Algorithmic Kids

Towards child-centred AI in Australia

unicef AUSTRALIA | for every child

A W O

This report has been developed by AWO for UNICEF Australia. UNICEF Australia thanks AWO for its research, analysis and drafting of this report. The conclusions and policy recommendations are UNICEF Australia's.

A W O

AWO is a consultancy that empowers individuals and organisations to uphold data rights, comply with the law and effect change in data protection and digital policy. AWO has offices in Sydney, London and Brussels and is also a registered law firm in the UK.

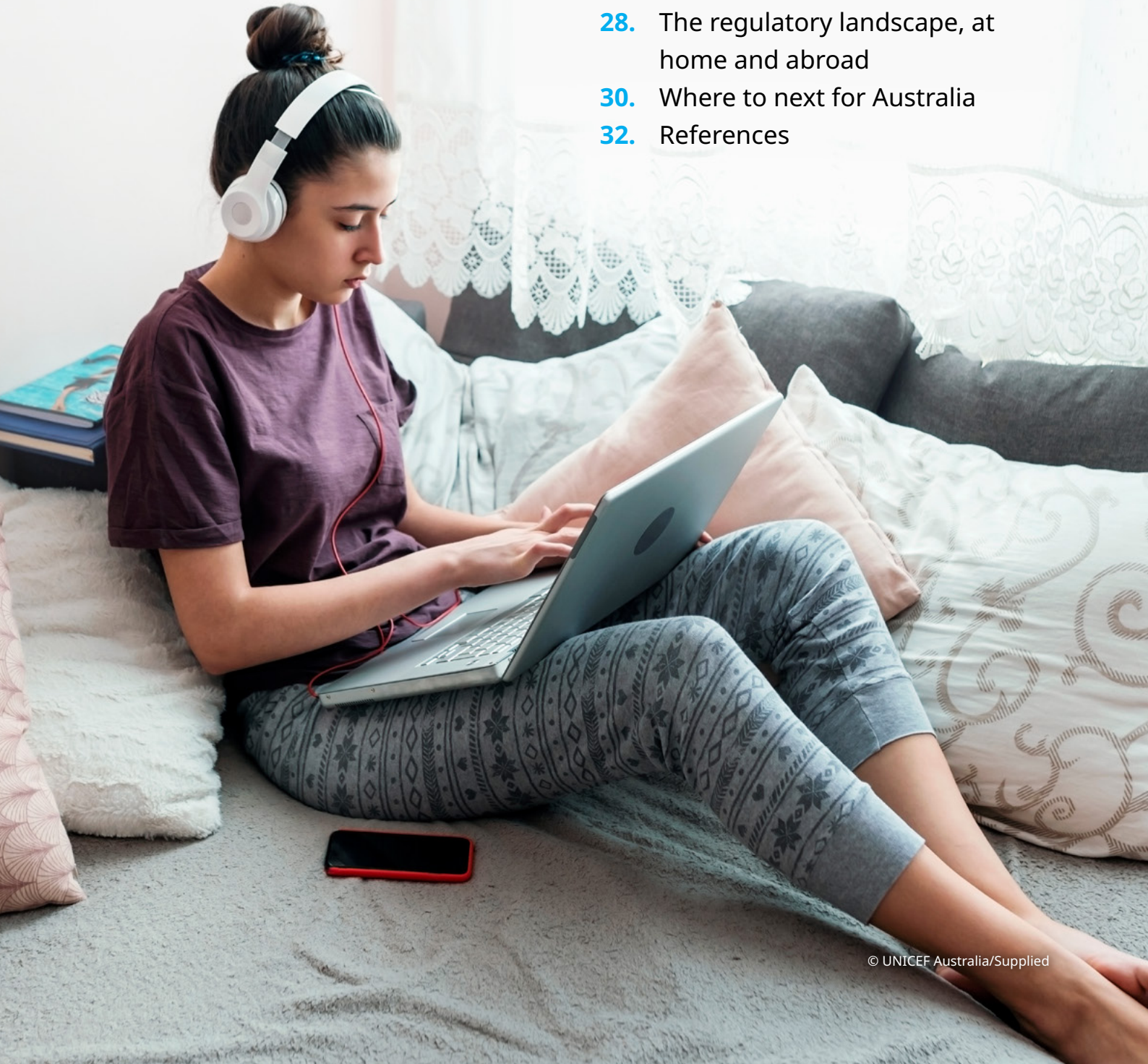
unicef
AUSTRALIA
for every child

UNICEF is the world's leading organisation working to protect and improve the lives of every child in over 190 countries. UNICEF is working around the world to deliver cutting-edge research and programming to improve digital environments and keep children and young people safe online. In Australia, we support this global ambition by working with local partners and governments to raise children's voices, defend their rights, and help them reach their potential at all stages of life.

UNICEF Australia acknowledges the Traditional Custodians of the lands and communities in which we and our partners work throughout Australia, and recognises their connection to their lands, waters, and communities. We pay respect to Aboriginal and Torres Strait Islander peoples and cultures, and to Elders past and present, and to the children of today who are the Elders of the future.

Suggested citation: UNICEF Australia (2025), Algorithmic Kids: Towards child-centred AI in Australia

Contents



Executive Summary

Artificial intelligence (AI) is rapidly changing childhood in Australia. From interacting with AI chatbots that are almost indistinguishable from a person, to generating new ideas, images and videos seemingly out of thin air, children in Australia are already living AI-augmented lives.

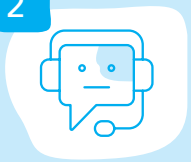
We know that children are early adopters of new technologies – 69% of Aussie teens have used a generative AI chatbot like ChatGPT.[i] Given their rapid adoption of new tech and often unique vulnerability to its effects, children also provide an early indication of where the latest digital frontier might take us. This is why UNICEF Australia has undertaken a rapid review of some of the most pressing impacts of AI on Australian children. We found that AI is an amplifier of both risk and opportunity across a few key areas:

1

Deepfakes, nudification and image-based abuse

Children are using GenAI to alter or create illicit images and videos, often of their peers. This image-based abuse is hurting children from the schoolyard to the home.

2



AI companions

AI companions and chatbots designed for adults are being used by children. Often romantic and sexualised in nature, a lack of safeguards is putting developmentally vulnerable children at risk.

3

Child exploitation material & AI

GenAI tools are being used to alter or create illegal child exploitation and abuse material. Ease of use is creating an explosion of illegal content, making an already complex problem worse.

4

AI-powered online safety

AI itself can be used to turn the tables on the new risks it's creating. AI-powered tools are helping tech platforms and law enforcement respond to new harms and offending more effectively.

5

Education & AI

AI can create personalised learning for students, help them test ideas and arguments, and provide answers to complex questions. For teachers, it can reduce their administrative burden, freeing them up to focus on what matters.

As the global AI race gathers speed, Australia has an opportunity now to put our unique stamp on this transformational new frontier. UNICEF Australia wants our country to be the international champion of child-centred AI – a global hub for innovation in safe and responsible AI for kids, mitigating its risks and maximising its opportunities. We offer some initial recommendations to put Australia on a pathway towards this vision

Recommendations

1. **Utilise existing digital reforms in Australia to make them fit-for-purpose for AI, then plug any gaps with new laws and regulations** - Better protection for children's data through the Children's Online Privacy Code, and a new duty of care in the Online Safety Act to ensure digital platforms are safe, will go a long way to managing AI risks for children. Where risks are left unaddressed, new laws should be adopted.

2. **Conduct a scoping study into age-gating AI companions, chatbots and nudify apps** - The findings of the Age Assurance Technology Trial could inform a scoping study on whether AI companions, chatbots and nudify apps should be age-gated in Australia, and how it would work in practice.

3. **Fund research to better understand the risks posed by AI, particularly for severe issues like AI-generated child exploitation material** - New evidence on severe AI harms will help Australia better respond to them. Funding research that tackles this, like the Disrupting Harm project for online child sexual exploitation, can help Australia stay ahead of the problem.

4. **Invest in AI-safety technologies and tools for Australian law enforcement and civil society** - We can use AI to turn the tables on some of the new risks it has introduced. Funding for new tech, like AI tools that help with the identification of child sexual abuse material, can help law enforcement and civil society be more effective and efficient.

5. **Establish an Australian AI Safety Institute, with a dedicated workstream on children** - An AI Safety Institute could lead technical research including into the impacts of AI on children, acting as a bridge between government, industry, and the public, and be a vehicle for sharing Australian innovation with the world.

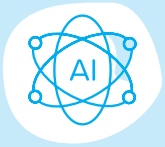
6. **Integrate AI literacy into the national curriculum and raise public awareness** - We need to equip Australia's next generation with the skills to both use AI safely and capitalise on the opportunities it provides. Raising public awareness will also help parents have the confidence to safely guide their children through an increasingly AI-augmented world.

Introduction

The rapid proliferation of artificial intelligence (AI) and particularly generative AI (GenAI) since the public release of tools like ChatGPT in late 2022 has reshaped digital experiences for children. Whether it is AI agents, image generators, facial recognition, smart toys, integration into health services, educational platforms, entertainment apps or more, GenAI is increasingly embedded in the everyday digital environments that Australian children navigate.

As UNICEF has explored in previous reports, the interaction between AI systems and children is complex and not only limited to those systems designed for and used by children[ii]. In many cases, even when AI systems are not specifically meant for children, children are interacting with them. In other situations, AI systems that are not used by children may still affect them in direct or indirect ways[iii].

A snapshot of data on AI and Aussie kids



69%

of teens have used a gen-AI chatbot like ChatGPT[a]

45%

of teens have used a gen-AI chatbot to generate ideas for schoolwork, the most common use[b]

47%

of teens aged 13-16 find it easy to use AI to help with study or work, rising to 67% for 17-18 year olds[c]

27%

of high school students think using AI for school work is cheating[d]

73%

of young people are aware that AI systems can be biased, discriminate, or provide false info[e]

56%

of teens have been exposed to deepfake content[f]

[a b f] UNICEF Australia, Protecting children in the online world: Reshaping the digital world for Aussie kids, 2024
[c e] Tesltra Foundation, 'Australian Youth Digital Index' 2024
[d] Cyber Safety Project 'Teens & Screens in 2024: The Digital Habits of 5145 Secondary Students from 22 Schools Across Australia', 2024

The current generation of children are born into a digital world with a unique ability to adapt to new technologies that often surpasses adults[iv]. AI offers a wide range of opportunities for children and young people. One example is in the field of education and learning where tailored learning could offer innovative ways for children and young people to interact with and contribute to digital environments. AI can also empower children to express their creativity in exciting new ways—enabling them to create stories, artwork, games, and other digital content without needing advanced coding skills. However, with this comes a range of risks for children and young people that must be considered and responded to. These risks are likely to have significant impacts on children and young people that we are still trying to grapple with and begin to understand.

UNICEF's global research tells us that Australia is undoubtedly a world-leader when it comes to many aspects of providing a safe world online for children[v] – it is important that Australia takes a proportionate and considered approach to GenAI and its impact on children. In doing so, it can also be helpful to consider the ways other countries are approaching these issues. This report briefly touches on this.

As GenAI continues to evolve, so too must our understanding of its impact on children. This calls for a collaborative approach—engaging educators, policymakers, technologists, and most importantly, children themselves—to ensure that AI is harnessed responsibly and inclusively in shaping the digital futures of Australia's youngest citizens.

This report focuses on certain key risks and opportunities that arise from AI (particularly GenAI) for children in Australia. Although there are undoubtedly other issues not covered here, this rapid review aims to provide an overview of some of the most pressing areas. The reality is that AI can present both opportunities and risks and that different children may be impacted in varying ways. This report covers:

AI risks and opportunities for Australian children

Deepfakes, nudification and image-based abuse

AI companions

Child exploitation material & AI

AI-powered online safety

Education & AI

Deepfakes, nudification and image-based abuse

One of the new risks raising major concerns for the safety and wellbeing of Australian children introduced by the rapid adoption of GenAI tools is related to the use of GenAI to manipulate or fabricate visual content – commonly referred to as deepfakes. These technologies, while not inherently malicious, are increasingly being used in ways that compromise children's privacy, dignity, and mental health. This is the case, for example, of a very specific type of tool used to generate deepfakes: nudification tools.

A nudification tool is a type of GenAI application – often in the form of an app or website – that uses AI to create sexually explicit deepfake images of real people, typically by digitally removing clothing from a source image. These tools are designed to distort reality and simulate nudity, often without the subject's knowledge or consent. While marketed as entertainment or novelty, they are increasingly being used to abuse, exploit, and humiliate children and young people.

The eSafety Commissioner has issued an urgent call for schools to report deepfake incidents to appropriate authorities as the rapid proliferation of 'nudify' apps online takes a growing toll on communities around Australia.[vi] eSafety has also issued a new Online Safety Advisory to alert parents and schools to the recent proliferation of open-source AI 'nudify' apps that are easily accessible by anyone with a smartphone.[vii] Reports to eSafety by people under 18 about digitally altered intimate images are growing rapidly. New data reveals these reports have more than doubled in the past 18 months. eSafety has noted that it is a "current crisis affecting school communities across Australia" and that they are having devastating impacts on young people in Australia.[viii] The Commissioner has issued a Toolkit for Schools which includes a guide for dealing with deepfake incidents directed at both parents, carers and schools.[ix]

Nudify Apps in Australia

In early 2025, it was reported that two students were suspended after fake sexually explicit images of up to 60 students from a school in Melbourne's north-west were circulated online. Victoria Police said an unknown number of school formal photos of female students at Gladstone Park Secondary College had been "inappropriately manipulated using artificial intelligence" before being released on the internet.

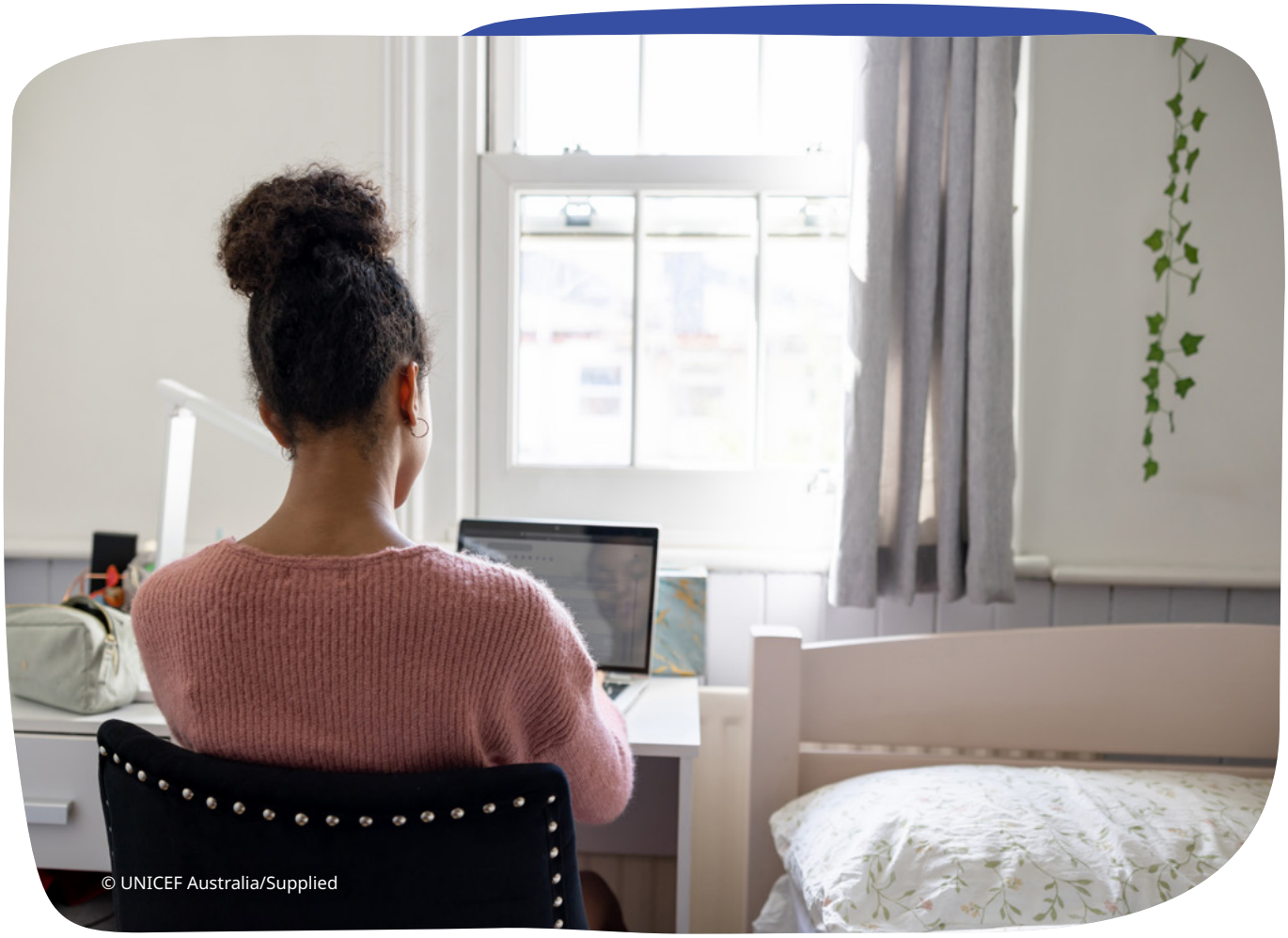
In 2024 a teenage schoolboy was arrested for allegedly distributing "incredibly graphic" deepfake images of 50 girls from Bacchus Marsh Grammar in Victoria. According to reports, the nude images were created by a nudify app and photos of the girls' faces taken from social media sites, and were then circulated online.

These stories are becoming increasingly common in Australia with Yourtown (the provider of Kids Helpline) reporting that in the past 6 years, there has been a significant increase in the number of teenagers approaching Kids Helpline for assistance regarding image-based abuse.

Similarly, the UK Children's Commissioner describes these tools as "apps that aid in the abuse and exploitation of children," noting their rapid evolution and the serious risks they pose to child safety and wellbeing.[x] The research of the UK Children's Commissioner provides a useful overview of the commercial ecosystem surrounding nudification tools to date. The report identifies several specific nudification tools and platforms, including Nudify, DreamGF.ai, and Undress AI. Some of these are embedded in broader 'AI girlfriend' platforms, while others are standalone nudification apps. They are either provided for free or at low-cost, require minimal technical skill, and are marketed using euphemistic language like 'photo enhancement' or 'AI art.' The report also notes that derivatives of the original DeepNude app, which was taken down in 2019, continue to circulate online in modified forms.[xvi]

Recent research[xvii] out of the US also reveals how apps can be marketed as 'AI girlfriends' but instead provide a wholly different service as nudify tools.[xviii] The research also determined that the creation and dissemination of synthetic non-consensual intimate imagery has moved from a custom service available on niche internet forums, to an automated and scaled online business that leverages a myriad of resources to monetise and market its services.[xix]

Children often encounter these tools through social media platforms such as X, Telegram, and Instagram, where some accounts are actively promoting them. They are also easily discoverable through simple web searches using terms like "undress app" or "deep nude ai," and are similarly available on appstores.[xx]



© UNICEF Australia/Supplied

> “Teenagers experience significant emotional distress from these violations, with many perceiving nude deepfake abuse as even more harmful than the sharing of real sexual images.”

The commercialisation of these tools—through advertising, referral schemes, and integration into broader AI entertainment platforms—has normalised their use and obscured their abusive potential[xxi]. This creates a dangerous environment in which children are both consumers and potential victims of sexually explicit deepfakes, often without adequate safeguards or awareness.[xxii]

Growing concern among child safety experts is expressed in a 2025 report by the UK’s National Society for the Prevention of Cruelty to Children (NSPCC), which notes that GenAI tools can be used to create synthetic child sexual abuse material including by manipulating existing images of children or generating entirely fabricated content[xxiii]. These tools can be used to place a child’s face onto a nude or sexualised body, or to create fake chat logs and screenshots that falsely depict a child engaging in sexual or inappropriate behaviour.[xxiv]

The report also highlights that children themselves may use these tools to target peers, sometimes without fully understanding the legal or ethical implications. For example, a 14-year-old boy reported to Childline that he had used an AI tool to place a friend’s face onto an AI body and was later blackmailed by someone putting that face on a naked body[xxv]. This illustrates how GenAI can be used both to perpetrate and to facilitate abuse, even among children.

The psychological and social consequences of image-based abuse are significant. Children who are targeted with deepfake or nudified images often experience shame, fear, and isolation. In some cases, these images are used to extort victims – threatening to release the content unless money or further images are provided. The NSPCC’s Childline service has received reports from children who were blackmailed with AI-generated nudes, including one case where a child was threatened with exposure unless he paid £200.[xxvi]

The Internet Matters report on children’s experiences of nude deepfakes reveals a clear gender imbalance in how this form of abuse affects young people. It finds that an overwhelming 99% of nude deepfakes feature women and girls, underscoring how these technologies are disproportionately affecting them. While boys are more likely to report having engaged with or created nude deepfakes, girls are far more likely to be the targets of such abuse.[xxvii]

This dynamic reflects broader patterns of gendered online harm, where girls are more frequently subjected to sexualised image-based abuse and are led to adopting precautionary strategies to avoid victimisation, such as reducing their online presence, much like the safety measures they take in the offline world (e.g. avoiding walking alone at night). This “chilling effect” is limiting girls’ participation in digital spaces.[xxviii]

Teenagers experience significant emotional distress from these violations, with many perceiving nude deepfake abuse as even more harmful than the sharing of real sexual images. This is due to the lack of control over the creation and distribution of the image, the anonymity of perpetrators, and the fear that others will believe the image is real.[xxix]

The intersection of GenAI and image-based abuse represents a rapidly evolving threat to the safety and wellbeing of Australian children. The increasing accessibility of deepfake and nudification tools, combined with the social dynamics of adolescence and the gendered nature of online abuse, is creating a high-risk environment.

AI companions

The emergence of AI companions – i.e. chatbots and virtual agents designed to simulate human-like relationships – has introduced a new dimension to children's digital lives. These systems, powered by large language models (LLMs), are increasingly accessible through mobile apps, web platforms, and social media-integrated services. While marketed as tools for emotional support, entertainment, or self-expression, AI companions raise significant concerns regarding their psychological, developmental, and social impacts on children and young people in Australia. Replika, one of the leading providers of AI chatbot relationships, boasts around 30 million users worldwide.[xxx]

AI companions are widely available through platforms such as Replika, Character.AI, Pi.AI, and JanitorAI. These services allow users to create and interact with AI personas that can simulate friendship, mentorship, or romantic relationships. Some platforms, like Replika, offer premium features that enable erotic roleplay, while others, such as Character.AI, allow users to design custom characters, including those based on fictional or real-world figures. These tools are often promoted on mainstream platforms and are visually appealing to younger users, with few effective mechanisms to enforce age restrictions.[xxxi] The personalities of these AI are often engineered to cater to the human desire for praise and validation which can mean that children's beliefs, behaviours, and even sense of identity could be shaped by these systems, at a time of their formation and development that is particularly sensitive.[xxxv] Conversations can be voice, text or even video.

Recent studies and inquiries, although sometimes not focusing on children, have revealed that children and young people are already engaging

© UNICEF Australia/Supplied

Children's use of AI companions

The Age has reported that many young Australians are using AI chatbots for therapy and companionship on a daily basis including disturbing stories around reactions and comments from the AI companions. In June last year, the eSafety Commissioner received an email from a group of concerned school nurses who noticed a spike in children as young as 10 spending hours a day talking to AI bots, often sexually.

Companion bots made global headlines last year after a 14-year-old boy in the United States died by suicide after forming a romantic attachment to an AI character he created on Character.AI. The AI reportedly encouraged his suicidal ideation. This incident has prompted legal action and renewed scrutiny of the emotional risks posed by AI companions.

with AI companions in diverse and complex ways. A study by the University of Sydney found that approximately half of AI companion users reported using them for friendship, a third for romance or sex, and nearly one in five for counselling.[xxxvi] The Australian Psychological Society has noted that teens may be drawn to AI companions for their non-judgmental nature and constant availability, particularly as a way to practice social skills or cope with loneliness.[xxxvii]

Australian youth mental health services have also begun to observe the growing presence of AI companions in young people's lives. The LifeChanger Foundation, which works with over 135,000 young people across Australia and New Zealand, reported that AI companions are increasingly mentioned in workshops as sources of support. While some teens use them as a form of journaling or emotional expression, others appear to rely on them as substitutes for human connection.[xxxviii] The eSafety Commissioner has noted that most AI-driven chatbots were not designed with safety in mind and children were not developmentally ready for the risks.[xxxix]

The emotional bonds that children form with AI companions can resemble parasocial relationships - one-sided emotional attachments typically formed with media figures. In the context of AI, these relationships are not only persistent but also interactive, which can intensify the illusion of mutual understanding.[xl]

Research has shown that many AI personas are designed to simulate romantic or sexual interactions, often with underage or sexualised minor characters. A 2025 investigation identified over 10,000 chatbots labelled as sexualised minor personas, many of which were accessible through platforms like Chub AI and JanitorAI. These bots often bypass moderation filters and are promoted in online communities that share tactics for evading safeguards.[xli] Some bots simulate

“While some teens use them as a form of journaling or emotional expression, others appear to rely on them as substitutes for human connection.”

grooming scenarios or depict children in incestuous or exploitative contexts.[xlii] These personas are not only harmful in themselves but may also serve as gateways to illegal content or normalise abusive behaviour.

The proliferation of AI companions is driven by a combination of technological, economic, and social factors. Open-source LLMs such as Meta’s LLaMA and Mistral’s Mixtral allow users to fine-tune models for specific purposes, including the creation of unmoderated AI personas. Meanwhile, proprietary models like OpenAI’s GPT-4 and Anthropic’s Claude are frequently jailbroken to bypass safety filters. Online communities on Reddit, 4chan, and Discord actively share jailbreaks, character cards, and chat logs to facilitate the creation of AI companions that can engage in harmful or explicit content.[xliii]

AI companions, as chatbots more generally, often rely on gamification, personalisation, and emotional responsiveness to retain users, which can inadvertently encourage overuse and emotional dependency. Heavy users of AI companions are eventually more likely to feel lonely, socially withdrawn, and emotionally dependent on their bots.[xliv] However, even though children are increasingly using bots, little is known about their effects – especially long-term – on children's social, emotional and cognitive development.[xlv]

Lastly it is worth noting that the vast data collection on children through AI companions raises significant privacy concerns – especially with the deeply sensitive nature of the personal information that may be shared.

© UNICEF Australia/Supplied

Child exploitation material and AI

Various reports and research have shown how GenAI, in particular image-generation models, can be used for child sexual abuse material (CSAM) related purposes. This presents a growing risk for Australian children. ICMEC Australia's SaferAI for Children Coalition has made it clear that the assumption that AI-generated CSAM is less harmful because it depicts 'fictional' children or does not involve physical harm to a child – is false.[xlvi] Instead, this content advances the exploitation and abuse of children and reiterates harmful behaviours among offenders by essentially contributing to the 'normalisation' of child sexual exploitation, blurring the lines between real and synthetic material.[xlvii]

There are two key types of CSAM that can be generated using GenAI:

a. **Creating Fully synthetic CSAM** – GenAI can be used to generate full synthetic CSAM. With the advanced capabilities of image-generation models, offenders can generate synthetic images that are virtually indistinguishable from real images.[xlviii] Through its tipline, the child safety organisation in the United States, the National Center for Missing & Exploited Children (NCMEC) has identified prompts that have been used to generate such images, including prompts instructing models to produce content depicting children in sexually explicit acts without involving any real child.[xlix]

© UNICEF Australia/Supplied

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.[liv] The project's primary objective is to develop an ethical AI tool designed to improve the efficacy of machine learning to counter child sexual exploitation.[lv]

This tool aims to significantly reduce the manual work required by investigators, enabling quicker identification of potential CSAM. To effectively train the model, the researchers did a call-out for adult Australians to contribute non-harmful childhood photos of themselves, targeting a collection of around 100,000 images from all ethnicities.[lvi] Importantly, data ethics were at the core of this project and all images were collected from adults who provided clear consent.

b. **Manipulating existing non-CSAM content** – GenAI can also be used to manipulate existing content of real children to create CSAM, even if the original content is not exploitative. This includes the use of deepfake technology capable of altering faces, bodies or voices to produce synthetic CSAM using real content. Internet Watch Foundation (IWF) has found that some deepfake CSAM videos shared on dark web forums involve taking adult pornography videos and adding a child's face. Offenders may also use AI to modify previously created abusive content to generate new CSAM, thereby increasing the volume and variety of material available. Importantly, this can result in evading detection by making it difficult for law enforcement to identify victims and remove them from harm.[li]

GenAI lowers the technical barriers for CSAM creation and enables offenders to generate large volumes of synthetic or partially synthetic CSAM quickly and anonymously. This material or its associated prompts can then be distributed on dark web platforms and accessed and/or used by other offenders. The data sources used to construct such datasets can include social media websites as well as popular adult video sites, and models trained on such material can more easily be used to generate CSAM using the right prompt.[lii] As the technology becomes more sophisticated, it is becoming "visually indistinguishable" from authentic child abuse images according to IWF.[liii]

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[lviii].

In 2023, the US-based NCMEC received via its tipline over 4,000 reports related to CSAM involving GenAI.[lix] In Australia, the Australian Centre to Counter Child Exploitation (ACCCE) noted to the

ABC that roughly 40,000 online child exploitation reports are received by the ACCCE per year.[lx] As at August 2024 fewer than 20 of those were AI-generated, however the prevalence is growing.[lxi]

While CSAM being created by the use of GenAI is clearly a significant risk, there are a number of global initiatives that are focused on using AI to detect CSAM, identify patterns in grooming tactics and offender behaviour, automate the review and categorisation of CSAM and present warning messages.[lxii] ICMEC Australia has noted that the rise in CSAM reports has created an urgent need for enhanced support to law enforcement agencies as 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.[lxiii] The eSafety Commissioner has noted that AI technology is essentially a "double-edged sword" as the same technology that can be used by perpetrators to exploit children is also being used to identify new offending material, support platforms in removing this material, and assist law enforcement by triaging suspected material for human review.[lxiv]

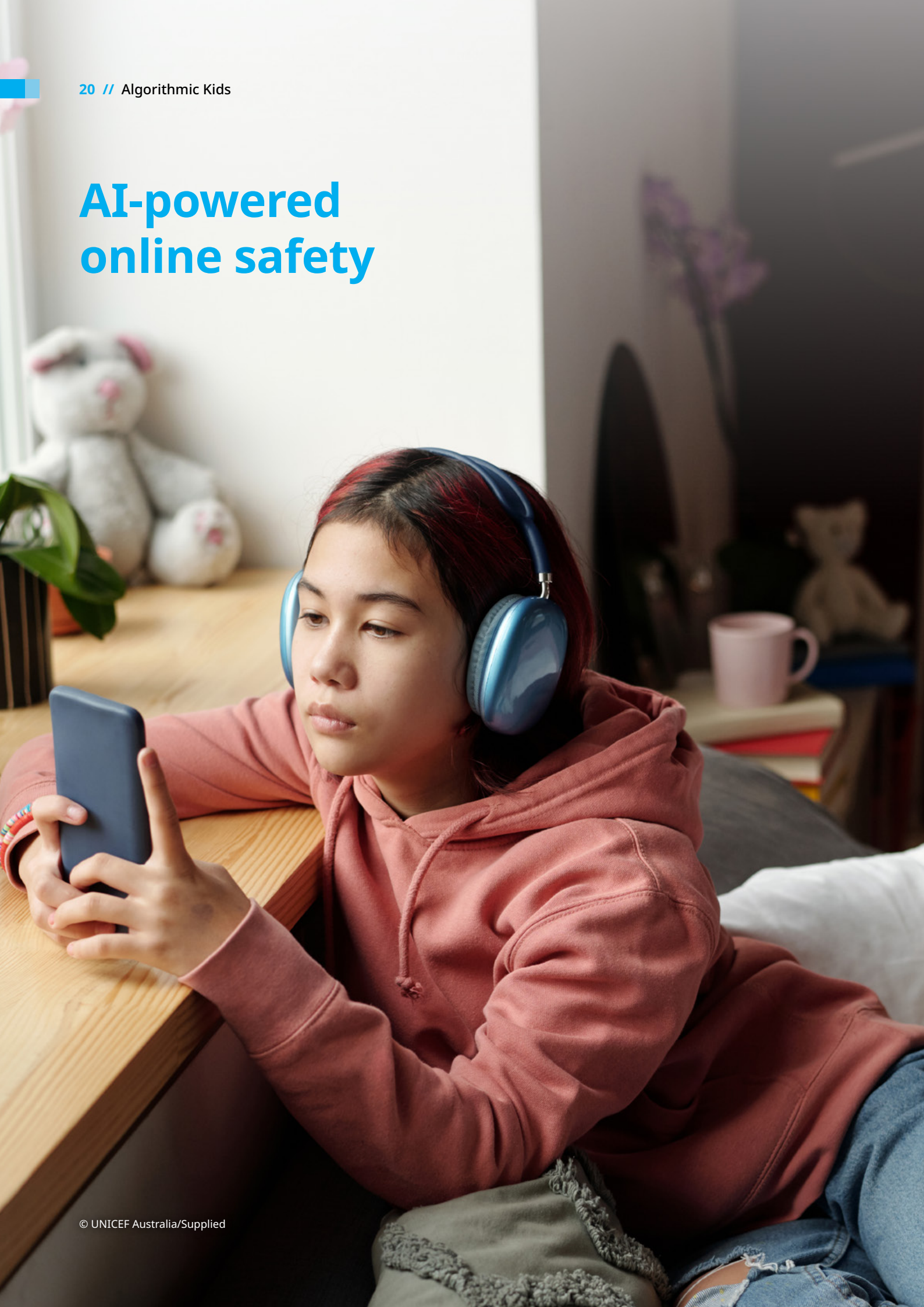
BBC Investigation on Stable Diffusion

In a 2023 UK report, the BBC found that Stable Diffusion which is an image generating AI model developed by Stability AI, was being used to create life-like images of child sexual abuse.[lvii] The Stable Diffusion app is described on the App Store as "an AI image generation app that can automatically generate images from text."

The increasing capability of video-generation models and related technology creates the potential for greater quantities of increasingly realistic CSAM. Without appropriate safeguards, there is a significant concern around the use of GenAI tools for the production of CSAM. These models also present unique challenges regarding the dissemination and detection of such material on the internet.

© UNICEF Australia/Supplied

AI-powered online safety




UNICEF Australia/2025/Andres

Despite the risks explored in this report, AI also presents opportunities to uphold children's rights and protect them in digital spaces. In trust and safety teams at tech platforms, for example, several applications of AI techniques have emerged that can be ultimately beneficial to children and make their online experiences safer. Over the years, social media platforms and third-party service providers have developed tools that help trust and safety teams in detecting content and behaviour that is illegal and/or violates their terms of service. This includes, for example, detecting hateful and extremist content, identifying networks of scammers, terrorist recruiters, or child abusers, or estimating the age of users.

A popular application of AI for trust and safety purposes is in the field of age estimation, where platforms want to understand the age of their users, apart from self-declared age, to provide them with age-appropriate content and user experience. Meta has trained an algorithm to predict whether a user is an adult (18 and over) or a teen (between 13 and 17).[lxv] In order to provide such an estimation, these systems process user generated content, including interaction with posts published by other users.

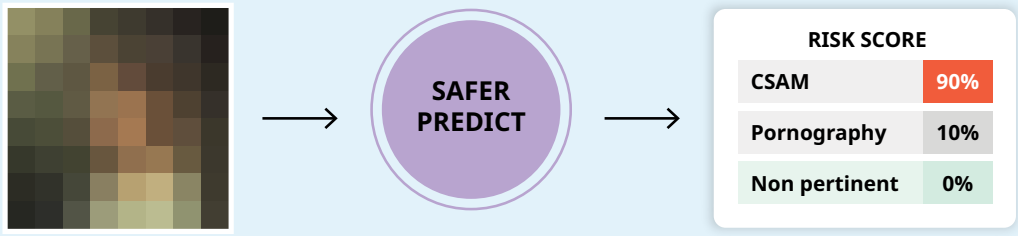
Meta has published information on several other technologies they employ for trust and safety purposes. An example is a technology used for countering child grooming. The company has released information on how they rely on data such as user profiles and group descriptions to estimate the occurrence of child grooming on WhatsApp.[lxvi] The technology they employ relies on the use of machine learning to process data patterns which possibly correlate with the sharing of CSAM – this includes, for instance, WhatsApp groups with large numbers of members and a frequent turnover.[lxvii]

Platforms are similarly relying on content identification and analysis of social graphs to combat extremist content and networks. These technologies usually rely on keyword lists developed by experts over time, including with the support of machine learning tools. In 2017, Meta also disclosed it relies on machine learning to detect terrorist groups on Facebook.[lxviii] The company uses trained models to identify user generated content, including posts, pages, groups, or profiles, that support terrorism.[lxix] These models then assess whether other users may be engaging in such activities. To do so, models process

Safer Predict built by Thorn

Safer Predict by Thorn uses AI tools like machine learning to detect child sexual abuse material (CSAM) including images, videos and text. Unlike other methods where previously known CSAM is used to look for matches and then remove the content, Safer Predict uses machine learning to analyse text, images and video, and then provide a risk score based on how likely the content is to be harmful.

The risk score signifies not just how likely the content is to be abusive, but also provides indicators and descriptions of the identified abuse, which in turn is meant to help human decision makers – like trust and safety teams at tech platforms, or law enforcement officials – respond to abusive content more effectively.[lxxiv]



A visualisation of how Safer Predict operates

Source – Thorn, https://get.safer.io/hubfs/Safer/Safer%20One%20Pagers/Safer%20Predict%20Overview.pdf, 2025

information such as whether a user is friends with several accounts that have been already blocked due to terrorism-related violations, or whether there are other characteristics in common between such accounts.[lxx]

These technologies can be used not only to identify perpetrators, but also those children and young people who are vulnerable to radicalisation online and can be used in coordination with other actors for countering radicalisation, such as law enforcement authorities, but also civil society organisations. This is particularly relevant as children have been found to be increasingly exposed to radicalisation in recent years.[lxxi]

AI can also help keep children away from hateful content, as social media platforms rely on models to identify users that are engaging in hate speech, trolling other users, or harassing them. In 2018, Twitter published a blogpost to

describe its use of machine learning models to identify trolling on its platform.[lxxii] These models relied on processing information such as the creation of multiple accounts, users "tagging" or "mentioning" accounts that they did not follow, and any interactions with accounts already found to violate the platform's terms of service. However, it is important to note that automated moderation systems should be rigorously and continually evaluated on their performance for users who are most vulnerable and most at risk, with appropriate safeguards implemented.[lxxiii]

AI can also be used for bot-detection – i.e. identifying social media profiles operated by bots. These kinds of accounts engage in coordinated behaviour that can serve several harmful practices: from disinformation campaigns to trolling and harassment, political manipulation, spamming and scamming. In 2020, Meta published a paper to introduce SybilEdge, a machine learning algorithm designed by the platform for the early identification of fake accounts.[lxxv] The system processes data such as friend requests and rejection rates of such friend requests, to determine if an account is fake even before any violation has occurred.

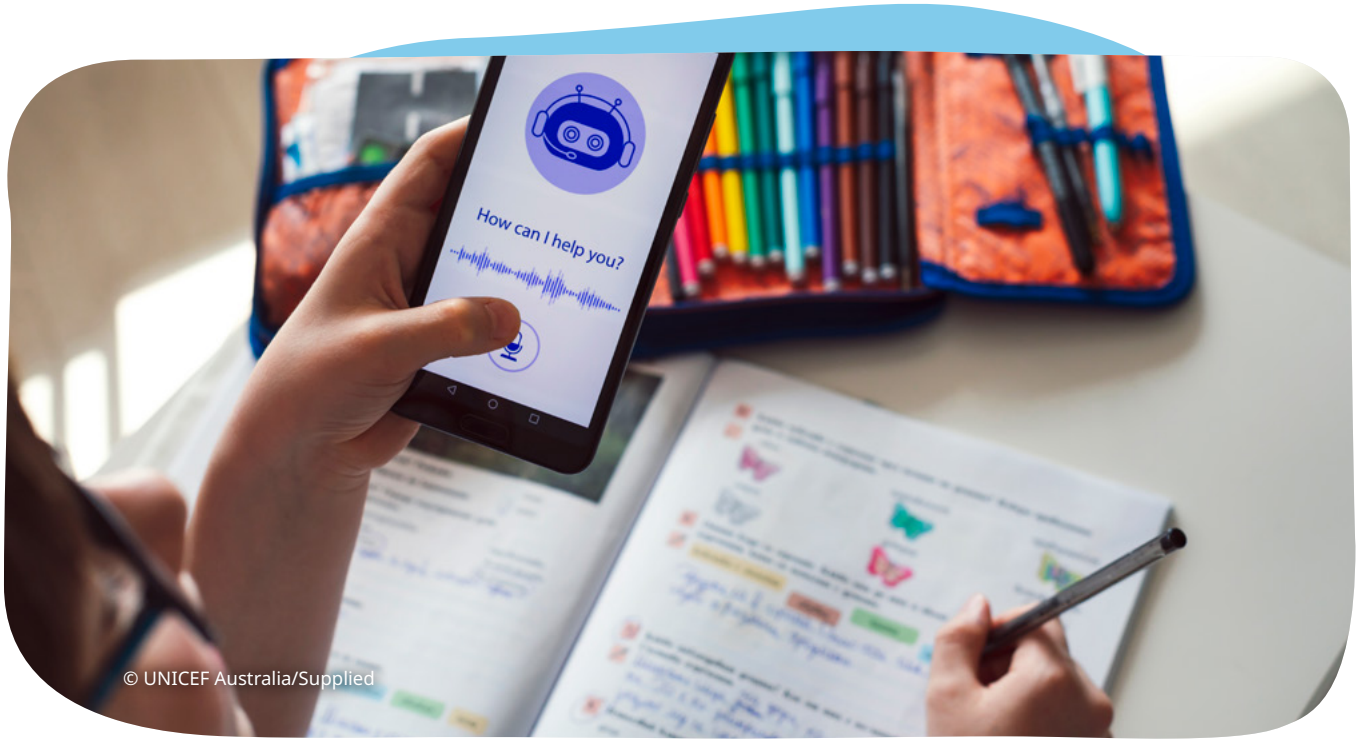
Another key aspect of the AI-powered trust and safety landscape is third-party providers who offer managed trust and safety services to support content moderation and other online safety practices. These are specialised service providers who use their research on AI and sector expertise to provide state of the art content identification and behavioural analysis tools.

These providers employ several AI technologies that can analyse not only content, but also its context, in order to make more precise predictions, deploying specialised algorithms that are trained to target a specific type of harm, and which allow customers to address several trust and safety issues.[lxxvi] Sometimes customers can fine-tune or even retrain algorithms based on their own specifications and data.[lxxvii]

A similar initiative has emerged under the auspices of the United Nations Interregional Crime and Justice Research Institute (UNICRI). The AI for Safer Children initiative hosts a Global Hub since 2022: an online platform where law enforcement officials can access a catalogue featuring comprehensive information on over 80 "cutting-edge AI tools" that are designed to support the investigation of crimes against children. It also provides a learning centre that equips law enforcement professionals with resources to improve their understanding of how to apply AI to investigations in relation to child abuse, how to improve their overall investigative techniques, and maintain their mental wellbeing. It also offers guidance on the responsible use of AI to streamline investigative processes.[lxxviii]

UNICEF Australia/2025/Andres

Education and AI



© UNICEF Australia/Supplied

Although data shows that almost one in two Australian children have used a GenAI powered chatbot to brainstorm ideas for their homework, more extensive data on concrete use cases for AI in education still needs to be collected. UNICEF's Policy Guidance on AI for Children acknowledges the potential of relying on AI to support children's education and help them collaborate, solve problems, and think critically.[lxxix] Other possible adoptions include personalised systems for smart tutoring, which could support teachers in helping children with learning difficulties.[lxxx]

Some insight in this area is provided by the parliamentary report on the inquiry into the use of GenAI in the Australian Education System.[lxxxi] As the report notes, the Australian education system is increasingly embracing GenAI, with the Australian Framework for Generative AI in Schools which came into effect in 2024 now encouraging its use in schools and universities actively integrating it into their curricula.[lxxxii] This shift reflects a broader recognition that GenAI is here to stay, prompting calls for supportive measures, including guidance and safeguards for all stakeholders—students, educators, parents, and developers.

The report mentions personalised learning, enhanced educational outcomes, and increased efficiency in managing administrative tasks as the key benefits GenAI may deliver.[lxxxiii] The Association of Heads of Independent Schools of Australia shared a survey which found that supporting student research, generating ideas for creative projects, and providing feedback to improve writing, were the most common use of AI tools by students. As to the perceived benefits from the use of AI, 43% of respondents reported improvements in terms of student participation or in their performances, or both.[lxxxiv]

In relation to the types of tools that are being used in Australian schools, ChatGPT emerged widely as the most utilised tool in the Australian education system, which may have implications in relation to quality. Be it ChatGPT or other popular chatbots powered by LLMs, these models have not been developed *for schools* so independent quality assurance is likely to be needed if these tools have to be relied upon to improve student learning.[lxxxv]

According to UNESCO, chatbots such as ChatGPT can enhance teaching and learning by taking on various roles that support both students

NSWEduChat

The NSW Department of Education is currently running a trial of NSWEduChat, the Department's GenAI tool, to provide staff with a safe way to explore the benefits of GenAI.[lxxxix] The tool aims to provide additional support to staff to develop and deliver teaching experiences, easing workload and helping users build their AI skills.[xc]

Initially trialled in 50 schools, feedback from teachers was so positive in terms of efficiencies and engagement, that the Department has now enabled access to the tool for all staff. Students are also involved in the trial with the results forthcoming. NSWEduChat has strong privacy and safety protections in place and aligns to the NSW AI Assessment Framework.[xci]

© UNICEF Australia/Supplied

and teachers. They can perform simple and technical tasks like research, calculations, and proofreading, while also acting as a creative and interactive tool. As a "possibility engine," it can help students explore alternative ways of expressing ideas; it can serve as a "Socratic opponent" to develop arguments, or as a "collaboration coach" for group problem-solving. It can support learning through data interpretation as an "exploratorium" or motivate students with games and challenges.[lxxxvi]

Students may rely on these technologies to consolidate their understanding of concepts, but also to seek answers to questions they may hesitate to ask teachers.[lxxxvii] Institutions have emphasised how GenAI can inspire students facing creative blocks by generating lists of ideas and supporting their creative process without replacing them. Other concrete examples on how GenAI is already being used by students includes to test story ideas, receive feedback, or build narrative frameworks. This may especially help those students who struggle with writing or feel anxious about their skills, to become more confident in their creative process.[lxxxviii]

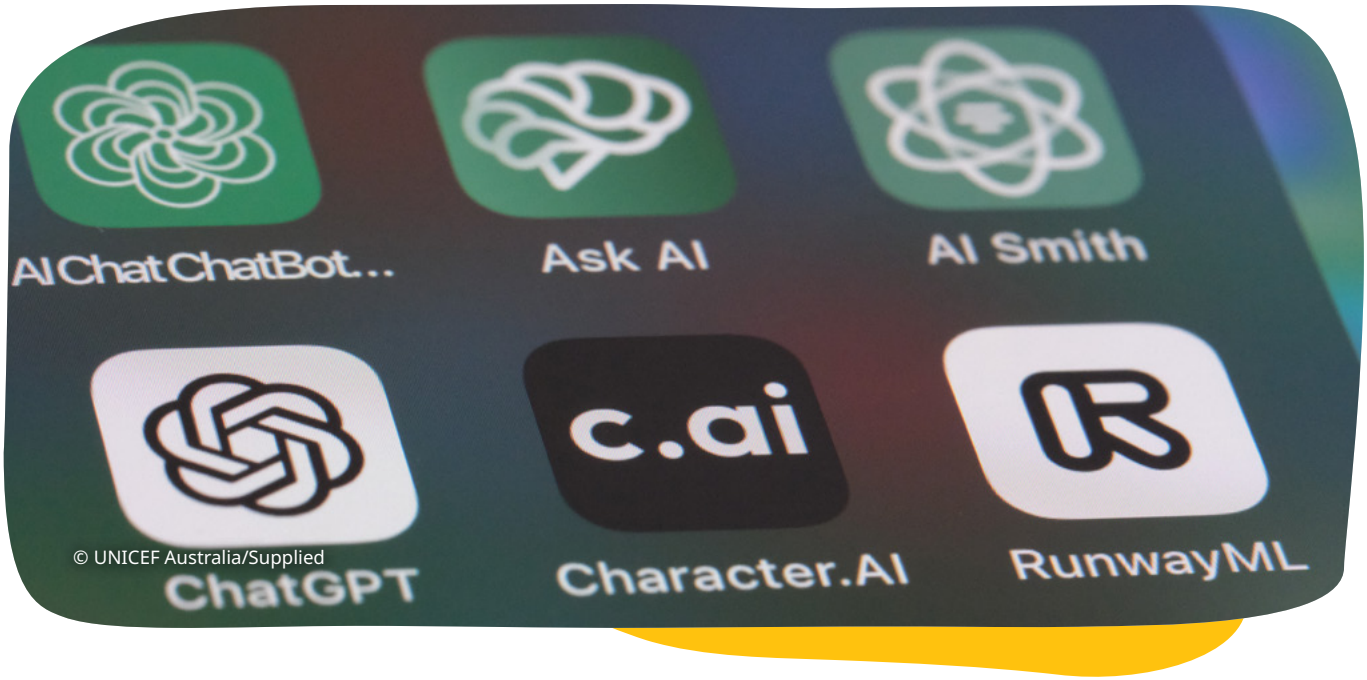
Another potential opportunity for AI and education is personalised learning. Algorithms can support mentorship and continuous learning by providing tailored resources and content to individual needs and interests, as well as level of knowledge. As GenAI can explain concepts in various styles, this can also augment teacher explanations while promoting independent learning and deeper comprehension. By adapting to student prompts and feedback, chatbots can offer one-on-one tutoring experiences that engage even disengaged learners. As a tutor, GenAI can also provide instant, personalised feedback on tasks like essays, helping students identify and improve on their weak spots. Again, this may be particularly beneficial for students with learning difficulties or with additional learning needs, as well as in regions with teacher shortages.[xciii] A concrete example is US-based Khan Academy's Khanmigo tool, which can scale personalised learning by generating custom content, analysing learning styles, and recommending strategies and resources tailored to each student's goals.[xciv]

Another specific use case for GenAI in education is in the context of teaching children how to code. For example, Tynker, a game-based coding platform, has developed an LLM-powered coding tool called Tynker Copilot which is designed to help children use visual block code to develop their own apps and games.[xcv] The tool is a fine-tuned model trained on thousands of kid-created projects. Children can input natural language commands such as "Teach me how to build a Fruit-ninja style game" and Tynker Copilot will generate the corresponding block code with step-by-step instructions to build the game.

In addition to these applications, AI tools can also support children's education by supporting those who teach children. This could include by helping to ease teacher workloads and administrative burden by automating time-consuming tasks such as grading, attendance tracking, lesson planning, and report writing. These efficiencies may free up time for teachers to focus on student engagement. GenAI may improve task efficiency and support differentiated learning, especially in under-resourced communities. However, successful adoption depends on providing teachers with adequate training and professional development to integrate GenAI effectively into their practice.[xcvi]

Finally, GenAI tools can provide educational institutions with data-driven insights into student growth and emerging trends, as well as concerning learning gaps.[xcvii] Chatbots can also support administrative teams and free up their time by supporting them in tasks such as providing administrative support in relation to generic queries, information on course requirements and sign up processes, retrieving information, scheduling reminders, and translating material for international students and staff.[xcviii]

UNICEF Australia/2025/Andres



The regulatory landscape, at home and abroad

Before exploring where Australia might need to make change to ensure we can maximise the potential of AI for children, it is helpful to first examine where things currently stand. For Australia at the moment, that means a patchwork of existing laws and policies that cover the digital world including AI, in addition to a few dedicated AI reform initiatives, some of which are explored here.

For the most severe impacts of AI in Australia, the Criminal Code Act 1995 criminalises possessing, controlling, producing, accessing, distributing, obtaining, and supplying AI-generated CSAM.[xcix] In addition, the Criminal Code has been amended to make clear that the offence of using a carriage service to transmit sexual material without consent applies even when an image has been created using deepfakes.[ci]

When it comes to data and privacy, the Australian Privacy Act and Australian Privacy Principles apply. The Australian Office of the Information Commissioner (OAIC) has released guidance for Australian organisations deploying AI, as well as for organisations who are training AI which provides guidance on privacy due diligence and the intersection of the Australian Privacy Principles with AI.[cii] The OAIC is also currently developing a new Children's Online Privacy Code,[ciii] which will provide a new legal framework for how children's data is collected and used, with obvious implications when it comes to AI.

In terms of broader online safety and AI, the online safety regime in Australia is regulated by the eSafety Commissioner. The Online Safety Act 2021 (Cth) (OSA) grants the eSafety Commissioner broad powers to investigate and act on harmful online content, including AI-generated child sexual exploitation material. Under the OSA, such content falls within defined categories of illegal and restricted material, and eSafety can compel its removal across a wide range of digital platforms.[civ]

The OSA also empowers eSafety to issue transparency notices under the Basic Online Safety Expectations (BOSE).[cv] These Expectations outline the minimum safety standards that online service providers are expected to meet. They cover a range of digital risks, including the use of generative AI, encrypted services, anonymous accounts, and recommender systems. Providers must demonstrate how they are ensuring these features are used safely, particularly in services likely to be accessed by children.

The OSA also provides for the development and enforcement of industry codes and standards for illegal and restricted online material which establish mandatory requirements on key services.[cvi] These codes apply to key sectors of the digital ecosystem, including those deploying generative AI. There are currently six registered industry codes and two enforceable standards, with more under development.[cvii] Where providers fail to comply, eSafety has the authority to impose penalties and take enforcement action to uphold safety standards. The BOSE and industry codes and standards are designed to protect Australians from illegal and restricted online content and are key mechanisms for industry to address the evolving online safety challenges posed by AI for children.

The Albanese Labour Government also announced in late 2024 that it will be legislating a Digital Duty of Care that will place the onus on digital platforms to proactively keep Australians safe and better prevent online harms.[cviii] It is unclear when or if this will legislated, but if implemented it stands to have wide-ranging impacts on the makeup of the online world including AI technologies.

In education, the Australian Framework for Generative AI in Schools seeks to guide the responsible and ethical use of generative AI tools in ways that benefit students, schools, and society.[cix] It includes 6 principles and 25 guiding statements.

Finally, Australia also released its proposal for Introducing Mandatory Guardrails for AI in High Risk Settings in September 2024[cx]. The paper called for feedback on the definition of high risk settings, and submissions have since been made and closed. The government is set to use the submissions as the basis of its thinking "on next steps, including how to best apply the proposed guardrails."[cxi]

International comparisons

The European Union (EU) has in place a comprehensive framework regulating digital technology. Regarding AI specifically, in 2024 the EU passed its AI Act which applies to AI systems developed and deployed in the Union. It uses a risk-based approach whereby AI systems of varying risk levels are subject to specific sets of obligations, with the main focus of the Act being on 'high-risk' AI systems including in the area of education and vocational training.[cxii]

It has been reported, however, that the European Commission is considering delaying the AI Act's entry into application and making targeted 'simplification' amendments to be proposed later this year.[cxiii] These considerations arise in the context of the Commission's 2025 work programme which is focused on cutting red tape and fostering innovation to support Europe's global competitiveness.[cxiv] The AI Act currently implements a staggered approach to its application whereby different provisions come into effect at different times. So far the provisions on AI literacy and prohibited AI systems are in effect, with provisions on general-purpose AI models set to come into effect in August 2025.

In the United Kingdom (UK), the government has not yet introduced broad AI legislation like the EU AI Act. The policy narrative regarding AI has so far mainly revolved around managing the risks of frontier models and how AI can be used to foster innovation and economic growth. This is evident from the publication of the UK government's AI Opportunities Action Plan in January 2025 which sets out plans for AI growth zones, new infrastructure and national data libraries for AI.[cxv]

Nevertheless, the UK's regulatory framework does contain legislation that applies to the development and deployment of AI systems. This includes the data protection regime, which was recently amended with the passing of the UK Data (Use and Access) Act 2025. Under the Act, those providing online services likely to be used by children must consider 'children's high protection matters' when implementing measures to protect their personal data.[cxvi]

Where to next for Australia

Australia has often been at the forefront of digital reform, but AI is a disruptive and transformative technology that is compelling us to consider our place in the world – what uniquely Australian imprint are we going to leave on the global AI race?

Children, due to their rapid adoption of new tech and unique vulnerability to its effects, often provide an early indication of what might be the trials and tribulations of the next technological frontier. When it comes to AI and children, we are seeing that it is an amplifier of both risk and opportunity. Be it through deepfakes and image-based abuse, AI companions, or AI-generated child exploitation material, there are some pretty loud alarm bells ringing. But we can also put that same technology to use in service of children, and whether it is online safety or education, the potential for positive impact is promising.

As we have done in the past, Australia has an opportunity to lead from the front and put our stamp on this exciting new technology. For UNICEF Australia, we want Australia to be the global champion for child-centred AI. We imagine a future where our neighbours look to us as the global hub for innovation in safe and responsible AI, the first true leaders in putting AI to work for the betterment of all in our society, including children. To that end, we offer a series of initial recommendations to help Australia realise this vision:

1

Utilise existing digital reforms in Australia to make them fit-for-purpose for AI, then plug any gaps with new laws and regulations

Multiple current digital reforms present an opportunity to minimise the risks posed by AI for children. The Privacy Act reforms, and particularly the development of Australia's first-ever Children's Online Privacy Code, are an obvious example. AI models survive on data. Through the privacy reforms, we can create new rules so children's data is only collected and used when in their best interests, protecting them at the source. Similarly, the Online Safety Act needs updating, namely through the creation of a new digital duty of care that compels tech companies to design their platforms in a way that is safe for children, which would include GenAI tools. These reforms should be seen through as a matter of priority. Where they leave risk unaddressed – say high-risk AIs that exploit vulnerable groups like children – then new laws and frameworks should be adopted.

2

Conduct a scoping study into age-gating AI companions, chatbots and nudify apps

AI companions, chatbots and nudify apps are presenting new risks for children, particularly those of a romantic and sexualised nature designed primarily for adults. Australia is currently running an Age Assurance Technology Trial. The findings of that trial could inform a short scoping study to help determine whether AI companions, chatbots and nudify apps should be age-gated in Australia, and how it would work in practice.

3

Fund research to better understand the risks posed by AI for children, particularly the most severe use cases such as AI-generated child exploitation material

Child sexual exploitation and abuse is a prominent issue that Australia needs to grapple with, and AI is only amplifying harm. GenAI tools in particular have the potential to create an explosion of abusive content, stretching the already limited resources of law enforcement and others. To better address the issue, we need to understand it in greater detail. Thankfully, tried and tested evidence generation projects already exist in this space, like the Disrupting Harm[cxvii] project. Funding research like this in Australia would put as at the forefront of tackling AI-facilitated child sexual exploitation, and help our local neighbours in the region do the same.

4

Invest in AI-safety technologies and tools for Australian law enforcement and civil society

We can use AI to turn the tables on some of the new risks the technology has introduced. An obvious area where this applies is augmenting law enforcement and civil society in Australia with AI that makes them more effective and efficient, for example, through AI tools that help with the identification and removal of child sexual abuse material. Investment now will help ensure Australia stays ahead of the game, given we expect the volume of this kind of illegal content to increase.

5

Establish an Australian AI Safety Institute, with a dedicated workstream on children

Australia has an opportunity to become a global leader in safe, responsible and child-centred AI. A dedicated AI Safety Institute could lead technical research including into the impacts of AI on children, and be a bridge between government, industry, and the public. It would also be a vehicle for sharing Australian innovation with the world, and be the focal point for participation in important global initiatives like the AI for Safer Children Initiative.

6

Integrate AI literacy into the national curriculum and raise public awareness

Equipping Australia's next generation with the skills to both use AI safely and capitalise on the opportunities it provides, is of great importance. For such a disruptive technology, nation-wide literacy and skills building is going to be needed. We also need to simultaneously raise public awareness to upskill those who are no longer in the education system, for example, parents, so they have the confidence to safely guide their children through an increasingly AI-augmented world.

Endnotes

i. UNICEF Australia 'Protecting children in the online world – reshaping the digital world for Aussie kids', 2024

ii. UNICEF 'Policy Guidance on AI for Children,' November 2021, page 20.

iii. Ibid.

iv. UNICEF Australia 'A Generation Online: Ensuring Children and young people thrive in a digital world', 2025, page 4.

v. UNICEF Australia 'Protecting children in the online world – reshaping the digital world for Aussie kids' , 2024, page 10.

vi. eSafety Commissioner website: 'eSafety urges schools to report deepfakes as numbers double', 27 June 2025.

vii. Ibid.

viii. eSafety Commissioner website: 'Deepfake damage in schools: how AI-generated abuse is disrupting students, families and communities', 27 June 2025.

ix. Ibid.

x. UK Children 's Commissioner, 'One day this could happen to me" Children, nudification tools and sexually explicit deepfakes', 28 April 2025.

xi. Sydney Morning Herald 'Year 11 students suspended after AI fake nudes circulated online', 21 February 2025.

xii. Ibid.

xiii. The Guardian 'Schoolboy arrested after 50 female students allegedly targeted in fake explicit AI photos scandal', 12 June 2024.

xiv. Ibid.

xv. Yourtown 'Yourtown (powering kids helpline) Annual Report 2024', 2024, page 12.

xvi. Ibid.

xvii. Graphika 'A Revealing Picture: AI-generated 'undressing' images move from niche pornography discussion forums to a scaled and monetized online business', December 2023 and Bellingcat website 'Meta's Suit Against Hong Kong Firm Was Just the Beginning – More Companies Linked to CrushAI 'Nudify' Apps', 18 June 2025.

xviii. Bellingcat 'Meta's Suit Against Hong Kong Firm Was Just the Beginning – More Companies Linked to CrushAI 'Nudify' Apps', 18 June 2025.

xix. Graphika 'A Revealing Picture: AI-generated 'undressing' images move from niche pornography discussion forums to a scaled and monetized online business', 2023

xx. UK Children 's Commissioner, 'One day this could happen to me" Children, nudification tools and sexually explicit deepfakes', 28 April 2025.

xxi. UK Children 's Commissioner, 'One day this could happen to me" Children, nudification tools and sexually explicit deepfakes', 28 April 2025.

xxii. Ibid.

xxiii. NSPCC, 'Viewing Generative AI and Children's Safety in the Round', January 2025, page 11.

xxiv. NSPCC, 'Viewing Generative AI and Children's Safety in the Round', January 2025, page 10.

xxv. NSPCC, 'Viewing Generative AI and Children's Safety in the Round', January 2025, page 11.

xxvi. NSPCC, 'Viewing Generative AI and Children's Safety in the Round', January 2025.

xxvii. Internet Matters, 'The new face of digital abuse: Children's experiences of nude deepfakes', 2024.

xxviii. UK Children 's Commissioner, 'One day this could happen to me" Children, nudification tools and sexually explicit deepfakes', 28 April 2025.

xxix. Internet Matters, 'The new face of digital abuse: Children's experiences of nude deepfakes', 2024.

xxx. The Verge 'Replika CEO Eugenia Kuyda says it's okay if we end up marrying AI chatbots' Nillay Patel, 13 August 2024.

xxxi. eSafety Commissioner website: Online safety advisory 'AI chatbots and companions – risks to children and young people' 18 February 2025.

xxxii. The Age 'Isabellea 'couldn't be without' her best friend. He wasn't real', Emily Kowal, 28 June 2025.

xxxiii. Ibid.

xxxiv. eSafety Commissioner website: Online safety advisory 'AI chatbots and companions – risks to children and young people' 18 February 2025 and European Conference on Information Systems 'A.I., All Too Human A.I.: Navigating the Companionship/Alienation Dialectic', Ciriello et al, 13 June 2025.

xxxv. UNICEF Innocenti 'Beyond Algorithms: three signals of changing AI-child interaction' Steven Vosloo and Cecile Aptel, 23 May 2025.

xxxvi. European Conference on Information Systems 'A.I., All Too Human A.I.: Navigating the Companionship/Alienation Dialectic', Ciriello et al, 13 June 2025.

xxxvii. Australian Psychological Society, 'APS discusses AI chatbot companions with ABC', 7 January 2025.

xxxviii. eSafety Commissioner website: Online safety advisory 'AI chatbots and companions – risks to children and young people' 18 February 2025.

xxxix. eSafety Commissioner website: 'New safety advisory warns unrestricted chatbots threaten child development', 18 February 2025.

xl. European Conference on Information Systems 'A.I., All Too Human A.I.: Navigating the Companionship/Alienation Dialectic', Ciriello et al, 13 June 2025.

xli. Graphika 'Character Flaws: School Shooters, Anorexia Coaches, and Sexualized Minors: A Look at Harmful Character Chatbots and the Communities That Build Them' Cristina Lopez G et al, March 2025.

xlii. Ibid.

xliii. Ibid.

xliv. AI Frontiers 'A Glimpse into the Future of AI Companions,' Vanessa Bates Ramirez, 29 May 2025.

xlv. UNICEF Innocenti 'Beyond Algorithms: three signals of changing AI-child interaction' Steven Vosloo and Cecile Aptel, 23 May 2025.

xlvi. ICMEC Australia Safer AI for Children Coalition 'Artificial Intelligence and Child Protection: a collaborative approach to a safer future', 2024, page 8.

xlvii. ICMEC Australia Safer AI for Children Coalition 'Artificial Intelligence and Child Protection: a collaborative approach to a safer future', 2024, pages 7 – 8.

xlviii. Catholic University Journal of Law and Technology, 'Understanding the Line Between Art and Abuse: How Generative AI Changes the Landscape of Child Sexual Abuse Materials', Vol. 33 2024, Amy Trivision, page 103.

xlix. Missing Kids Blog 'Generative AI CSAM is CSAM', 3 November 2024.

l. Internet Watch Foundation 'What has changed in the AI CSAM landscape?', July 2024.

li. ICMEC Australia Safer AI for Children Coalition 'Artificial Intelligence and Child Protection: a collaborative approach to a safer future', 2024, page 7.

lii. Standford Cyber Policy Center Blog 'Investigation Finds AI Image Generation Models Trained on Child Abuse', 20 December 2023.

liii. Internet Watch Foundation 'Artificial Intelligence and the Production of Child Sexual Abuse Imagery' 2024 Report Update.

liv. See: https://mypicturesmatter.org/

lv. Ibid.

lvi. Ibid and as mentioned in ICMEC Australia Safer AI for Children Coalition 'Artificial Intelligence and Child Protection: a collaborative approach to a safer future', 2024, page 13.

lvii. BBC 'Illegal trade in AI child sexual abuse images exposed', Angus Crawford and Tony Smith, 29 June 2023.

lviii. Criminal Code Act 1995 (Cth).

lix. Missing Kids Blog 'Generative AI CSAM is CSAM', 3 November 2024.

lx. ICMEC Australia Safer AI for Children Coalition 'Artificial Intelligence and Child Protection: a collaborative approach to a safer future', 2024, page 15.

lxi. Ibid.

lxii. ICMEC Australia Safer AI for Children Coalition 'Artificial Intelligence and Child Protection: a collaborative approach to a safer future', 2024, pages 9 and 15.

lxiii. ICMEC Australia Safer AI for Children Coalition 'Artificial Intelligence and Child Protection: a collaborative approach to a safer future', 2024, page 15.

lxiv. eSafety Commissioner website: 'Generative AI and child safety: a convergence of innovation and exploitation', 11 June 2025.

lxv. Meta website: 'How Meta uses AI to better understand people's ages on our platforms', Facebook Tech 22 June 2022.

lxvi. Whatsapp website: 'How WhatsApp helps fight child exploitation' FAQ.

lxvii. WIRED 'Police caught one of the web's most dangerous paedophiles. Then everything went dark', Matt Burgess, 12 May 2021.

lxviii. Meta website: 'Hard questions: how we counter terrorism', About Facebook, 15 June 2017.

lxix. Ibid.

lxx. Ibid.

lxxi. UNODC 'Handbook on Children Recruited and Exploited by Terrorist and Violent Extremist Groups" The role of the Justice System', 2017 and EUROPOL 'European Union Terrorism Situation and Trend report 2024', 17 April 2025.

lxxii. X (then Twitter) blog: 'Serving healthy conversation' by Del Harvey and David Gasca, 15 May 2018.

lxxiii. Oversight Board 'Content Moderation in a new era for AI and Automation', September 2024.

lxxiv. Thorn, Protect your platform from hosting new CSAM, 2025.

lxxv. Proceedings of the Web Conference 2020 'Friend of Faux: Graph-Based early detection of fake accounts on social media', Adam Breur et al, 2020.

lxxvi. See for example https://www.activefence.com/solutions/harmful-content-detection/; https://www.unitary.ai/articles/introducing-unitarys-garm-plug-and-play-move-beyond-mere-object-recognition and https://www.unitary.ai/articles/what-are-multimodal-algorithms.

lxxvii. Unitary AI 'Do we still need humans to moderate online content', Tim Bernard, 2 November 2023.

lxxviii. UNICRI 'AI for Safer Children', 21 June 2025.

lxxix. UNICEF 'Policy Guidance on AI for Children,' November 2021, page 21.

lxxx. UNICEF 'Policy Guidance on AI for Children,' November 2021, page 21.

lxxxi. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024.

lxxxii. Department of Education 'Australian Framework for Generative Artificial Intelligence in Schools', endorsed in 2025.

lxxxiii. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024.

lxxxiv. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024, page 66.

lxxxv. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024, page 15.

lxxxvi. UNESCO 'ChatGPT and Artificial Intelligence in higher education: quick start guide' 2023.

lxxxvii. Harvard Center for Digital Thriving 'What teens are saying about AI – right now. Ten timely insights from our youth advisors to help adults stay in step', 2025, page 2.

lxxxviii. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024, page 68.

lxxxix. NSW Government, NSWEduChat, 2025

xc. NSW Government, NSWEduChat, 2025

xci. NSW Government, NSWEduChat, 2025

xcii. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024.

xciii. See https://www.khanmigo.ai/

xciv. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024, page 71.

xcv. See https://www.prweb.com/releases/introducing-tynker-copilot--the-first-ever-llm-powered-coding-companion-for-young-coders-301959002.html.

xcvi. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024, page 96.

xcvii. Parliament of Australia 'Study Buddy or Influencer: Inquiry into the use of generative artificial intelligence in the Australian education system', House of Representatives Standing Committee on Employment, Education and Training, August 2024, page 68.

xcviii. UNESCO 'ChatGPT and Artificial Intelligence in higher education: quick start guide' 2023.

xcix. Criminal Code Act 1995 (Cth).

c. By the Criminal Code Amendment (Deepfake Sexual Material) Act 2024.

ci. Section 474.17A of the Criminal Code Act 1995 (Cth).

cii. See OAIC 'Guidance on privacy and the use of commercially available AI products', updated 17 January 2025 and OAIC 'Guidance on privacy and developing and training generative AI models', updated 23 October 2024.

ciii. OAIC website: 'Children's Online Privacy Code', updated 22 April 2025.

civ. eSafety Commissioner website: 'Generative AI and child safety: A convergence of innovation and exploitation', 11 June 2025.

cv. Ibid.

cvi. eSafety Commissioner website: 'Industry codes and standards', updated May 2025.

cvii. eSafety Commissioner website: 'Generative AI and child safety: A convergence of innovation and exploitation', 11 June 2025.

cviii. See https://minister.infrastructure.gov.au/rowland/media-release/new-duty-care-obligations-platforms-will-keep-australians-safer-online

cix. Department of Education 'Australian Framework for Generative Artificial Intelligence in Schools', endorsed in 2025.

cx. Department of Industry, Science and Resources 'Introducing mandatory guardrails for AI in high-risk settings: proposals paper', September 2024.

cxi. Ibid.

cxii. See Regulation (EU) 2024/1689 EU AI Act, Annex III, point 3.

cxiii. MLex 'EU Commission eyes pausing AI Act's entry into application,' 5 June 2025 and Politico 'EU Commission opens door for 'targeted changes' to AI Act', 13 May 2025.

cxiv. European Commission, 'Commission work programme 2025', February 2025.

cxv. UK Department for Science, Innovation and Technology 'AI Opportunities Action Plan' 13 January 2025.

cxvi. See UK Data (Use and Access) Act 2025, s.81(2).

cxvii. See https://www.unicef.org/innocenti/projects/disrupting-harm

unicef
AUSTRALIA
for every child

Australian Committee for UNICEF Limited
ABN 35 060 581 437
PO Box 587, Pyrmont, NSW 2009
Phone: 1300 884 233 | Fax: 1300 780 522
Email: support@unicef.org.au | www.unicef.org.au