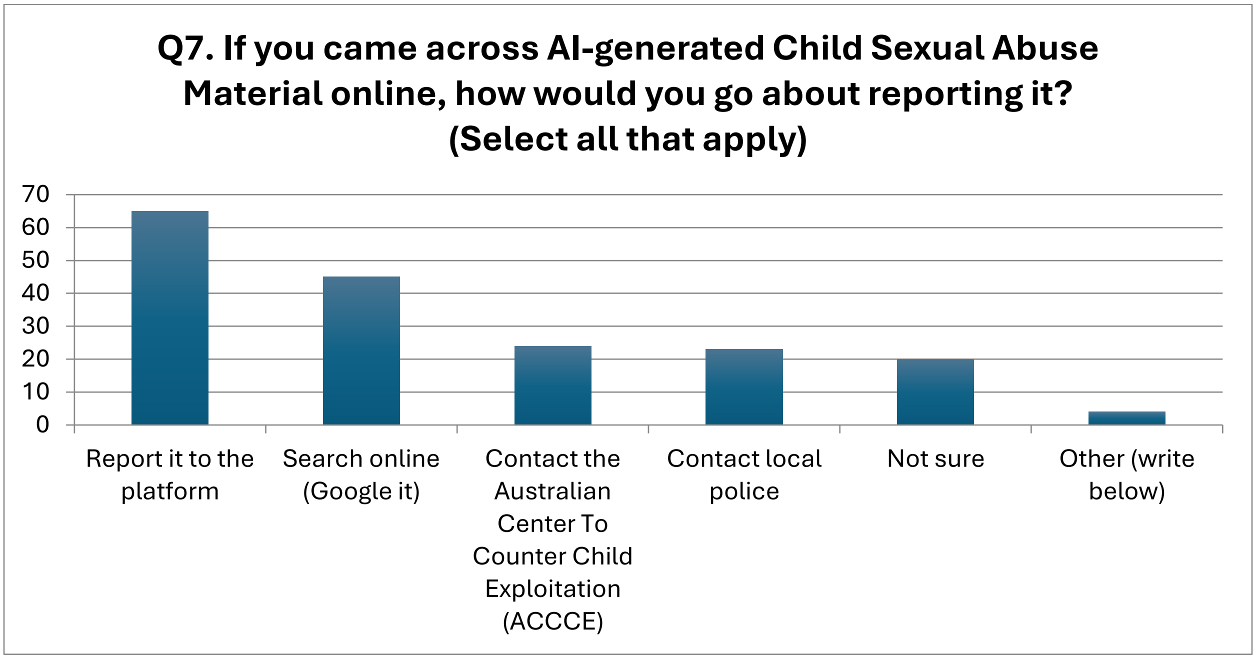


Figure 14. Participant reporting pathway suggestions

Systemic Gaps

The data also revealed multiple systemic gaps in Australia regarding the governance of AI-generated CSAM. This theme has been separated into three main categories: Legislative and Regulatory Reforms, Aligning Australia with Emerging International Norms, and Technological Reform and Accountability.

Legislative and Regulatory Reforms

Online survey participants expressed a need for legislative and regulatory improvements to reduce AI-generated CSAM. In the written response section, several respondents advocated for imposing stricter legal consequences to deter both creators and consumers of this material. Additionally, some suggested the establishment of new legislation specifically designed to address AI-generated CSAM. One respondent emphasised the critical role of international agreements and partnerships in strengthening legislative efforts against AI-generated CSAM. Furthermore, others stressed the need for regulations that would hold generative AI providers accountable for monitoring and preventing the production and dissemination of AI-generated CSAM.

Similar reforms were also suggested by the experts who highlighted gaps in the current Australian system and identified areas for improvement. Expert participants indicated that the existing Australian legislation, including the *Online Safety Act*, *Criminal Code Act*, and *Privacy Act*, requires reform to enhance the protection of children against the online harm associated with AI-generated CSAM. Experts further emphasised the importance of establishing a well-defined threshold in the law to clarify what constitutes AI-generated CSAM. They advocated for improved legal responses that consider the root causes of the material and the age of the individual generating it, rather than relying on a one-size-fits-all legislative approach:

> *“Well, how are we going to treat these individuals who are actually perpetrating, and how do we differentiate between the root causes of, whether it’s a sexual offence or if it's 100% bullying within the context of the school environment.”* – Expert 2

Aligning Australia with Emerging International Norms

One expert called for the adoption of a duty-of-care approach, whereby platforms will have to proactively remove AI-generated CSAM. There is a need to transition away from the current notice-and-take-down model in the *Online Safety Act*, which is reliant on AI-generated CSAM being reported primarily by users. This shift would position Australia in alignment with emerging international norms, particularly those established by the European Union and the United Kingdom. Two experts also encouraged Australia to examine the current international frameworks, learn from its practices, and adapt accordingly. This perspective aligns with the insights provided by Judson and colleagues (2024) regarding the duty-of-care approach mandated by the UK *Online Safety Act 2023*. The UK places responsibility onto platforms to identify, mitigate, and manage risks of harm, encompassing various risk mitigation tools that extend beyond merely removing content (Judson et al., 2024). For instance, platforms may enhance user friction or utilise chatbot interventions to manage illegal content effectively and facilitate prompt removal (Judson et al., 2024). However, the study also revealed that this method may inadvertently increase the risk of removing legitimate content, potentially infringing upon users' freedom of expression (Judson et al., 2024). Kira (2024) further emphasised the need to clarify the threshold issue within duty-of-care policies regarding the classification of AI-generated CSAM. Specifically, it is crucial to establish a clear legal threshold to determine whether such content falls under specific categories of prohibited material, such as nudity, bullying, harassment, or abuse (Kira, 2024).

Technological Reform & Accountability

Additionally, three of the four interviewed experts urged AI companies to enhance their awareness of the implications of developing and deploying AI technology, ensuring that it cannot be used to produce abuse material. They asserted that technology companies must prioritise "Safety by Design," which is an approach to address anticipated risks and safety issues of new technology early in the research, development, and design phases (van de Poel & Robaey, 2017). During the creation of apps and online platforms in which minors may access, developers must strengthen their focus on child protection (Okolie, 2023). Flynn and colleagues (2022) suggest that technology companies have a responsibility to educate their users on the dangers of deepfakes and connecting users to abuse support services. Overall, technological reform suggestions involved companies taking more responsibility by designing their software to be unable to produce abusive content, in addition to creating tools to identify, mitigate, and proactively remove it under a duty-of-care approach.

Conclusion

This study explored the social, policy and legislative considerations of AI-generated CSAM through balancing insights from the public and industry experts.

The first research question regarding public perceptions of AI-generated CSAM revealed that people have significant concerns regarding the risks and vulnerabilities imposed by AI-generated CSAM, but also revealed gaps in awareness. These gaps in awareness amongst the public involved a lack of familiarity with the issue of AI-generated CSAM itself, and where to go for help to access support services. Experts further shared the view that there is confusion from the public regarding the risks of AI-generated CSAM, in addition to local authorities often lacking the expertise to handle cases of online abuse.

The second research question regarding the evolving risks of AI-generated CSAM found that both experts and the public shared concerns regarding the psychological and social impacts on both children and adults. The increasing accessibility and ease of generation of AI-generated CSAM was seen to increase risks to children’s wellbeing and safety as AI is being increasingly used in school bullying.

The third research question regarding the inadequacies of the current Australian system found numerous gaps in current regulation practices. Expert insights revealed the difficulties for law enforcement to cope with the increasing cases of digital abuse including AI-generated CSAM, the need for Australia to align with emerging international legislation, and the need for policy reforms to address cases of AI-generated CSAM through implementing well-defined thresholds of what is and is not AI-generated CSAM.

The authors of this study acknowledge that the scope of this study has been limited by the convenience and purposive sampling of its participants and is thus not indicative of the entire Australian population. Regardless, the vulnerability concerns, systemic gaps and suggestions shared by the participants may help inform future policies, practices and research in the online child protection space.

Recommendations

Based on the findings, the researchers make the following recommendations for future policies, practices and research:

- **Awareness efforts**: Educational and awareness campaigns should be targeted towards specific populations such as teachers and public awareness campaigns must be done in such a way to avoid unnecessary alarmism.
- **Legislative and regulatory reforms**: There is a need for better-defined thresholds in the law to clarify what constitutes AI-generated CSAM. Reforms are also required on national and international levels to specifically address AI-generated CSAM and prevent it from being produced in countries with lax regulation.
- **Accountability**: Tech companies must be held accountable to ensure safety through prioritising efforts such as “Safety-by-Design".
- **Improvements in reporting pathways and support services**: The awareness, quality and accessibility of reporting pathways and support services regarding online abuse need to be improved. Schools and educational institutions are best placed to educate young people about the dangers and reporting pathways of AI-generated CSAM, as evident through current initiatives by the eSafety Commissioner, but must continue to be updated to address the evolving nature of AI-generated CSAM.
- **Future research:** Given the limited scope of this study regarding participant recruitment from the public, future research may include more insights from people belonging to older demographics. This may involve more insights from parents and caregivers with children of various ages, who can share further insights regarding the firsthand risks of AI-generated CSAM on children and their families.

References

Adams, C., Pente, P., Lemermeyer, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education. *Computers and Education. Artificial Intelligence*, *4*, 100131. https://doi.org/10.1016/j.caeai.2023.100131

Australia Federal Police. (2024, June 31). *Victorian man jailed for producing almost 800 AI-generated child abuse images | Australian Federal Police*. Afp.gov.au. https://www.afp.gov.au/news-centre/media-release/victorian-man-jailed-producing-almost-800-ai-generated-child-abuse-images

Australian Federal Police. (2024, March 30). *Tasmanian jailed for possessing AI-generated child abuse material | Australian Federal Police*. Www.afp.gov.au. https://www.afp.gov.au/news-centre/media-release/tasmanian-jailed-possessing-ai-generated-child-abuse-material

Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. *IEEE Access*, *10*, 77110–77122. https://doi.org/10.1109/ACCESS.2022.3191790

Brierley, J. A. (2017). The role of a pragmatist paradigm when adopting mixed methods in behavioural accounting research. *International Journal of Behavioural Accounting and Finance*, *6*(2), 140–154. https://doi.org/10.1504/IJBAF.2017.086432

Caddle, X. V., Naher, N., Miller, Z. P., Badillo-Urquiola, K., & Wisniewski, P. J. (2023). Duty to Respond: The Challenges Social Service Providers Face When Charged with Keeping Youth Safe Online. *Proceedings of the ACM on Human-Computer Interaction*, *7*(GROUP), 1–35. https://doi.org/10.1145/3567556

Chen, J. J., & Lin, J. C. (2024). Artificial intelligence as a double-edged sword: Wielding the POWER principles to maximize its positive effects and minimize its negative effects. *Contemporary Issues in Early Childhood*, *25*(1), 146–153. https://doi.org/10.1177/14639491231169813

Christensen, L. S., Rayment-McHugh, S., Prenzler, T., Chiu, Y.-N., & Webster, J. (2021). The theory and evidence behind law enforcement strategies that combat child sexual abuse material. *International Journal of Police Science & Management*, 146135572110269. https://doi.org/10.1177/14613557211026935

Dennehy, R., Meaney, S., Cronin, M., & Arensman, E. (2020). The psychosocial impacts of cybervictimisation and barriers to seeking social support: Young people’s perspectives. *Children and Youth Services Review*, *111*, 104872. https://doi.org/10.1016/j.childyouth.2020.104872

ESafety Commissioner. (2023, October 20). *Image-based abuse*. ESafety Commissioner. https://www.esafety.gov.au/key-topics/image-based-abuse

Flynn, A., Powell, A., Scott, A. J., & Cama, E. (2022). Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse. *British Journal of Criminology*, *62*(6), 1341–1358. https://doi.org/10.1093/bjc/azab111

Fong, H. F., Bennett, C. E., Mondestin, V., Scribano, P. V., Mollen, C., & Wood, J. N. (2020). The Impact of Child Sexual Abuse Discovery on Caregivers and Families: A Qualitative Study. *Journal of interpersonal violence*, *35*(21-22), 4189–4215. https://doi.org/10.1177/0886260517714437

ICMEC Australia. (2023, June 27). What does Generative AI mean for CSE? https://icmec.org.au/blog/what-does-generative-ai-mean-for-cse/

ICMEC Australia. (2024). ABOUT. https://icmec.org.au/about/#section-16-13

Janjeva, A., Harris, A., Mercer, S., Kasprzyk, A., & Gausen, A. (2023). The Rapid Rise of Generative AI. *CETaS Research Reports*.

Judson, E., Kira, B., & Howard, J. W. (2024). The Bypass Strategy: platforms, the Online Safety Act and future of online speech. *Journal of Media Law*, 1–22. https://doi.org/10.1080/17577632.2024.2361524

Kemp, K. (2024, July 3). Photos of Australian kids have been found in a massive AI training data set. What can we do? The Conversation. http://theconversation.com/photos-of-australian-kids-have-been-found-in-a-massive-ai-training-data-set-what-can-we-do-233868

Kira, B. (2024). When non-consensual intimate deepfakes go viral: The insufficiency of the UK Online Safety Act. *Computer Law & Security Review*, *54*, 106024–106024. https://doi.org/10.1016/j.clsr.2024.106024

Krishna, A. (2021). Internet.gov: Tech companies as government agents and the future of the fight against child sexual abuse. *California Law Review*, *109*(4), 1581–1635. https://doi.org/10.15779/Z38KW57J9B

Maier, B. (2010). How Has the Law Attempted to Tackle the Borderless Nature of the Internet? *International Journal of Law and Information Technology*, *18*(2), 142–175. https://doi.org/10.1093/ijlit/eaq001

Martin, J. (2014). “It’s Just an Image, Right?”: Practitioners’ Understanding of Child Sexual Abuse Images Online and Effects on Victims. *Child & Youth Services*, *35*(2), 96–115. https://doi.org/10.1080/0145935X.2014.924334

Martin, N. (2021). Image-Based Sexual Abuse and Deepfakes: A Survivor Turned Activist's Perspective. *The Palgrave Handbook of Gendered Violence and Technology*, 55–72. https://doi.org/10.1007/978-3-030-83734-1_4

Martineau, K. (2023, April 20). *What is generative AI?* IBM Research Blog; IBM. https://research.ibm.com/blog/what-is-generative-AI

McGlynn, C., Johnson, K., Rackley, E., Henry, N., Gavey, N., Flynn, A., & Powell, A. (2021). 'It's Torture for the Soul': The Harms of Image-Based Sexual Abuse. *Social & Legal Studies*, *30*(4), 541–562. https://doi.org/10.1177/0964663920947791

Nasser, M., & Welch, V. (2013). Prioritization of systematic reviews leads prioritization of research gaps and needs. *Journal of Clinical Epidemiology*, *66*(5), 522–523. https://doi.org/10.1016/j.jclinepi.2012.09.007

Okolie, C. (2023). Artificial Intelligence-Altered Videos (Deepfakes) and Data Privacy Concerns. Journal of International Women's Studies. 25. 13.

Oxford English Dictionary. (2023). *Deepfake, N.* https://doi.org/10.1093/OED/9547101155.

Rimer, J. (2024, April 17). Child Sexual Exploitation. *Oxford Research Encyclopedia of Criminology.* Retrieved 3 Nov. 2024, from https://oxfordre.com/criminology/view/10.1093/acrefore/9780190264079.001.0001/acrefore-9780190264079-e-780.

Robinson, K. A., Saldanha, I. J., & Mckoy, N. A. (2011). Development of a framework to identify research gaps from systematic reviews. *Journal of Clinical Epidemiology*, *64*(12), 1325–1330. https://doi.org/10.1016/j.jclinepi.2011.06.009

SACE. (2022). *Victim Blaming*. Sexual Assault Centre of Edmonton. https://www.sace.ca/learn/victim-blaming/

Sarma, A. (2022). Promoting Child Safety Online in the Time of COVID-19: The Indian and Australian Responses. ORF Issue Brief, 557. Observer Research Foundation.

Schmidt, F., Varese, F., & Bucci, S. (2023). Understanding the prolonged impact of online sexual abuse occurring in childhood. *Frontiers in Psychology*, *14*, 1281996–1281996. https://doi.org/10.3389/fpsyg.2023.1281996

School Community Engagement Plan | eSaftey commissioner. (2022)Retrieved November 4, 2024, from https://www.esafety.gov.au/sites/default/files/202202/Engage%201%20-%20School%20community%20engagement%20plan.pdf

Selwyn, N., & Cordoba, B. (2022). Australian public understandings of artificial intelligence. *AI & Society*, *37*(4), 1645–1662. https://doi.org/10.1007/s00146-021-01268-z

Thiel, D., Stroebel, M., & Portnoff, R. (2023). Generative ML and CSAM: Implications and Mitigations. *Standford Internet Observatory Cyber Policy Center*. https://fsi.stanford.edu/publication/generative-ml-and-csam-implications-and-mitigations

van de Poel, I., & Robaey, Z. (2017). Safe-by-Design: from Safety to Responsibility. NanoEthics, 11(3), 297–306. https://doi.org/10.1007/s11569-017-0301-x

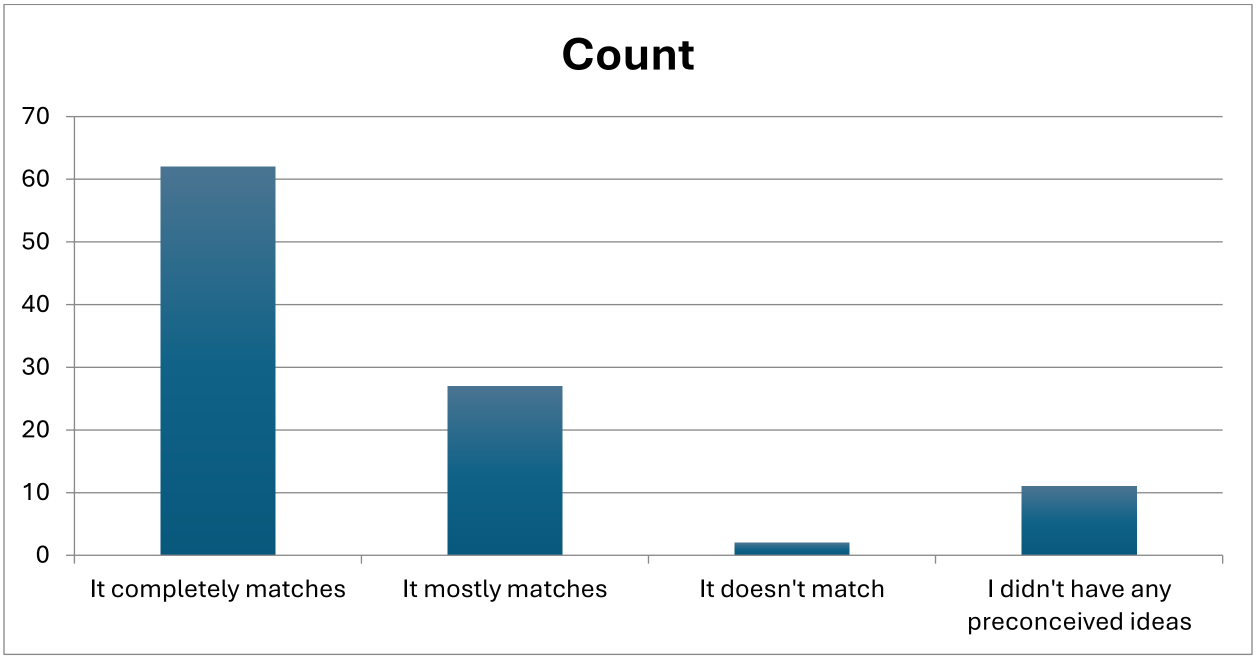
Yasmine, H. B. (2024). Online Child Sexual Abuse: Exploring Psychological and Social Impacts, Prevention, and Intervention Strategies. *Journal of Studies in Deviation Psychology*. https://www.asjp.cerist.dz/en/downArticle/704/9/1/249151

Appendix A

Quantitative survey results

Survey data

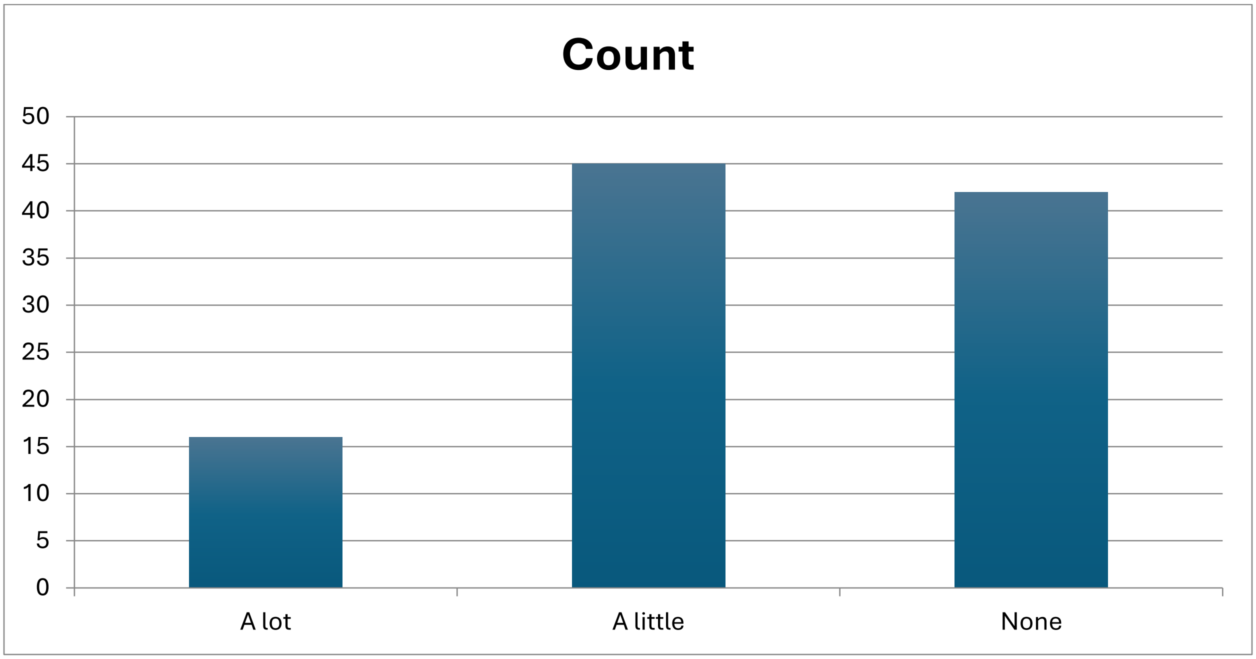
Q1. Does the definition given (AI-generated Child Sexual Abuse Material (CSAM) can be defined as the use of Artificial Intelligence to create sexualised images or videos of children that are either: Digitally altered to depict a real child (deepfakes); or Completely generated depicting a child that doesn't exist. match your preconceived ideas of what AI-generated Child Sexual Abuse material is?)



(Figure 7.)

Categorical Summary			
Sample Size			
102			
Preconceived ideas	Count	Percent of Data	Confidence Interval (Percent of Data)
It completely matches	62	60.8%	51.1% to 69.7%
It mostly matches	27	26.5%	18.9% to 35.8%
It doesn't match	2	2.0%	0.5% to 6.9%
I didn't have any preconceived ideas	11	10.8%	6.1% to 18.3%

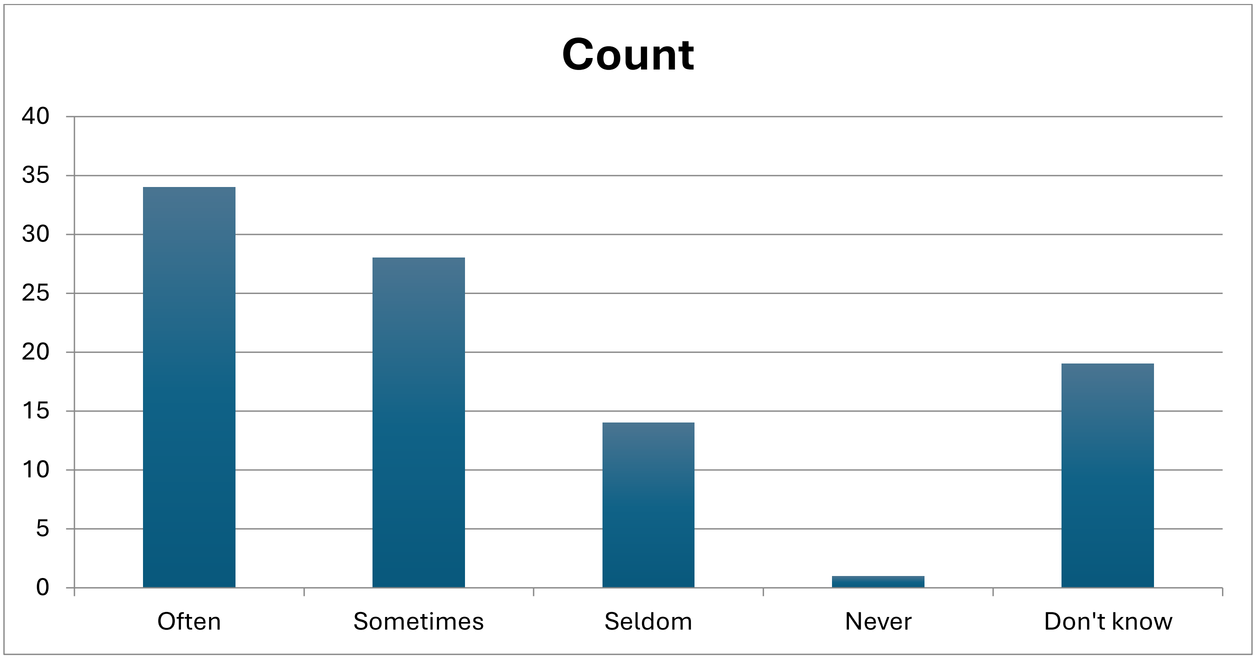
Q2. Noting this, how much have you seen, read or heard about AI-generated Child Sexual Abuse Material as an issue?



(Figure 6.)

Categorical Summary			
Sample Size			
103			
Hearing about AI CSAM as an issue	Count	Percent of Data	Confidence Interval (Percent of Data)
A lot	16	15.5%	9.8% to 23.8%
A little	45	43.7%	34.5% to 53.3%
None	42	40.8%	31.8% to 50.4%

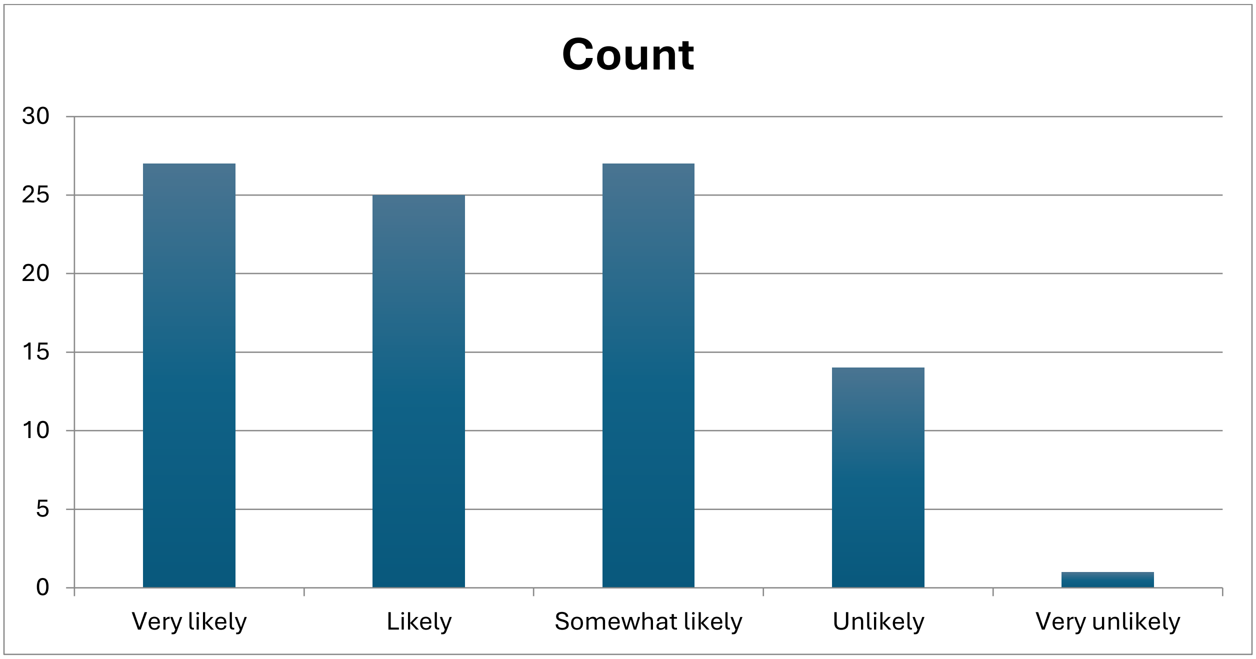
Q3. How often do you think AI-generated Child Sexual Abuse Material is occurring in Australia?



(Figure 11.)

Categorical Summary			
Sample Size			
96			
Occurrence	Count	Percent of Data	Confidence Interval (Percent of Data)
Often	34	35.4%	26.6% to 45.4%
Sometimes	28	29.2%	21.0% to 38.9%
Seldom	14	14.6%	8.9% to 23.0%
Never	1	1.0%	0.2% to 5.7%
Don't know	19	19.8%	13.1% to 28.9%

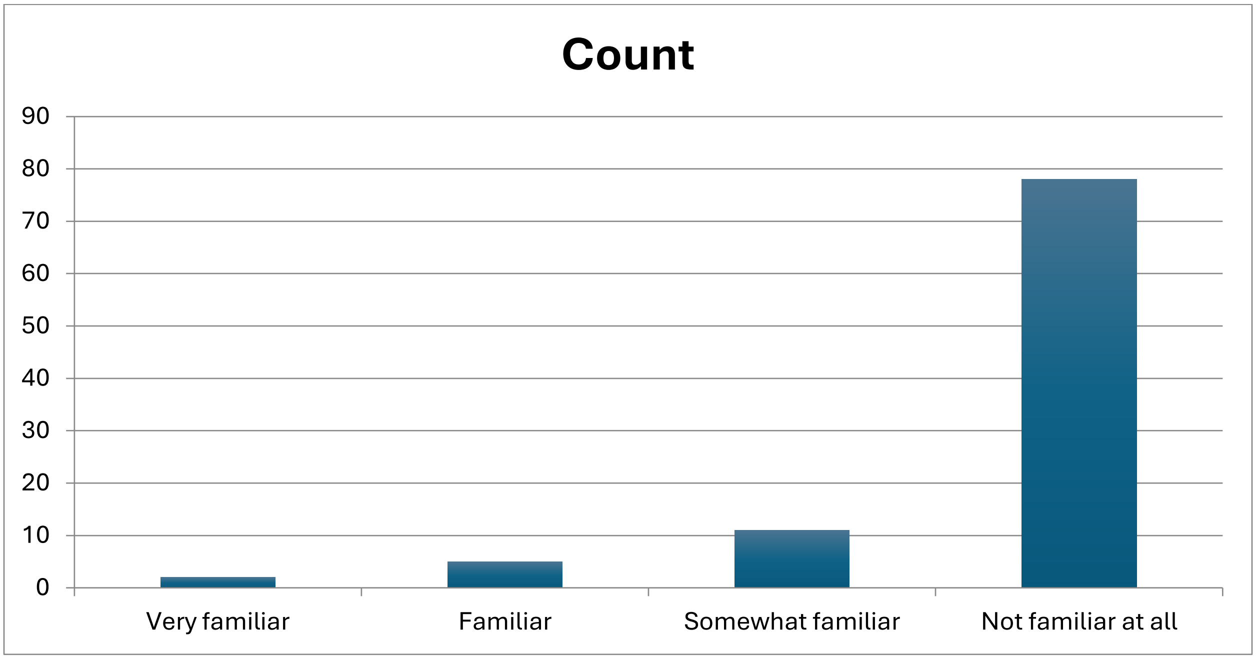
Q4. What do you think is the likelihood of someone's photos under the age of 18 being used by AI to generate abuse material?



(figure 12.)

Categorical Summary			
Sample Size			
94			
Likelihood	Count	Percent of Data	Confidence Interval (Percent of Data)
Very likely	27	28.7%	20.6% to 38.6%
Likely	25	26.6%	18.7% to 36.3%
Somewhat likely	27	28.7%	20.6% to 38.6%
Unlikely	14	14.9%	9.1% to 23.5%
Very unlikely	1	1.1%	0.2% to 5.8%

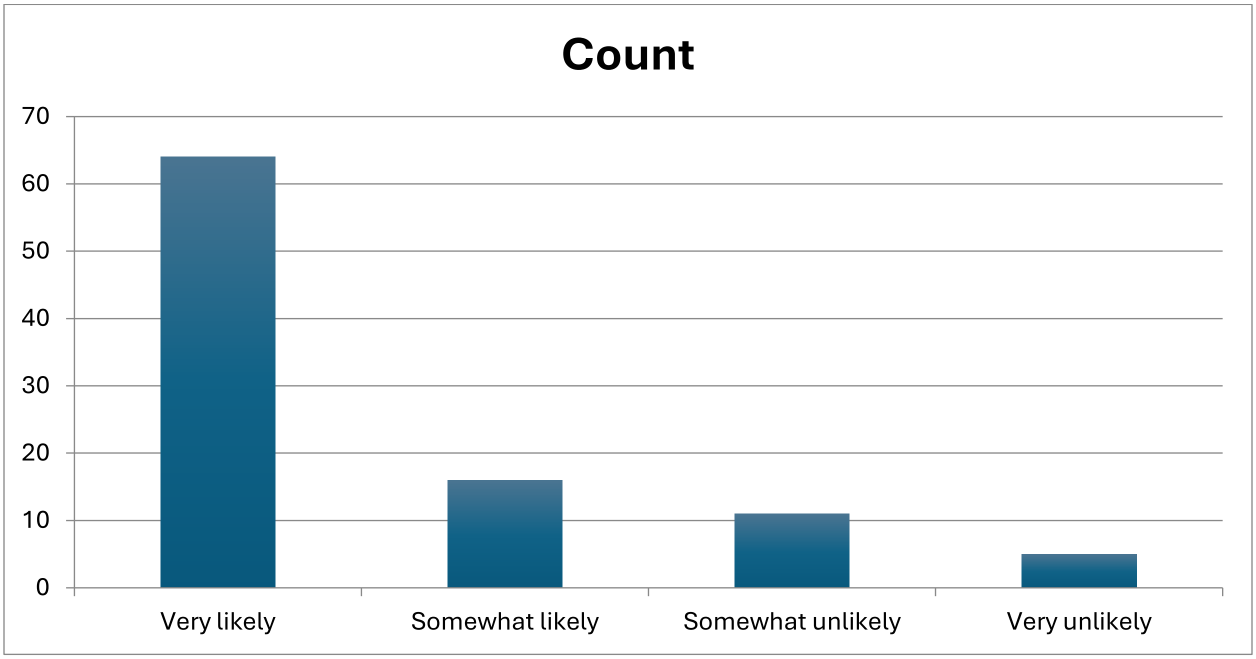
Q5. How familiar are you with existing support services and resources available for individuals affected by AI-generated Child Sexual Abuse Material?



(Figure 9.)

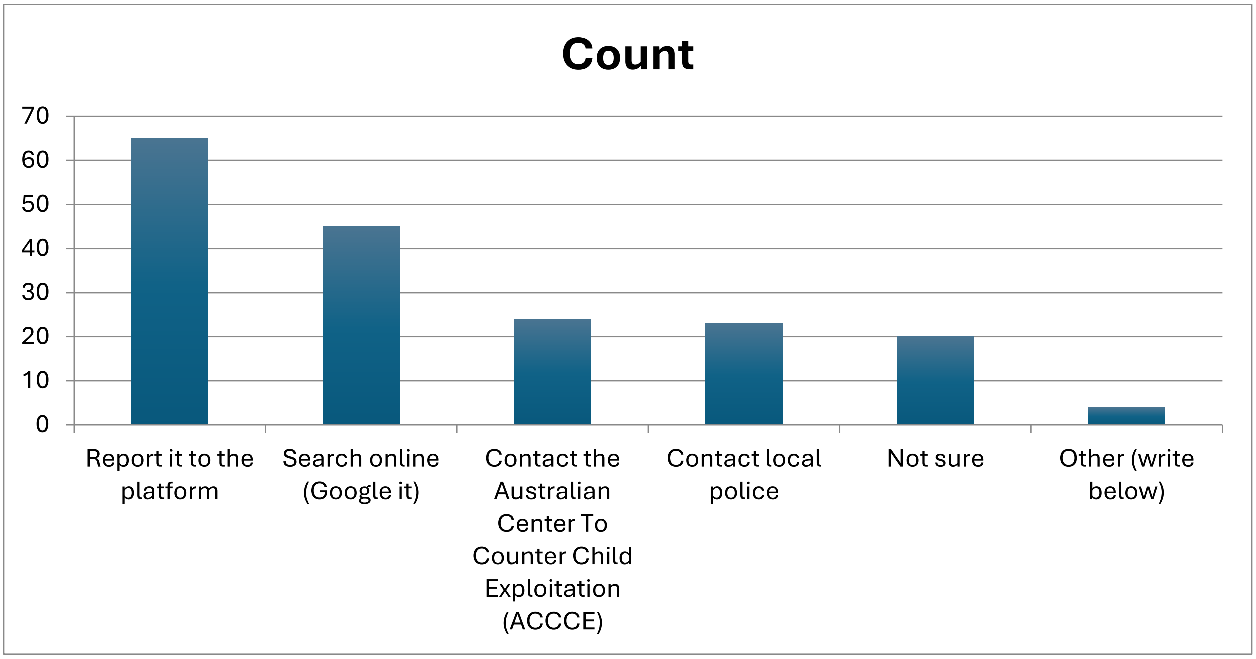
Categorical Summary			
Sample Size			
96			
Support services	Count	Percent of Data	Confidence Interval (Percent of Data)
Very familiar	2	2.1%	0.6% to 7.3%
Familiar	5	5.2%	2.2% to 11.6%
Somewhat familiar	11	11.5%	6.5% to 19.4%
Not familiar at all	78	81.3%	72.3% to 87.8%

Q6. How likely are you to report suspected instances of AI-generated Child Sexual Abuse Material if you encounter them online?



Categorical Summary			
Sample Size			
96			
Reporting content	Count	Percent of Data	Confidence Interval (Percent of Data)
Very likely	64	66.7%	56.8% to 75.3%
Somewhat likely	16	16.7%	10.5% to 25.4%
Somewhat unlikely	11	11.5%	6.5% to 19.4%
Very unlikely	5	5.2%	2.2% to 11.6%

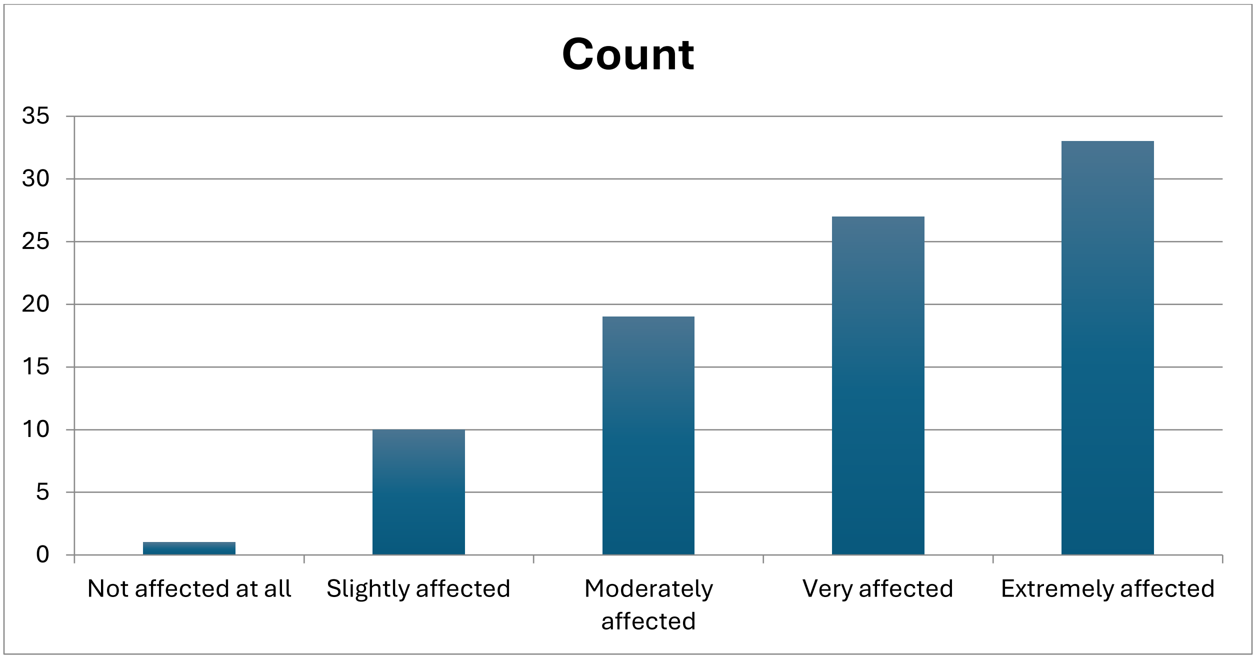
Q7. If you came across AI-generated Child Sexual Abuse Material online, how would you go about reporting it? (Select all that apply)



(Figure 14.)

Categorical Summary			
Sample Size			
93			
Choice	Checked Percent	Confidence Interval	Checked Count
Report it to the platform	67.7%	57.7% to 76.4%	63
Search online (Google it)	48.4%	38.5% to 58.4%	45
Contact the ACCCE	24.7%	17.1% to 34.4%	23
Contact local police	23.7%	16.2% to 33.2%	22
Not sure	20.4%	13.5% to 29.7%	19
Other (write below)	4.3%	1.7% to 10.5%	4

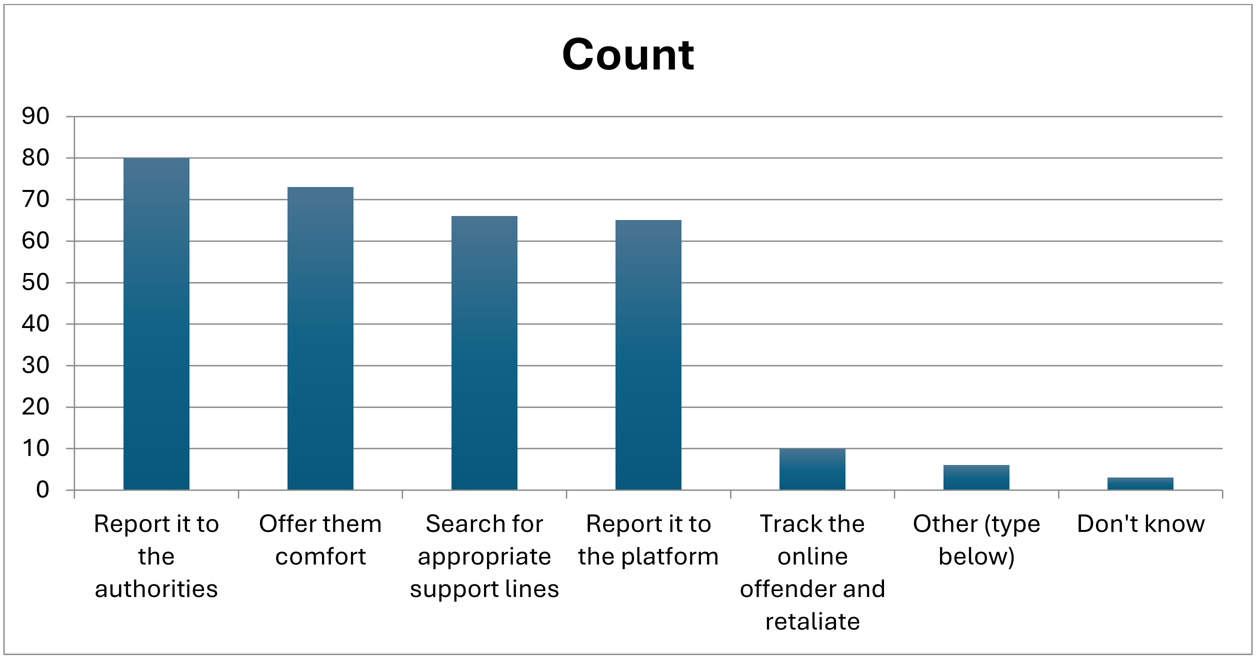
Q9. On a scale of 1 to 5, If you encountered AI-generated Child Sexual Abuse Material online, to what extent would it affect you?



(figure 5.)

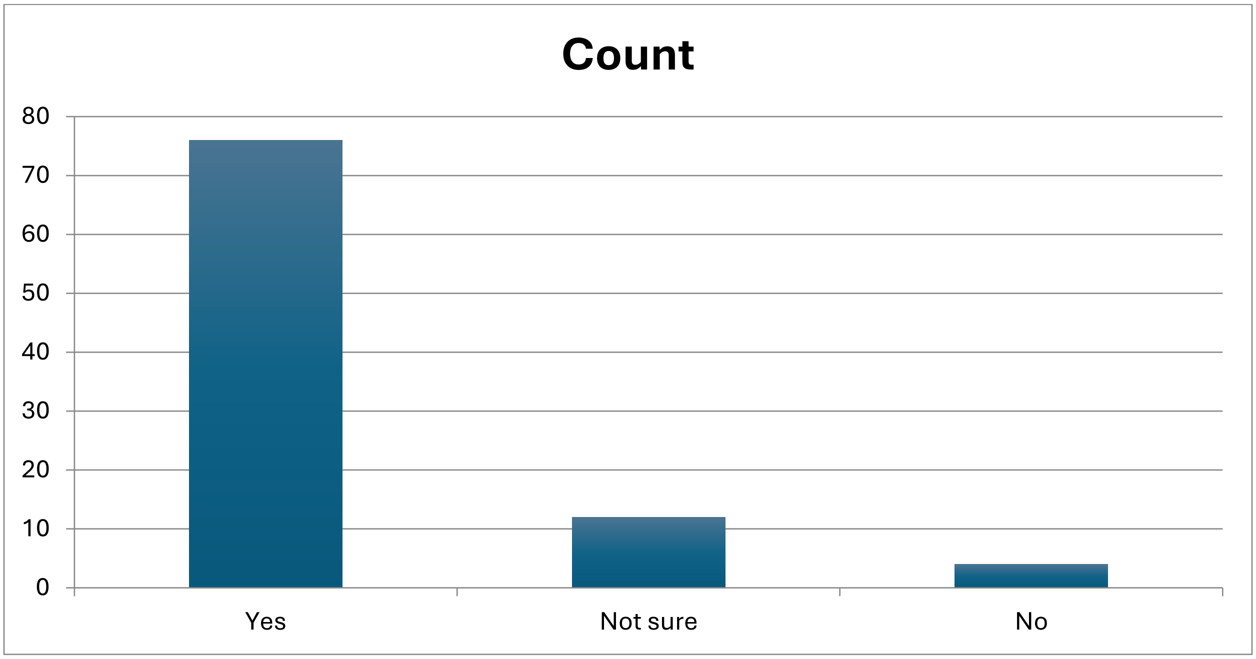
Categorical Summary			
Sample Size			
90			
Affect	Count	Percent of Data	Confidence Interval (Percent of Data)
Not affected at all	1	1.1%	0.2% to 6.0%
Slightly affected	10	11.1%	6.1% to 19.3%
Moderately affected	19	21.1%	14.0% to 30.6%
Very affected	27	30.0%	21.5% to 40.1%
Extremely affected	33	36.7%	27.4% to 47.0%

Q10. How would you react if someone you knew was a victim of Deepfaked Child Sexual Abuse Material? (Select all that apply)



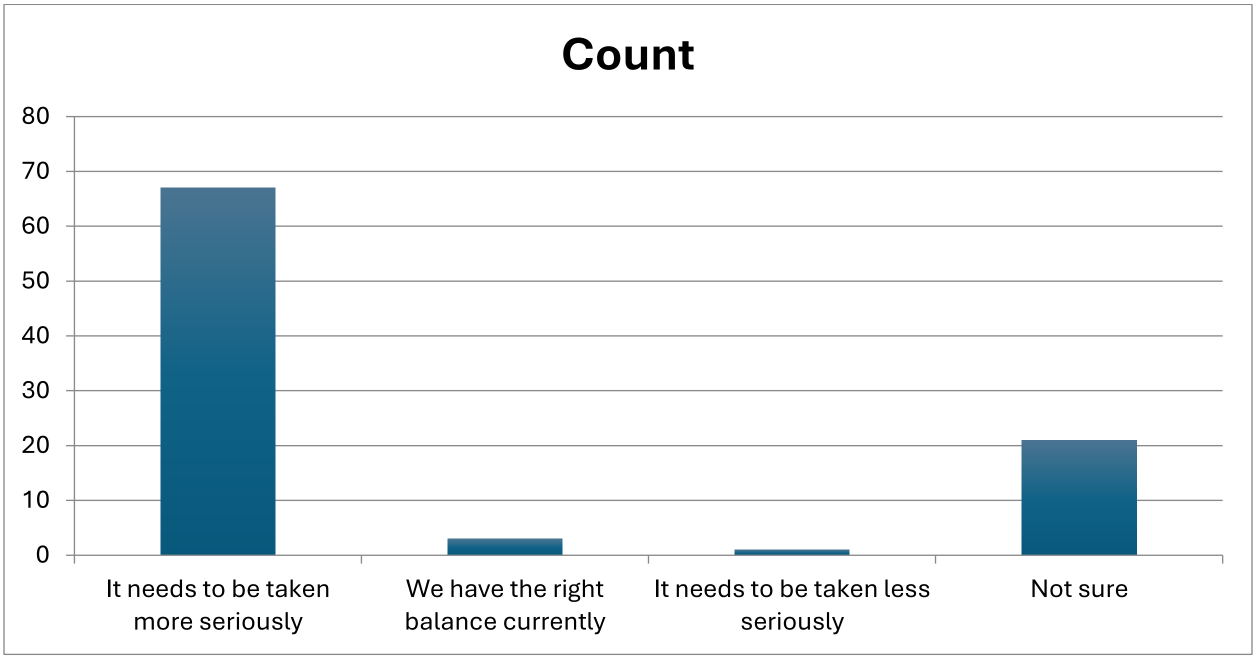
Categorical Summary			
Sample Size			
90			
Reaction	Checked Percent	Confidence Interval	Checked Count
Report it to the authorities	86.7%	78.1% to 92.2%	78
Offer them comfort	77.8%	68.2% to 85.1%	70
Search for appropriate support lines	71.1%	61.0% to 79.5%	64
Report it to the platform	70.0%	59.9% to 78.5%	63
Track the online offender and retaliate	11.1%	6.1% to 19.3%	10
Other (type below)	6.7%	3.1% to 13.8%	6
Don't know	2.2%	0.6% to 7.7%	2

Q11. Do you believe AI-generated Child Sexual Abuse Material should legally be treated in the same way as other forms of child sexual abuse material?



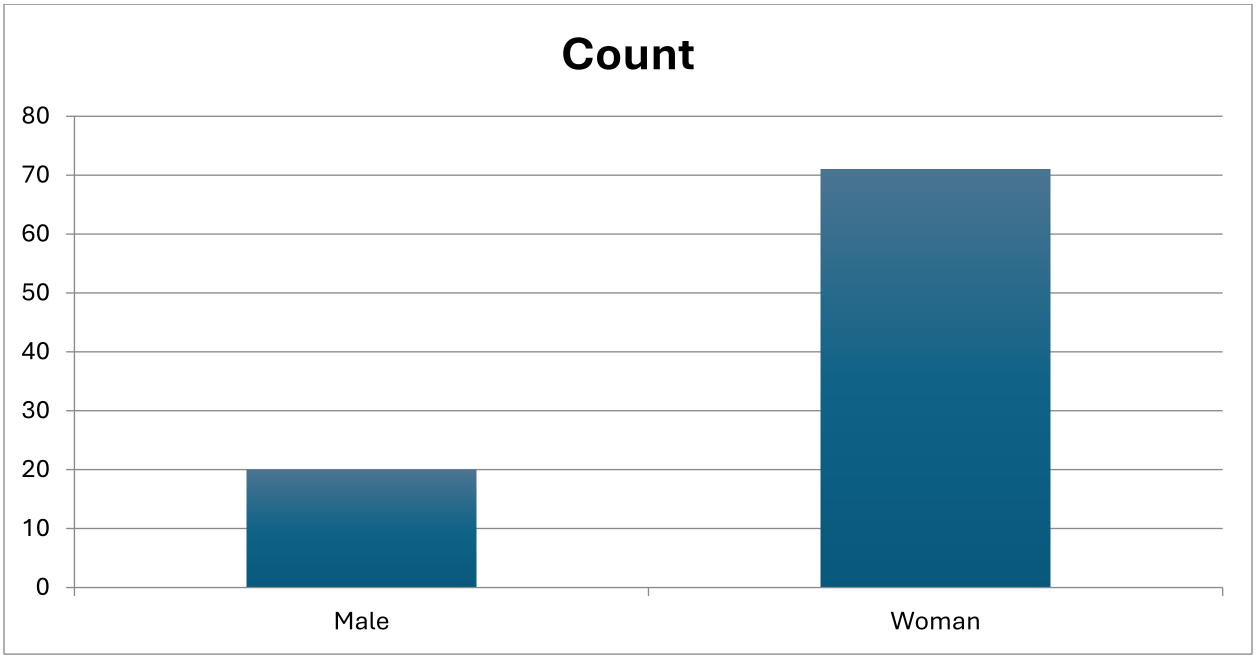
Categorical Summary			
Sample Size			
92			
Treatment as compared with non-AI forms	Count	Percent of Data	Confidence Interval (Percent of Data)
Yes	76	82.6%	73.6% to 89.0%
Not sure	12	13.0%	7.6% to 21.4%
No	4	4.3%	1.7% to 10.7%

Q12. Do you think AI-generated Child Sexual Abuse Material should be treated more seriously in Australia?



Categorical Summary			
Sample Size			
92			
Treatment more/less	Count	Percent of Data	Confidence Interval (Percent of Data)
It needs to be taken more seriously	67	72.8%	63.0% to 80.9%
We have the right balance currently	3	3.3%	1.1% to 9.2%
It needs to be taken less seriously	1	1.1%	0.2% to 5.9%
Not sure	21	22.8%	15.4% to 32.4%

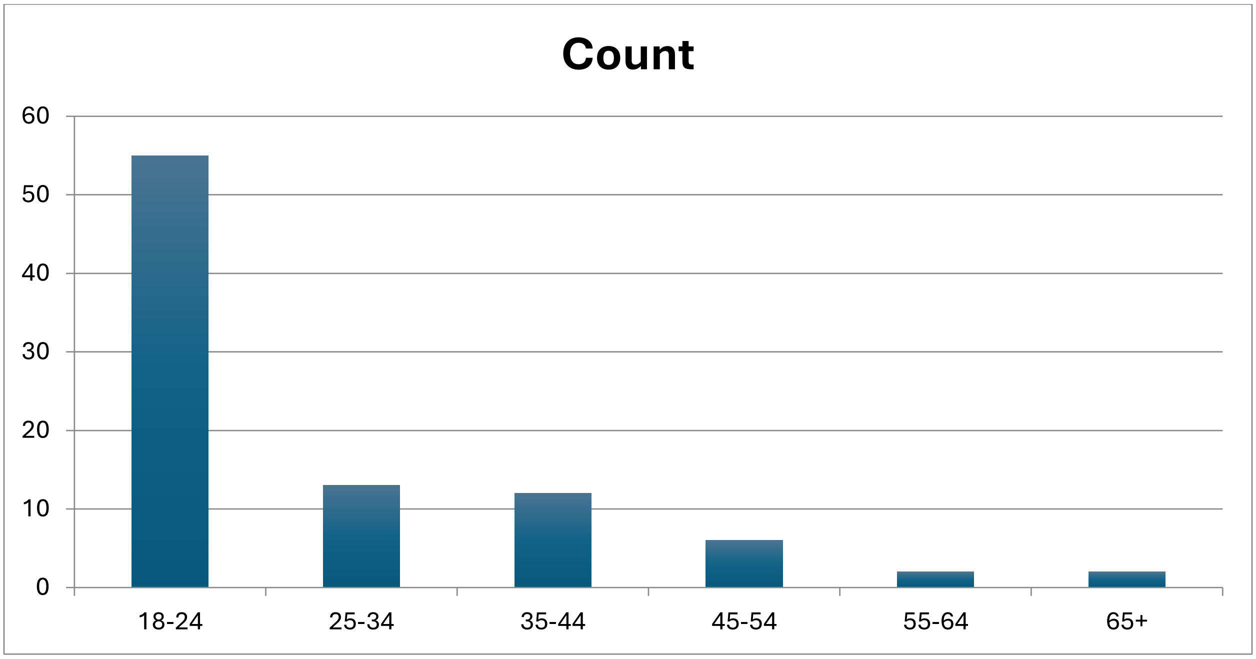
Q14.What gender do you identify as? (Zero respondents selected non-binary or other)



(Figure 1.)

Categorical Summary			
Sample Size			
91			
Gender - Selected Choice	Count	Percent of Data	Confidence Interval (Percent of Data)
Male	20	22.0%	14.7% to 31.5%
Woman	71	78.0%	68.5% to 85.3%

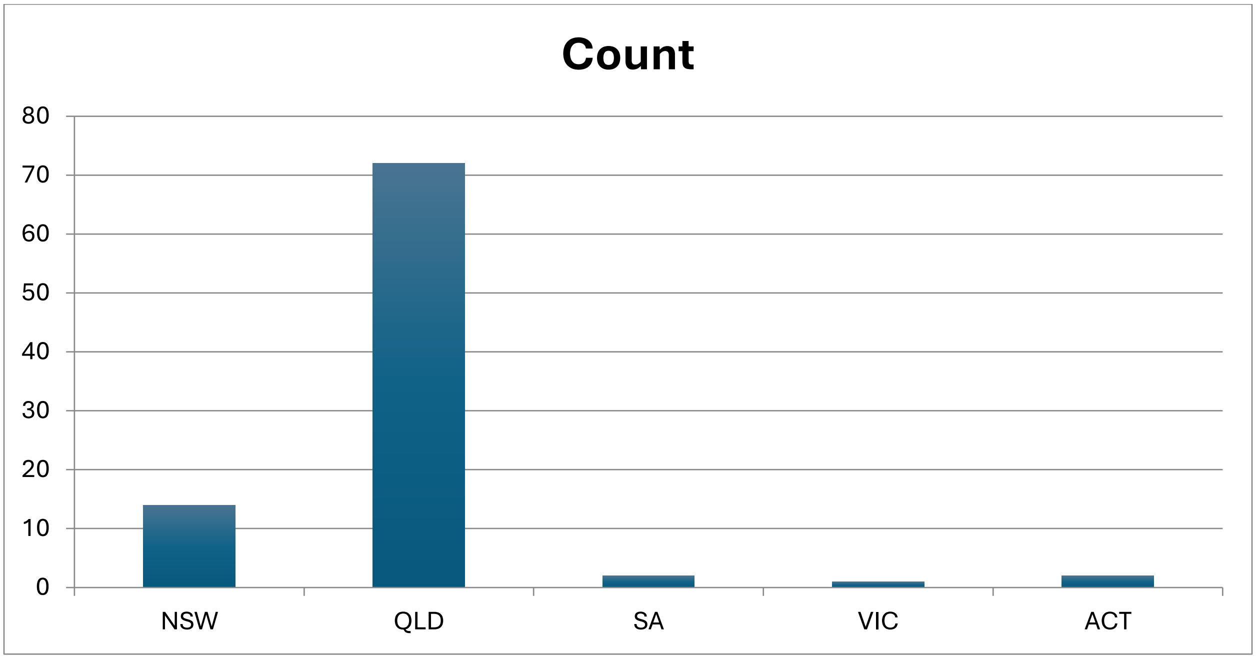
Q15. Your age



(Figure 2.)

Categorical Summary			
Sample Size			
90			
Age	Count	Percent of Data	Confidence Interval (Percent of Data)
18-24	55	61.1%	50.8% to 70.5%
25-34	13	14.4%	8.6% to 23.2%
35-44	12	13.3%	7.8% to 21.9%
45-54	6	6.7%	3.1% to 13.8%
55-64	2	2.2%	0.6% to 7.7%
65+	2	2.2%	0.6% to 7.7%

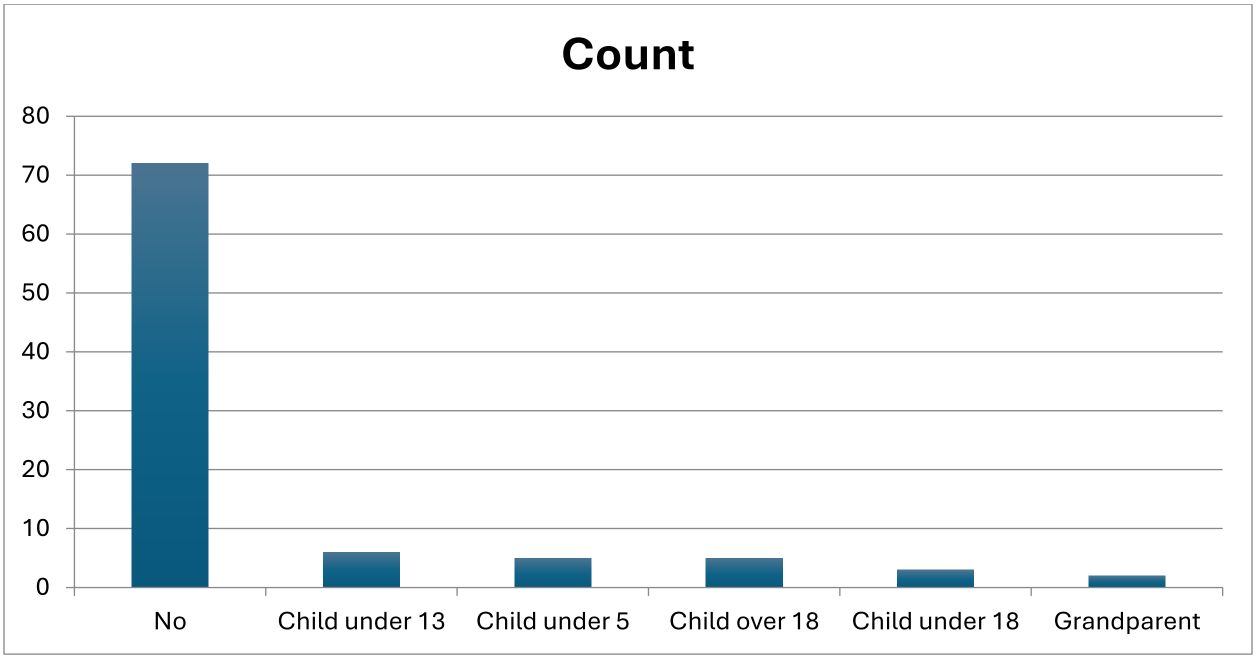
Q16. State of residence



(Figure 3.)

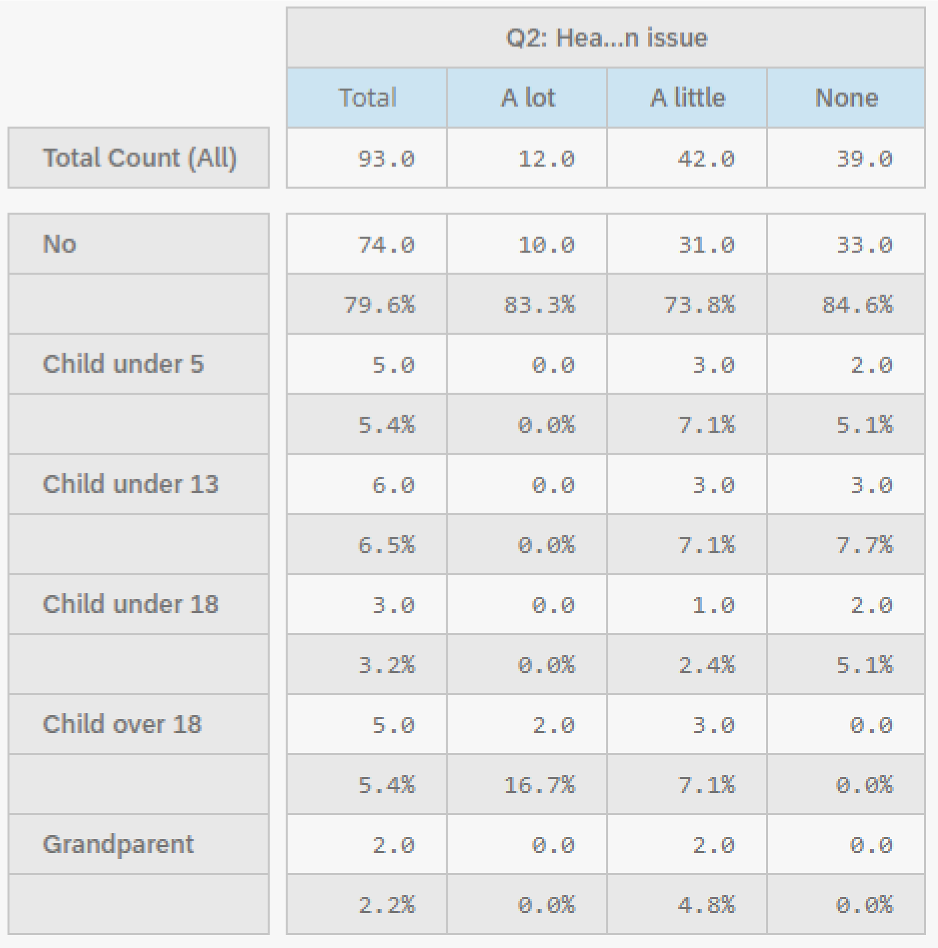
Categorical Summary			
Sample Size			
91			
State	Count	Percent of Data	Confidence Interval (Percent of Data)
NSW	14	15.4%	9.4% to 24.2%
QLD	72	79.1%	69.7% to 86.2%
SA	2	2.2%	0.6% to 7.7%
VIC	1	1.1%	0.2% to 6.0%
ACT	2	2.2%	0.6% to 7.7%

Q17. Are you a parent, grandparent or caregiver? (Select all that apply)



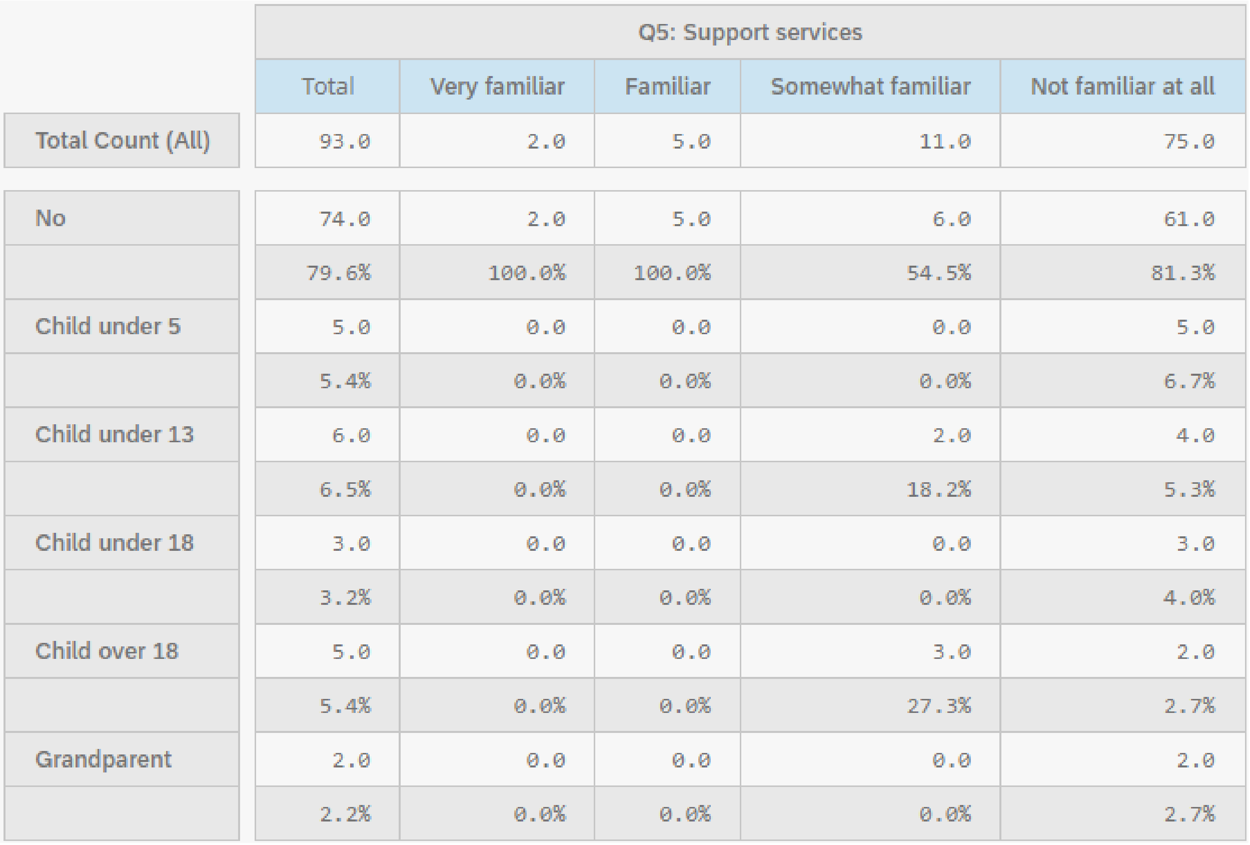
(figure 4.)

Categorical Summary			
Sample Size			
91			
Parent/caregiver	Checked Percent	Confidence Interval	Checked Count
No	79.1%	69.7% to 86.2%	72
Child under 13	6.6%	3.1% to 13.6%	6
Child under 5	5.5%	2.4% to 12.2%	5
Child over 18	5.5%	2.4% to 12.2%	5
Child under 18	3.3%	1.1% to 9.2%	3
Grandparent	2.2%	0.6% to 7.7%	2



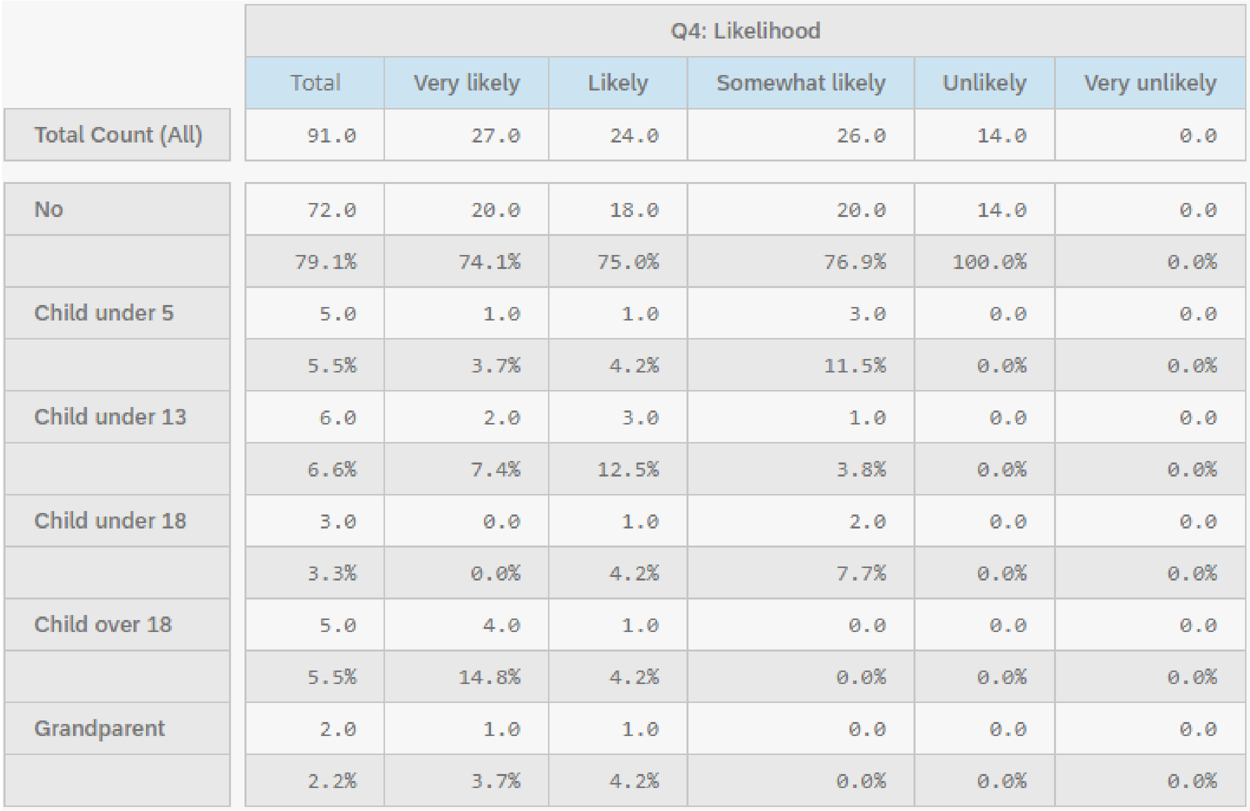
	Q2: Hea...n issue			
	Total	A lot	A little	None
Total Count (All)	93.0	12.0	42.0	39.0
No	74.0	10.0	31.0	33.0
	79.6%	83.3%	73.8%	84.6%
Child under 5	5.0	0.0	3.0	2.0
	5.4%	0.0%	7.1%	5.1%
Child under 13	6.0	0.0	3.0	3.0
	6.5%	0.0%	7.1%	7.7%
Child under 18	3.0	0.0	1.0	2.0
	3.2%	0.0%	2.4%	5.1%
Child over 18	5.0	2.0	3.0	0.0
	5.4%	16.7%	7.1%	0.0%
Grandparent	2.0	0.0	2.0	0.0
	2.2%	0.0%	4.8%	0.0%

(Figure 8. Crosstabulation of question 2 and parental status)



	Q5: Support services				
	Total	Very familiar	Familiar	Somewhat familiar	Not familiar at all
Total Count (All)	93.0	2.0	5.0	11.0	75.0
No	74.0	2.0	5.0	6.0	61.0
	79.6%	100.0%	100.0%	54.5%	81.3%
Child under 5	5.0	0.0	0.0	0.0	5.0
	5.4%	0.0%	0.0%	0.0%	6.7%
Child under 13	6.0	0.0	0.0	2.0	4.0
	6.5%	0.0%	0.0%	18.2%	5.3%
Child under 18	3.0	0.0	0.0	0.0	3.0
	3.2%	0.0%	0.0%	0.0%	4.0%
Child over 18	5.0	0.0	0.0	3.0	2.0
	5.4%	0.0%	0.0%	27.3%	2.7%
Grandparent	2.0	0.0	0.0	0.0	2.0
	2.2%	0.0%	0.0%	0.0%	2.7%

(Figure 10. Crosstabulation of question 5 and Parental status)



	Q4: Likelihood					
	Total	Very likely	Likely	Somewhat likely	Unlikely	Very unlikely
Total Count (All)	91.0	27.0	24.0	26.0	14.0	0.0
No	72.0	20.0	18.0	20.0	14.0	0.0
	79.1%	74.1%	75.0%	76.9%	100.0%	0.0%
Child under 5	5.0	1.0	1.0	3.0	0.0	0.0
	5.5%	3.7%	4.2%	11.5%	0.0%	0.0%
Child under 13	6.0	2.0	3.0	1.0	0.0	0.0
	6.6%	7.4%	12.5%	3.8%	0.0%	0.0%
Child under 18	3.0	0.0	1.0	2.0	0.0	0.0
	3.3%	0.0%	4.2%	7.7%	0.0%	0.0%
Child over 18	5.0	4.0	1.0	0.0	0.0	0.0
	5.5%	14.8%	4.2%	0.0%	0.0%	0.0%
Grandparent	2.0	1.0	1.0	0.0	0.0	0.0
	2.2%	3.7%	4.2%	0.0%	0.0%	0.0%

(Figure 13. Crosstabulation of question 4 and parental status)

Appendix B

Interview questions

1. Do you encounter cases regarding AI-generated Child Sexual Abuse material? If so, how often? Has the emergence of AI led to any increases compared to previous cases?
2. What are, in your experience or based on empirical evidence, the risks or challenges with AI-generated Child Sexual Abuse Material?
3. In your work, have you ever felt the need for the reality and risks of AI-generated CSAM to be taken more seriously?
4. How do you see that AI-generated CSAM affects children, their families and the wider community?
5. What challenges are there in detecting, investigating, and prosecuting AI-generated CSAM cases in Australia?
6. How often do you think AI-generated Child Sexual Abuse Material is occurring in Australia? How does this compare to what is happening internationally?
7. How do you believe current Australian legal and societal responses are addressing the issue of AI-generated CSAM, and what improvements do you suggest?
8. Thinking about the policies adopted by other countries to address the risks and challenges of AI-generated CSAM, what would work in Australia and what are the gaps?
9. Do you think technological advancements are introducing new risks to Australian society? If yes, can you tell me more about the type of advancements and risks these pose to society? What can we do to safeguard Australian society?
10. What do you think can be done to prevent AI-CSAM from a software standpoint e.g. what can the developers of Gen AI platforms do to prevent the creation of AI CSAM in particular the open-source environment.