ICMEC AUSTRALIA

SaferAI for Children Coalition

Artificial Intelligence and Child Protection:

A Collaborative Approach to a Safer Future

Discussion Paper

Led by ICMEC Australia, the **SaferAI for Children Coalition** is a collaborative body of Australian law enforcement, not-for-profit organisations, academics, AI experts, and other public sector stakeholders, all dedicated to creating better outcomes for children in the age of AI.

Table of Contents

Executive summary

Protecting children is one of the most crucial and urgent problems facing our world today. The 2023 Australian Childhood Maltreatment Study revealed a stark reality: **1 in 4 Australians have experienced sexual abuse as a child**. Such statistics are a glaring call to action – one child harmed is one too many.

With the advent of the internet, and now the growing prominence of Generative AI (GenAI), we have witnessed the rapid evolution of technology's role in society. The remarkable capabilities of GenAI, exemplified by large language models (LLMs) like ChatGPT and image generators like Stable Diffusion, have demonstrated potential for significant societal benefit. However, these technologies also present new avenues for misuse. We know that child sex offenders have already manipulated AI to scale their abuse, from generating child sexual abuse material (CSAM) to orchestrating grooming and sextortion.

The quick pace of this technology's growth has unfortunately meant that its been rolled-out without the necessary safety considerations in place. This has led to the rapid manipulation of GenAI tools to create child abuse material. The UK's Internet Watch Foundation found a singular dark web CSAM forum with 20,254 AI-generated images in a one-month period in 2023 - with 11,108 of these selected for assessment as they were most likely to be criminal. These are frightening statistics that continue to escalate as the technology becomes more capable and accessible.

This paper has been collaboratively written with insights from the members of the SaferAI for Children Coalition, a collective effort comprising child protection not-for-profit organisations, academic experts, and law enforcement agencies. From outlining the risks posed by AI-enabled tools in facilitating child sexual exploitation (CSE) to advocating for legislative reforms and innovative AI-driven solutions, this paper navigates the complex terrain of protecting children in the digital age. By synthesising research findings, case studies, and international best practices, it aims to inform policymakers, industry leaders, and the Australian public on the importance of harnessing AI responsibly for the protection and wellbeing of our most vulnerable - our children.

About the SaferAI for Children Coalition

Established in March 2024, the SaferAI for Children Coalition is a united front of child protection focused non-government organisations, academic experts, business, government, and law enforcement agencies, led by ICMEC Australia. We are committed to protecting children in a landscape increasingly dominated by advanced technologies. As we navigate through both the challenges and opportunities presented by AI, our mission is clear: to ensure that these innovations serve to protect rather than endanger our most vulnerable community members.

The SaferAI for Children Coalition embodies a diverse array of expertise, dedicated to fostering safer outcomes for children in this new digital age.

Our Members

alannah & madeline foundation

Australian Childhood Foundation

AiLECS lab

A21

Bravehearts
bravehearts.org.au

the Carly Ryan foundation.

CQUniversity AUSTRALIA

Daniel Morcombe Foundation Inc.
Keeping kids safe!

THE GRACE TAME FOUNDATION

ICMEC AUSTRALIA

IJM

National Centre for Action on Child Sexual Abuse

Rigr

unicef AUSTRALIA
for every child

UNIVERSITY of TASMANIA

Defining key terms

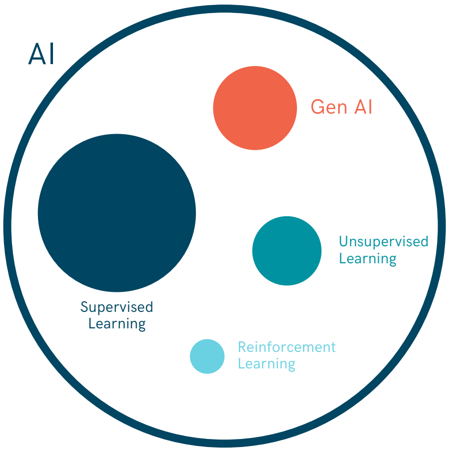
The complexities of emerging technologies are challenging to navigate, but crucial to understand, particularly as they rapidly integrate into various aspects of society. Understanding terms such as Artificial Intelligence, Generative AI, and Machine Learning is fundamental to addressing their implications. However, it is important to note that there are still no universally agreed-upon definitions for many of these terms, making it essential to establish clear and consistent meanings within the context of this paper. Similarly, defining child protection-related terms ensures a comprehensive and focused discussion on the intersection of technology and child safety.

Artificial Intelligence (AI)

AI involves computer systems with the capacity to execute tasks that would typically require human intelligence, such as problem-solving, learning patterns, speech recognition, and language understanding.

Generative Artificial Intelligence (GenAI)

GenAI is a branch of AI that focuses on creating new content, such as text, images, audio, and video, by learning from existing data.



(Andrew Ng, Stanford, 2023)

Machine Learning (ML)

ML is a subset of AI that involves the development of algorithms and statistical models that enable computers to perform tasks without explicit instructions. Instead, these systems learn from patterns and inferences drawn from data.

Child Sexual Abuse Material (CSAM)

CSAM is any representation, by whatever means, of a child engaged in real or simulated explicit sexual activities or any representation of the sexual parts of a child for primarily sexual purposes.

Online Child Sexual Exploitation (OCSE)

According to the Australian Centre to Counter Child Exploitation, Online Child Sexual Exploitation is "The use of technology or the internet to facilitate the sexual abuse of a child, including the production and sharing of child sexual abuse material online."

Self-Generated Child Sexual Abuse Material (SG-CSAM)

Thorn defines SG-CSAM as "explicit imagery of a child that appears to have been taken by the child in the image. It can result from consensual or coercive acts. Kids often refer to consensual experiences as 'sexting' or 'sharing nudes'."

A snapshot of the issue

Understanding the scope and scale of technology-facilitated child sexual exploitation and abuse is crucial to grasp the severity of this pervasive issue - and understand the complexities of how AI is changing this crime. While the focus of this paper is on the role of AI in child protection, it is important to provide an overview of the broader landscape. The following statistics highlight the alarming prevalence of child sexual abuse, and the impact of technology in facilitating these crimes. These figures not only underscore the urgent need for innovative solutions, but also set the stage for exploring how AI can be harnessed to combat this global crisis.

In Australia,
1 in 4
children are sexually abused.

(Australian Child Maltreatment Study, 2023)

More than 1 in 3 girls experience child sexual abuse in Australia.

(Australian Child Maltreatment Study, 2023)

Almost 1 in 5 boys experience child sexual abuse in Australia.

(Australian Child Maltreatment Study, 2023)

On average, it will take 24 years for child sexual abuse victims to disclose.

(Royal Commission into Institutional Responses to Child Sexual Abuse, 2017)

In 2023, NCMEC received almost
36.2 million reports of CSAM
an increase of 4.2 million from 2022

(NCMEC, 2023)

60%
rise in sextortion cases in Australia in the final months of 2022.

(Australian Centre to Counter Child Exploitation)

The Internet Watch Foundation found one dark-web forum:

- 20,254 AI-generated photos
- 12 analysts spent 87.5 hours assessing 11,108 of these images
- 2,975 were criminal
- 2,562 of these were realistic enough to be treated in the same way as CSAM
- The thousands of other material were non-criminal in nature, but contained children.

(Internet Watch Foundation UK, 2023)

Interpol's CSE database has 4.3 million unique images and videos

+ adding 15 victims on average every day

(Interpol, 2023)

Why is this issue important for Australia?

The Issue:

Child sex offenders manipulate any means available to aid their offending. With the invention and rollout of the internet, offenders began shifting their abuse and exploitation of children to leverage the capability of new technology such as chat applications and the dark web - leading to crimes like grooming, sextortion, distribution and exchange of CSAM, and live-streaming. Now, with emerging technologies like GenAI and end-to-end encryption, the ways this crime is facilitated is evolving again. Artificial Intelligence is being used to facilitate child sexual exploitation in many forms:

- **Increased CSAM Production**
  AI tools can be misused to create realistic child sexual abuse material, significantly increasing the volume and accessibility of illegal content. Not only can new abuse material be created, but real material can be manipulated to appear AI-generated in attempts to evade detection and frustrate law enforcement efforts to identify victims and remove them from harm. It is also being used to create new material of existing victims.
- **Deepfakes**
  Advances in AI enable the creation of hyper-realistic fake videos and images, which can be used to generate non-consensual explicit content involving children, compounding the problem of child sexual exploitation (CSE).
- **Sextortion**
  Sexual extortion, often referred to as sextortion, is impacted as offenders are using AI to impersonate minors or trusted adults online, manipulating victims into sharing explicit content or engaging in sexual acts, which are then used to extort or further abuse them. Perpetrators can exploit LLMs (large language models) to mimic natural human language. This capability allows offenders to groom children at scale in automated and more targeted ways.
- **‘Nudifying’ apps**
  Certain AI apps can digitally alter images to create nude or sexualised photos of children, posing serious privacy and safety risks. These apps also make it alarmingly easy for underage adolescents to commit image-based abuse against their peers.

AI-Generated CSAM is harmful and abusive.

The assumption that AI-Generated material is less harmful because it depicts 'fictional' children, or does not involve physical harm, is false. Such content perpetuates the exploitation and abuse of children, reinforcing harmful behaviours and attitudes among offenders. AI-Generated CSAM may contribute to the 'normalisation' of CSE, blurring the lines between real and synthetic material and complicating efforts to combat this crime. Additionally, AI can be used to take images of real children and make them sexualised, further exacerbating the harm and abuse of children.

In Australia, AI-Generated CSAM is illegal under Commonwealth law. Possessing, controlling, producing, accessing, distributing, obtaining, and supplying AI-generated CSAM is a criminal offence under the Commonwealth Criminal Code.

The Opportunity:

AI is immensely capable and transforms the way that we work, learn, and interact with information. The growth of Generative AI over the past 12 to 18 months has already transformed our world, and presents us with a unique opportunity to leverage positive use of the technology to protect children. Support and investment is critical to push forward the development of innovative solutions, and to ultimately save children from harm.

The opportunities are expansive, and possible initiatives to support include:

- Research and consequential development of AI-powered tools designed to detect and prevent OCSE.
- Leveraging AI to analyse CSAM faster, without human eyes in the first instance.
- Ongoing collaboration between tech companies, law enforcement, schools, and child protection organisations to develop and implement AI-driven protections.
- Training for law enforcement and child protection professionals on the latest available AI technologies (and their applications) in preventing and investigating CSE.
- Ethical deployment of AI tools for law enforcement and other agencies.
- Resourcing expert support for victim-survivors affected by these crimes to help them recover and heal.
- Deploying AI to detect and report known, new, or livestreamed CSAM (including within encrypted environments), or detect and block CSAM creation, storage and distribution at device level.

We are at a pivotal moment where we can address this issue proactively before it escalates uncontrollably - a rare and significant opportunity. Australia has the chance to become a global leader in this arena. The SaferAI for Children Coalition, with its extensive group knowledge, exemplifies the power of cross-sector collaboration in tackling these emerging challenges. By leveraging AI for child protection, Australia can continue to set a global standard and lead the way in protecting children in the digital age.

What actions are being taken internationally?

Addressing the global prevalence of online child sexual exploitation remains a significant international priority. As the world grapples with this issue, the intersection of responding to and legislating AI has emerged as a crucial consideration in child protection. Learning from international initiatives is paramount for Australia as we explore various approaches.

European Union (EU) AI Act

The EU's AI Act represents a milestone as the world's first specific legal framework for AI. Following its formal adoption in May 2024, the European Council is set to publish the Act in the Official Journal of the European Union by mid-2024. This landmark legislation introduced stringent regulations aimed at ensuring the safe and ethical deployment of AI technologies, with a dedicated focus on protecting children.

The Act is noteworthy for its approach to balancing innovation with the protection of fundamental rights. By addressing concerns around AI's potential risks, especially in relation to vulnerable groups such as children, the EU AI Act sets a global precedent in AI governance. It aims to foster a more secure and trustworthy AI environment, promoting accountability and transparency in the development and deployment of AI technologies.

The Act mandates requirements for high-risk AI applications, whilst also emphasising the importance of human oversight and accountability mechanisms. By establishing clear guidelines for AI developers, the EU seeks to mitigate potential harms and promote the responsible use of AI technologies across its member states.

reThink Chatbot

The reThink Chabot is a groundbreaking trial program aimed at combatting CSAM on pornography websites, and was recently conducted on Pornhub UK's platform in collaboration with UK-based child protection organisations and the University of Tasmania. The Chatbot directed users to vital support services provided by Stop it Now! (a program provided by the Lucy Faithfull Foundation in the UK), and partnered with Aylo (formerly MindGeek) and the Internet Watch Foundation. Through the display of warning messages and the provision of support via a chatbot prompted by child abuse related searches, this initiative aimed to deter individuals from accessing CSAM and seek help instead. This is a strong example of leveraging technology for enhanced online child protection efforts.

'Safer' by Thorn

Safer uses advanced AI and machine learning to help content-hosting platforms detect CSAM on their platforms. The tool leverages proprietary research, a vast database of known CSAM hashes, and issue expertise to protect content-hosting platforms. Safer's tools enable accurate detection and removal of CSAM, enhancing the safety of the online environment.

TakeItDown Act (United States)

The newly proposed TakeItDown Act represents a significant step towards addressing the pressing issue of non-consensual intimate images (NCII), which includes media commonly known as deepfakes, within the United States. While still under development, this legislation aims to provide crucial protections for victims while upholding the principles of free speech.

The Act proposes several key measures to protect victims of 'real' and 'deepfake' NCII, by:

- Criminalising the knowing publication of NCII on social media and online platforms.
- Clarifying that a victim's consent to the creation of an image does not mean consent to its publication.
- Protecting 'good faith' efforts to help victims
- Requiring online platforms to take down NCII when made aware by the victim within 48 hours.
- Protecting lawful speech by stipulating computer-generated NCII realistically depicts an individual.

This Act represents a critical legislative effort to combat the growing threat of NCII, providing protections for victim-survivors and providing clear responsibilities for online platforms. If adopted, this Act could have a significant impact on victim-survivor rights.

UNICEF's Policy Guidance on AI for Children

The UNICEF Policy Guidance on AI for Children aims to ensure that AI systems are developed and deployed in ways that respect and uphold children's rights. The guidance acknowledges the transformative impact of AI on children's lives both directly and indirectly. UNICEF emphasises the dual nature of AI's potential: it can support children's development, but also poses significant risks to their privacy, safety, and overall wellbeing.

The guidance outlines nine key requirements for child-centred AI, and highlights the importance of inclusive consultation processes to lead to impactful outcomes, aiming to influence global AI policies and strategies to be more inclusive and protective of children's rights.

UNICRI AI for Safer Children Hub

The United Nations Interregional Crime and Justice Research Institute (UNICRI) AI for Safer Children initiative, launched in 2020 by the UNICRI Centre for AI and Robotics in collaboration with the Ministry of Interior of the United Arab Emirates, aims to combat online child sexual exploitation and abuse using AI. This initiative focuses on supporting law enforcement agencies by helping them explore and utilise AI tools in investigations.

The AI for Safer Children Global Hub is a secure platform designed specifically for law enforcement. The hub provides information on AI tools that can be used to tackle CSE, and guides ethical usage of these tools. The initiative aligns with Target 2 of Goal 16 of the 2030 Agenda for Sustainable Development, which aims to end abuse, exploitation, trafficking, and all forms of violence against children.

World Childhood Foundation

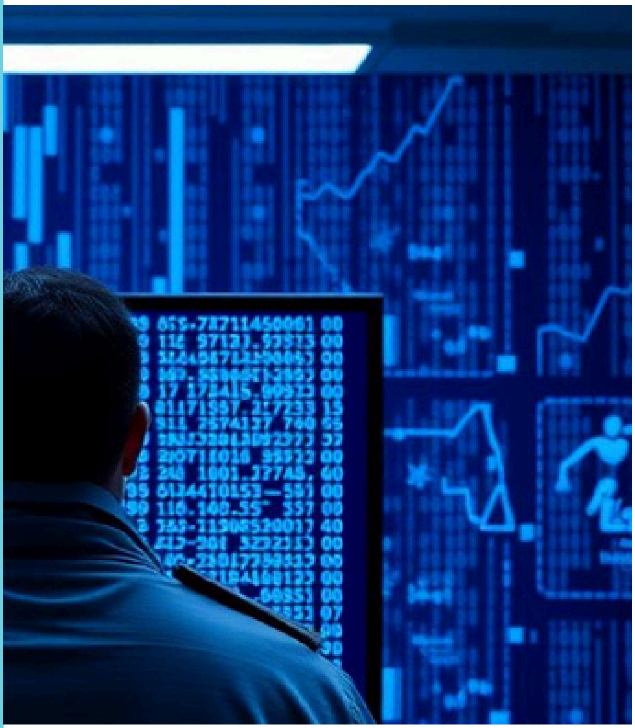
The World Childhood Foundation, based in Sweden, has been a pioneering not-for-profit organisation at the forefront of working at the intersection of AI with child protection on a global scale. Their *Stella Polaris* initiative, launched in 2021, stands out as a collaborative effort aimed at harnessing AI technologies to enhance child protection measures. These initiatives are designed to mitigate risks and safeguard children from online exploitation and other forms of harm. This approach underscores the importance of leveraging existing AI tools, and fostering the creation of new technologies tailored to address specific challenges in child protection.

The current Australian landscape

Australia has recognised the urgent need to address the challenges posed by OCSE, and the intersection of AI with child protection. Positive steps are being taken within our country's borders that indicate a proactive approach to these critical issues. Across Australia, from government to not-for-profits, there are a range of initiatives and legislation aimed at enhancing online safety for children, and leveraging AI to protect them. These efforts reflect a commitment to protecting children from the evolving threats in the digital landscape.

In Australia, we're world-renowned in our child protection response efforts. This is in large part due to the dedication and expertise of our law enforcement partners who work tirelessly to combat this crime - bringing offenders to justice and saving children from harm.

However, from an AI perspective, our national capabilities are still developing. Australians also rank among the most cautious globally regarding AI's role, making the intersection of AI and child protection relatively new and in need of further support and exploration.



ANZPAA's Australia New Zealand Police Artificial Intelligence Principles

The Australia New Zealand Policing Advisory Agency, ANZPAA, has developed these principles to guide the ethical and responsible use of AI by law enforcement agencies across Australia and New Zealand - a critical step towards creating synergy across jurisdictions. These principles are an important framework for law enforcement to refer to when adopting new AI technologies. Although these are not specific to the child protection response ecosystem, the principles provide an apt baseline for sector-specific frameworks to be created - and ensure they align with existing frameworks to avoid duplication.

Department of Industry and Science (DISR)

DISR, through its oversight of the National AI Safety Standards, plays an essential role in guiding AI practices toward ethical and safe outcomes, especially for applications impacting children. By expanding these standards to address child protection specifically, DISR has the opportunity to establish clear guidelines for AI tools used in monitoring, identifying, and mitigating online risks to children. With such standards in place, AI technologies in child protection can operate responsibly and effectively, supporting a national framework that prioritises safety and the ethical use of AI in sensitive areas.

National AI Centre (NAIC)

Now part of DISR, the NAIC serves as a collaborative hub for advancing AI solutions that address societal needs, including child protection. Through partnerships across government, industry, and academia, the NAIC can promote the development of AI tools capable of identifying harmful content, detecting grooming behaviors, and preventing exploitation. By prioritizing projects focused on safeguarding children, the NAIC could play a critical role in fostering ethical AI applications that serve the public good, positioning Australia as a leader in AI innovation with strong protective standards for vulnerable populations.

Office of the Australian Information Commissioner (OAIC)

The OAIC's October 2024 guidelines on privacy and the use of commercially available AI products reinforce the importance of transparency, accountability, and data protection, particularly in AI applications involving children's data. By setting and promoting these privacy standards, the OAIC enables AI technologies in child protection to operate with strict data safeguards, fostering public trust. These guidelines provide a valuable framework for ensuring that AI systems respect children's privacy rights and can be safely integrated into protective initiatives, reinforcing AI's role as a reliable, ethical asset in child protection efforts.

My Pictures Matter – AiLECS Lab

The 'My Pictures Matter' project is a collaborative effort involving the Australian Federal Police and Monash University's AiLECS Lab. The project's primary objective is to develop an ethical AI tool designed to detect child sexual abuse material in videos or photos shared on the dark web, or seized during criminal investigations. This tool aims to significantly reduce the manual work required by investigators, enabling quicker identification of potential CSAM. To effectively train the model, the initiative calls for adult Australians to contribute non-harmful childhood photos of *themselves*, targeting a collection of around 100,000 images from all ethnicities.

The eSafety Commissioner's Holistic Approach: Prevention, Protection, and Proactive and Systemic Change

The eSafety Commissioner (eSafety) is Australia's independent regulator and educator for online safety. eSafety's functions are governed by the Online Safety Act 2021, which came into effect in January 2022 and has been independently reviewed in 2024. As Australia's leader in online safety, eSafety provides information, guidance and advice to the whole Australian community and a broad range of stakeholders within the online safety ecosystem, including government, NGOs, and industry and subject matter experts.

In Australia, the Basic Online Safety Expectations (BOSE) - established by the Minister for Communications through a legislative determination - and industry codes and standards - designed to protect Australians from illegal and restricted online content - are key mechanisms for industry to address the evolving online safety challenges posed by AI technologies in child protection. Recent updates related to generative AI include the amendment of the BOSE Determination to include an explicit expectation related to generative AI and the registration of the Designated Internet Services (DIS) industry standard which includes specific obligations for certain generative AI services.

Through the BOSE and the Online Safety Act, the eSafety Commissioner can require online service providers to report on how they are meeting the expectations, encouraging transparency and accountability in safety efforts. The amendments in early 2024 placed new expectations on service providers, particularly the additional considerations of generative AI tools to be safe in their design and operation. The updates also specify that there needs to be proactive minimisation of generative AI being used to facilitate deepfake images and other illicit, harmful material.

The DIS Industry Standard will commence in December 2024, regulates specific generative AI services in a targeted manner. For example, it includes a category called High Impact Generative AI DIS, defined as those which use machine learning models to enable end-users to produce material, where the service has not incorporated sufficient controls to reduce the risk of generating synthetic high impact material such as child sexual exploitation material. In terms of detection and identification, High Impact Generative AI DIS may be required to assess whether inputs into the service contain known (previously verified) CSAM. From 22 December 2024, Australian residents will be able to make complaints to eSafety where they consider an online service is not complying with a code or standard under the Online Safety Act which, as noted above, covers generative AI services. If known CSAM is identified, providers must remove it from the service as soon as practicable. This hinges on technical feasibility from the provider.

These updates aim to create a safer online environment for children while holding service providers accountable for a number of risks associated with generative AI. These regulatory updates demonstrate the importance of a proactive stance to mitigate the risks of generative AI being misused or exploited to create abusive and exploitative material. It also reinforces Australia's commitment to protecting children in the digital age. The ongoing evolution of these regulatory developments reflect the dynamic nature of online safety challenges, emphasising the need for continuous adaptation and vigilance in the face of technological advancements.

To read more about eSafety's multidimensional approach and work as anticipatory regulator, please see Appendix A.

What are Australian law enforcement seeing?

The rise in CSAM reports has created an urgent need for enhanced support to law enforcement agencies (LEAs). AI is now a pivotal tool not only in addressing the challenges of traditional CSAM investigations but also in combatting the emerging complexities of AI-generated CSAM.

The ability of AI to automatically generate, manipulate, and distribute abuse material means that LEAs are handling more cases that are both larger in volume and more technologically sophisticated. This surge demands resources and tools that can alleviate the strain on investigators, enabling them to focus on high-priority cases and child rescue operations.

How can AI help?

- Automate the review and categorisation of CSAM
- Prioritise cases based on urgency and threat levels
- Identify patterns in grooming tactics and offender behavior

> **‘Under the influence’**
> ABC, Friday 16 August 2024
>
> The use of AI tech to create or edit child abuse material remains relatively rare, according to Detective Superintendent Frank Rayner from the Australian Centre To Counter Child Exploitation.
>
> Roughly 40,000 online child exploitation reports are received by the ACCCE per year.
>
> To date this year, fewer than 20 of those were AI-generated.
>
> Despite this, Superintendent Rayner said **the material was out there, and its prevalence was growing.**”

Our Coalition is dedicated to supporting our law enforcement partners by advocating for the development, funding, and deployment of advanced AI solutions. These include image and video analysis tools that automate the identification and categorisation of CSAM, which reduces the need for manual review by human officers. Such tools not only speed up processing but also protect investigators from trauma.

Where to next for Australia?

The Coalition recognises the significant role that collaboration plays in creating a safer world for our children, which is why we are all dedicated to working together at this intersection of child protection and AI. Engaging a variety of leaders and decision makers in this space is imperative to our work. We're reaching across sectors and borders to emphasise the central role that child protection needs to have in various aspects of Australian society.

Raising community awareness

Raising awareness about the potential dangers of AI in relation to child safety and educating the Australian public and professionals about preventive measures should be a cornerstone of our joint efforts. This awareness must be twofold - exploring both the risks and opportunities presented by AI. This technology has transformed our world. Australians need the tools to understand how AI works, and realistically understand both its threats and opportunities. Being empowered with information will help our world best tackle this new technology, and ensure it's used for good going forward.

Legislative change

Comprehensive legislation and regulatory frameworks are key to implementing an effective, whole-of-system approach to child safety. AI has certainly posed a barrier to these efforts, necessitating amendments to legislation and government considerations to navigate a comprehensive response. The steps taken already by the eSafety Commissioner in this regard are commendable and set a strong foundation for future actions in other areas of the Australian government.

Investing in innovative solutions

Innovation has driven significant impact in the fight against CSE. Leveraging AI is critical to enhancing our ability to protect children from online harms. By harnessing the power of technology, we can develop advanced tools for both protection and prevention. AI can be utilised to identify and remove harmful content swiftly, predict and prevent potential abuses, and support law enforcement in apprehending offenders. Investing in research and the development of new technological solutions can lead to more effective methods to protect children online. Continued support of innovation will ensure we stay ahead of emerging threats and continue to improve our protective measures.

Consultation and co-design with victim-survivors

Incorporating the voices of victim-survivors in the development of policies and technologies is vital to ensuring that solutions are effective and sensitive to the needs of those affected. Australia should prioritise appropriate consultation and co-design with victim-survivors, enabling their lived experiences to guide and shape the measures put in place. Integrating the survivor perspective can significantly enhance the relevance and impact of initiatives aimed at preventing and responding to child sexual abuse.

Supporting critical research and implementing findings

Understanding this complex and evolving issue requires up-to-date research. Supporting research initiatives that focus on developing AI tools specifically designed to detect and prevent child exploitation is essential. Encouraging interdisciplinary studies that combine insights from technology, psychology and criminology will help better understand and counteract offenders' methods. Promoting partnerships between academic institutions, government agencies, and not-for-profits ensures that research findings are translated into practical tools and policies. Additionally, encouraging research initiatives that identify best practices in supporting children and families affected by these crimes to recover and heal is crucial, as is promoting these insights to government and relevant services.

Engaging a variety of sectors and perspectives

Expanding collaboration across sectors is essential to effectively address AI and child protection. Beyond traditional focus areas like law enforcement and technology, engaging education and psychology sectors is critical. Schools can equip young people with digital resilience, while psychologists offer insights into AI's behavioral impacts on children, informing tailored interventions. Partnering with these sectors enables a holistic response, combining preventative education and psychological expertise to strengthen child protection in an evolving AI landscape.

Our commitment as a Coalition

The SaferAI for Children Coalition suggests the following considerations to ensure that child protection remains effectively intertwined with the adoption of AI in Australia. As a cross-industry group, we are dedicated to leveraging our expertise and networks to create better outcomes for children.

Developing community support resources

The Coalition advocates for the creation of educational materials and resources to help parents/guardians, carers, educators, and children understand the risks and benefits of AI. Establishing support networks and service responses for victims of online exploitation is crucial, providing them with the necessary resources and assistance to feel supported.

Social media awareness

Launching awareness campaigns to educate the public about the dangers of AI in the context of child exploitation, and the measures that can be taken to mitigate these risks, is vital. Utilising social media platforms to disseminate information about the latest threats and safety tips related to AI technologies is also recommended.

Cross-sector collaboration

Fostering partnerships across government agencies, tech companies, and child protection organisations is essential for sharing knowledge and resources. Organising conferences and workshops to facilitate the exchange of best practices and innovative solutions for protecting children from AI-related threats is encouraged. Additionally, advocating for international cooperation to address the global nature of OCSE in order to harmonise legal frameworks and enforcement efforts is important.

Legislative and Policy Change

Legislative and policy change is a crucial aspect to protecting children from the potential harms of AI. We will advocate for the development and implementation of strong, comprehensive laws and policies that address the intersection of AI and child protection. By influencing policy discussions, providing expert recommendations, and participating in consultations, we aim to ensure that child protection is a priority in AI regulation - particularly around applications targeted at children or known to be used by children and young people. Additionally, we will monitor the perceived effectiveness of existing laws, identify gaps, and promote international cooperation to harmonise legal frameworks and enforcement mechanisms across borders.

Future direction

By implementing these recommendations and taking decisive action, the SaferAI for Children Coalition aims to ensure that AI technologies are harnessed to protect, rather than endanger, our children. The Coalition is committed to working with all relevant stakeholders to create a safer digital environment for future generations. The significant and specific impact of AI on child safety underscores the critical need for specialised consultation in this arena.

We are eager to discuss these proposals in more detail and explore the ways in which the SaferAI for Children Coalition can support government in turning these ideas into action. By aligning our resources and expertise, we can ensure that AI development in Australia keeps the safety of our children a paramount concern. Through collaborative efforts, innovative thinking, and unwavering commitment, we can build a future where technology serves as a shield for our children, safeguarding their innocence and well-being in an increasingly digital world.

The Coalition hopes to implement actionable items from these recommendations to increase awareness and improve outcomes for Australian children in the increasingly AI-driven age. We are dedicated to developing solutions, creating comprehensive roadmaps, and collaborating with stakeholders to ensure the protection and wellbeing of our children.

Together, we can forge a path towards a safer, more secure environment for our youngest and most vulnerable, ensuring that every child is protected from the risks posed by emerging technologies. The SaferAI for Children Coalition stands ready to lead this crucial endeavour, fostering a society where AI is a force for good, dedicated to the protection of our children.

Appendix A

As part of its work as an anticipatory regulator, eSafety has a Tech Trends workstream, and conducts horizon scanning and works with subject matter experts. This allows eSafety to identify the online safety risks and benefits of emerging technologies, as well as the regulatory challenges and opportunities they may present. In August 2023, eSafety published a position statement on generative AI, which provides an overview of the generative AI lifecycle, examples of its use and misuse, consideration of online safety risks and opportunities, as well as regulatory challenges and approaches including an explanation of how the Online Safety Act 2021 applies. It also provides specific Safety by Design interventions that industry can adopt immediately and other approaches to improve user safety.

eSafety's work to prevent AI-related harm through education and awareness raising includes updates to its professional learning program, which now includes a webinar about online safety considerations for generative AI in education. eSafety also actively promotes AI related issues within our community-based work.

eSafety's regulatory schemes cover both real and synthetic child sexual abuse material, deepfake image-based abuse, AI enabled-content used to target Australian children through cyberbullying and adult cyber abuse. eSafety is starting to receive reports from the public about AI-driven abuse and has taken regulatory action against an individual for creating and posting deepfake intimate images of Australian women without their consent.

References

ABC News, Artificial intelligence is being used to create child abuse material, and police are worried about its spread, https://www.abc.net.au/news/2024-04-18/artificial-intelligence-child-exploitation-material/103734216

AFP, *Tasmanian jailed for possessing AI-generated child abuse material*, https://www.afp.gov.au/news-centre/media-release/tasmanian-jailed-possessing-ai-generated-child-abuse-material

Australian Centre to Counter Child Exploitation, *What is online child sexual exploitation?,* https://www.accce.gov.au/help-and-support/what-is-online-child-exploitation

Australian Government, *Australia's AI Action Plan*, https://wp.oecd.ai/app/uploads/2021/12/Australia_AI_Action_Plan_2021.pdf

End Violence Against Children, *combatting Online Child Sexual Violence with the help of AI,* https://www.end-violence.org/articles/combatting-online-child-sexual-violence-help-ai

Human-Centred Artificial Intelligence, *Artificial Intelligence Definitions*, https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

ICMEC Australia, *Keeping Children Safe in the Age of AI: Examining the risks and benefits of AI implementation for CSE*, https://icmec.org.au/keeping-children-safe-in-the-age-of-ai/

INHOPE, *What is generative AI?*, https://www.inhope.org/EN/articles/what-is-generative-ai

International Centre for Missing and Exploited Children, *Terminology*, https://www.icmec.org/resources/terminology/

Internet Watch Foundation, *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/

Lucy Faithful Foundation, *A call to end AI-generated child sexual abuse*, https://www.lucyfaithfull.org.uk/featured-news/a-call-to-end-ai-generated-child-sexual-abuse.htm

Reuters, *US receives thousands of reports of AI-generated child abuse content in growing risk*, https://www.reuters.com/world/us/us-receives-thousands-reports-ai-generated-child-abuse-content-growing-risk-2024-01-31/

New York Times, *A.I.-Generated Child Sexual Abuse Material May Overwhelm Tip Line,* https://www.nytimes.com/2024/04/22/technology/ai-csam-cybertipline.html

Scanlan, Joel; Prichard, Jeremy; Hall, Lauren; Watters, Paul; Wortley, Richard (2024). *reThink Chatbot Evaluation*. University of Tasmania. Report. https://hdl.handle.net/102.100.100/608878

Stanford Cyber Policy Center, *Investigation Finds AI Image Generation Models Trained on Child Abuse*, https://cyber.fsi.stanford.edu/news/investigation-finds-ai-image-generation-models-trained-child-abuse

Singh S & Nambiar V, *Role of Artificial Intelligence in the Prevention of Online Child Sexual Abuse: A Systematic Review of Literature,* https://www.tandfonline.com/doi/full/10.1080/19361610.2024.2331885?scroll=top&needAccess=true

Thiel D, Stroebel M, and Rebecca Portnoff, *Generative ML and CSAM: Implications and Mitigations*, https://doi.org/10.25740/jv206yg3793

Thorn, *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, *Thorn and All Tech Is Human Forge Generative AI Principles with AI Leaders to Enact Strong Child Safety Commitments,* https://www.thorn.org/blog/generative-ai-principles/

Thorn*, UNDERSTANDING AND MITIGATING RISKS TO YOUR PLATFORM: Emerging Online Trends in Child Sexual Abuse,* p4. Accessed 14 March 2024 https://get.safer.io/hubfs/2023%20Emerging%20Trends%20Report/23_Safer_EmergingTrends_FNL.pdf

Time, *As Tech CEOs Are Grilled Over Child Safety Online, AI Is Complicating the Issue*, https://time.com/6590470/csam-ai-tech-ceos/

UK Government, *Joint Statement: Tackling child sexual abuse in the age of Artificial Intelligence*, https://www.gov.uk/government/publications/tackling-child-sexual-abuse-in-the-age-of-artificial-intelligence/joint-statement-tackling-child-sexual-abuse-in-the-age-of-artificial-intelligence

University of Adelaide, *Artificial Intelligence*, https://libguides.adelaide.edu.au/artificial_intel

US Department of Homeland Security, *Artificial Intelligence and Combatting Online Child Sexual Exploitation and Abuse*, https://www.dhs.gov/sites/default/files/2024-04/24_0408_k2p_genai-bulletin.pdf

U.S. Senate Committee on Commerce, Science, and Transportation, *Sen. Cruz Leads Colleagues in Unveiling Landmark Bill to Protect Victims of Deepfake Revenge Porn*, https://www.commerce.senate.gov/2024/6/sen-cruz-leads-colleagues-in-unveiling-landmark-bill-to-protect-victims-of-deepfake-revenge-porn

ICMEC
AUSTRALIA

SaferAI for Children Coalition

About ICMEC Australia

ICMEC Australia is a not-for-profit organisation with a clear mission: to support and strengthen the professionals who detect, report, prosecute and prevent online child sexual exploitation. We collaborate with various stakeholders, such as financial services and corporate entities, law enforcement, policymakers, academics, and NGOs, to develop strategies to protect children from harm. ICMEC Australia turns the use of online technology to exploit children on its head. We support frontline workers in combatting child sexual abuse by leveraging technology and data-driven strategies to detect, report, prosecute, and prevent child exploitation.

www.icmec.org.au   prevention@icmec.org.au   ICMEC Australia