Coria

Addressing the Risks of Al-enabled CSAM and Explicit Content in Education

Includes insights from over 600 schools across North America, UK, Australia and New Zealand.

Essential reading for:

School / district and MAT leaders, pastoral staff, IT and anyone with a responsibility for, or an interest in, child digital safety.

qoria.com



Contents

3
4
5
8
9
10
25
26
28
30
31
31
31
32
33
34
35
36
42
43
43
44
45
46



Foreword

Tim Levy Managing Director, Qoria

In recent times, we've seen generative AI become an integral part of our lives. Tools like ChatGPT and Midjourney are ensuring powerful technologies, once reserved for specific sectors, are now accessible to the masses.

There is no doubt that AI is a game-changing technological enhancement that is, and will continue to affect all our lives significantly.

For schools, AI is transforming the classroom by offering teachers tools to enhance student engagement, streamline lesson planning and better support individual student needs.

It offers schools the long, sought-after opportunity to free up admin time and focus teachers more on instruction and student interaction - a welcome relief for these overburdened institutions.

However, whilst AI provides impressive benefits, it does present a darker side that each of us must carefully consider and prepare for. A quarter of children are using Al apps to do their schoolwork! It's critical therefore that stakeholders have access to insights and information to better understand how these tools are shaping students' learning, digital habits and safety - both inside and outside the classroom.

It is our duty to support that journey. Bringing challenges to the forefront in papers like this not only raises awareness but also encourages active dialogue between educators, parents and policymakers. And that's critical to Al tools being implemented thoughtfully, benefiting educational outcomes while protecting children's wellbeing.

Only by working hand in hand can we help make the online world a place where every child can safely explore, learn, and thrive. Because none of us is as powerful as all of us.





A message from Catherine Brown,

Chair of the Internet Watch Foundation (IWF)

"In the last few months, we have really seen AI image generators become much more sophisticated, to the point where almost anyone can make or manipulate life-like sexual or nude imagery of children at the click of a button.

So-called 'nudifying' apps have increased the ease with which this imagery can be made, and we have seen what happens when this technology is abused. The resultant imagery can be used to bully and distress children and young people. It can be used by predators to groom, entrap, and ensnare them into cycles of online sexual abuse.

About this report

This report puts a spotlight on the impact the abuse of this technology can have. It is vital reading for anyone who works with children and young people, and is an important asset in helping understand the evolving threats they face online."





Section 1

Executive Summary

While there is a wealth of information available on AI in education, much less is known on the risks around child sexual abuse material (CSAM), predation and associated strategies exacerbated by AI technologies.

This first-of-its-kind report explains the risk Al-enabled CSAM poses to students and the mitigation strategies schools can implement.

In June 2024 we asked schools around the world to tell us what they were seeing and how they were addressing this growing concern.

Our goal was twofold. Firstly, to better understand their challenges so we might find ways to provide more targeted support and guidance. Secondly, to foster a sense of shared experience - if there was one - so that no school would feel isolated in navigating this new and emerging landscape.

603 schools responded to our survey from across the United States, UK, Australia and New Zealand. Participants included superintendents, MAT leaders, principals, headteachers, IT leaders, and pastoral staff. And their settings varied in size from primary and secondary schools, to colleges, large school districts and multi-academy trusts.

What schools told us:

- Most respondents were concerned about the potential for adult perpetrators to use Al to groom their students.
- More than 30% were unfamiliar with common online grooming tactics perpetrators use.
- Around a third of US, UK and Australian respondents said they were now experiencing incidents of students possessing, sharing or requesting nude images every month - with the average age reported between 11-13 years.
- Approximately 12% of US and Australian respondents and approximately 20% of UK and New Zealand respondents said they were seeing incidents of students possessing, sharing or requesting nude images in children as young as 8-10 years.



Other insights:

- The AI tactics used to facilitate CSAM and explicit content sharing are growing exponentially, requiring schools to remain vigilant and continuously update their safety protocols.
- Increasingly, teachers and school staff are also finding themselves victims of image-based manipulation through the misuse of AI technology. It is an important step for schools to recognize this as a developing HR concern.



Key takeaways:

- The findings suggest a clear need for more support to help schools understand and address the risks around CSAM and explicit content sharing and to more effectively engage their parent communities.
- Early detection of risk is key. Many technologies that schools already possess, (such as filtering and monitoring), or may be aware of, (such as digital student check-in tools and others), can all greatly assist in the detection of risk, although greater education to appropriate staff about what to look for is likely needed.
- Technology solutions must be supported by ongoing education and collaboration.
- Cultivating an 'abuse-aware' culture across the school community where staff, students and parents work together to spot signs of harm can strengthen child safety.
- Schools should not feel helpless or alone in trying to deal with Al-enabled CSAM. While the challenges may, at times, seem daunting, there are many proactive and proven strategies contained in this report schools can adopt now to help them get out and ahead.





Section 2

Introduction

It's no secret that schools around the world are variously incorporating and grappling with the potential of Al.

While there is a wealth of information on the many educational benefits of AI, as well as risks such as cheating, privacy and data concerns, much less is known on the risks to student safety and wellbeing. Specifically, child sexual abuse material (CSAM), predation and associated strategies exacerbated by AI technologies.

Child sexual abuse material (CSAM) refers to any visual depiction of sexually explicit conduct involving a child. This includes images, videos, or any digital content that shows children engaged in sexual activities or poses.

Al-enabled CSAM is a real and growing risk and it comes against a backdrop of many schools and educators already feeling overwhelmed and unsure about how to navigate the evolving digital scape.

This first-of-its-kind report explains the risk Al-enabled CSAM poses to students and the mitigation strategies schools can implement.

It provides a roadmap for creating safe and supported environments where every student can thrive in their digital life, alongside the people and communities who care for them.

With insights from 603 schools, and from Qoria's own student safety experts, as well as other credible research, this report is a must-read for anyone responsible for, or interested in, child digital safety and wellbeing.

While the survey sample is not representative of the global schools population, it provides valuable insights into emerging trends and patterns, offering a fair indication of broader issues based on the data collected.



Thank you

We extend our gratitude to the schools who participated in our survey. Their open and honest contributions will provide valuable insight to other schools, as well as reassurance that they are not alone in their concerns or challenges surrounding Al-enabled CSAM and online safety in general.

Note: Most respondents chose to remain anonymous in exchange for their open and honest inputs. Their comments have been included throughout this report.



Section 3

AI, CSAM and Schools - Foundational Overview

Any organised use of AI in education is still in its early stages, and as with any emerging technology, schools are seeking to understand its application and value.

What we do know is that as AI evolves, so do the tactics of those who exploit it for harmful purposes, particularly in the area of CSAM and predation. This requires schools to be aware and vigilant.

By understanding how AI facilitates the spread of CSAM and child exploitation, schools can start to build a more protected environment, and equip their communities with the knowledge and tools to navigate this new landscape safely.





3.1 The AI-enabled CSAM challenge for schools

CSAM is a serious form of exploitation and abuse, and its production, distribution, possession, or viewing is a criminal offence in most countries. There are 2 key areas schools should be aware of:

(i) Grooming, solicitation and dissemination of CSAM

(ii) Child-generated CSAM

(i) Grooming, solicitation and dissemination of CSAM

What's the issue?

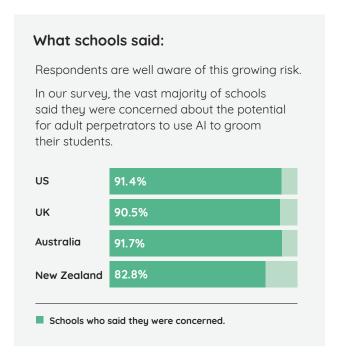
The proliferation of CSAM created and searched for by adult perpetrators has increased significantly in recent times.

A new report by the Childlight Global Child Safety Institute² found that 1 in 8 children (302 million young people) have experienced online child abuse and exploitation in the past 12 months.

In 2023, the Internet Watch Foundation found 20,254 Al-generated images³ within a single month on a dark web forum dedicated to CSAM, causing them to declare 2023 as the most significant year on record for the sharing of child abuse material.

Perpetrator processes and AI enablement

Perpetrators are intentionally interacting with AI to personalize their approaches and manipulate children's immature development and psychology to improve their chances of success.





Grooming methods

Understanding how perpetrators are using AI to target children can assist schools in recognizing the early warning signs and strengthen their response and prevention efforts.

There are several stages to grooming:

(1) Targeting the victim

Perpetrators often target children who appear vulnerable or lack supervision. This can include those with indicators of low self-esteem, insecurity, or a lack of attachment to their families.

Children who are vulnerable or suffering from mental health⁴ concerns are at particular risk.

E2E (end-to-end encryption) services, including Snapchat, WhatsApp, Signal and Telegram are common platforms perpetrators use.

How is AI an enabler?

Al algorithms support perpetrators by analysing vast amounts of data to:

- Identify vulnerable children with greater precision.
- Detect patterns of behavior, interests, sentiment and emotional states to use when approaching a child to discredit them or people close to them.

Perpetrators are versatile in the ways they manipulate algorithms to target children. A past example was the use of the hashtag #dropboxlinks which was used as a (now defunct) network for predators on Instagram to find and share explicit photos of underage children.

(2) Gaining trust

Abusers pose as friends, celebrities or other people to pique the interest of children and ask questions to understand their home life and situation. They often offer gifts or rewards to gain the child's trust.

How is AI an enabler?

- Abusers often use generative AI to create fake and convincing personas, and ensure they are tailored to appear realistic, relatable, and trustworthy.
- Al can generate fake images or certifications for example to convince a child that someone is really who they say they are. Examples include the creation of certifications, stories or websites to convince a child they are credible.

(3) Filling a need

Perpetrators often strive to fill a need in the child's life by providing emotional support, gifts, or rewards. This is designed to create a sense of dependence and loyalty to the perpetrator. They may also simultaneously discredit the child's support circle, parents or friends. They may also try and compromise the child's relationships with others to solidify a close attachment between them and the child.

How is AI an enabler?

- Al tools can be used to create fictitious information that discredits people close to the child.
- Perpetrators also use AI to begin generating and sharing explicit or pornographic content with the child to desensitize them toward sexual content generally. This is an attempt to "normalize" further stages of the grooming process down the track.



Grooming methods, continued.

(4) Isolating the child

Perpetrators may discourage contact with others, further alienating the child from their parents or friends. This can make the child feel trapped and unable to reach out for help or support.

How is AI an enabler?

- Similar to the ease with which investment scams work, perpetrators can use AI to direct targetted children to spoofed websites and use misleading information to undermine the advice, support or conversations children may have with their support networks, eroding trust.
- Alternatively, they may threaten to share Al-created content that could embarrass the child.

(5) Sexualising the relationship

As previously mentioned, perpetrators may expose children to explicit material, normalizing sex, and sexual requests. A child's natural curiosity is exploited so that when the abuse is actually initiated, it is less shocking. The perpetrator may reinforce the message that this activity is what the child wants, making the child appear to be the one who initiated this process through coercion.

This can be subtle and gradual, making it difficult for the child to recognize the signals of abuse.

How is AI an enabler?

 Abusers may freely and colloquially speak about or share sexually explicit material with the child through Al tools like character.ai, or apps like hotify, or using coded language such as:

NP4NP - Naked pic for naked pic

LMP - Like my pic

Snacc - A person you find attractive

Rule 34 - Any topic can be made into

a pornographic topic

GNOC - Get naked on camera

TDTM - Talk dirty to me

• Abusers may invite the child to explore nudification tools together such as clothoff.ai or nudifier.ai.

(6) Maintaining control

Perpetrators often employ tactics like confusion, threats, and secrecy to maintain control. They may claim they have a "special relationship" or withdraw or threaten the child's safety.

How is AI an enabler?

 Threats to publicly expose private information, such as personal or embarrassing photos, or private conversations, is a primary strategy. Tools include:

Doxxing: Al algorithms scour the internet and databases to compile personal information about people to intimidate, threaten or harm a person. This information is then shared online to cause distress or harm.

Deepfakes: These images and videos can be used to fabricate evidence of wrongdoing or to discredit individuals, turning others against them.



Grooming methods, continued.

What schools said:

We asked schools if they had observed a rise in incidents of students engaging with adult strangers online. The findings indicate a very real and present concern. Equally significant is the number of schools who are unsure if any students are engaged in this behavior, underlining the need for increased visibility.

	UK	US	AU	NZ
Yes	23.3%	20%	16.7%	3.1%
No	52%	42.9%	29.2%	58.8%
Unsure / Don't know	24.7%	37.1%	54.2%	28.1%



How schools can stay vigilant to signs of grooming

Students may exhibit noticeable behavioral changes as a result of this grooming process. Any, or all of these changes, do not definitively indicate a child is being groomed. But they are warning signs that should not be ignored and recognizing them early can be essential in preventing escalation of potential harm.

Possible indicators can include:

Changes in behavior and mood

- Increased secrecy around online activity, such as guarding their devices, closing screens when staff or other students are near, or becoming defensive about who they're speaking with.
- Mood swings that are sudden or unexplained, including increased anxiety, irritability, or withdrawal from favorite teachers, friends, family, and normal activities.
- Signs of fear or distress when receiving messages or notifications, or reluctance to discuss their online interactions.

Social withdrawal and changes in relationships

- Distancing from friends and trusted adults and an increase in secrecy about their social interactions.
- Isolation from usual peer groups and reluctance to participate in group activities.
- Attachment to a new, possibly unknown person who may be described vaguely or kept secret, especially if this individual seems to be influencing the student's actions or perspectives.

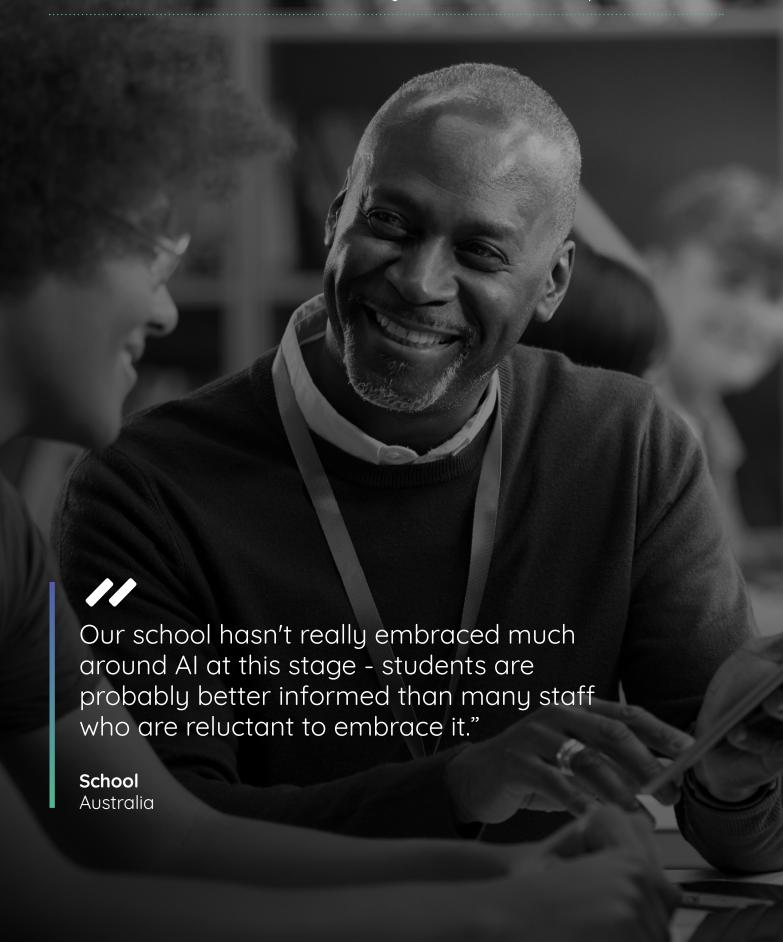
Signs of gifts or new possessions

- Unexplained new items (such as gifts, clothing, or devices) that the child cannot or will not explain.
 Perpetrators sometimes send gifts as a way to build trust and to control.
- Requests for money or access to a credit card, which may be linked to requests from the perpetrator.

Changes in academic performance and engagement

- Sudden drop in academic performance or a lack of interest in class, which may indicate distraction or emotional distress.
- Skipping school or missing classes more frequently, possibly to engage in online conversations or meet someone in person.
- Lack of concentration or increased fatigue, which could stem from staying up late engaging with someone online.









Perpetrator AI tooling

Perpetrators are increasingly using AI tools to manipulate, create, and disseminate harmful content, including deepfakes and synthetic media. This makes it easier to groom, exploit, and target vulnerable children:

By understanding these tooling methodologies schools are better places to spot their use and initiate targeted digital safety education for students to do the same. Examples include:

Fake profile generators:

Al tools are now being used to create sophisticated fake profiles with convincing photos, bios, and a whole history of posts and activity on the platform. Perpetrators use these fake profiles to masquerade as someone else and deceive children into trusting them.

Nudification tools:

Nudification tools are Al-powered apps that can digitally remove clothing from images, creating simulated nude versions of the subjects. These tools use advanced deep learning algorithms to analyze the clothing in a photo and then digitally erase it, producing an edited image that appears to show the person nude.

The legality of these tools is a grey area in the global legislative system. While not permitted on app stores, users can still encounter ads for these tools on platforms like Instagram and Facebook and may continue accessibility through site URLs.

It is important to note that these tools vary in sophistication and realism, making it difficult for both parents and schools to determine whether images or videos depict a real young person, or whether they are Al-generated.



I have seen firsthand the impact on our female students of being sent deepfake nudes derived from their social media profiles, by male students they counted as friends (hiding behind the anonymity of AI tools)."

School

United Kingdom



Perpetrator AI tooling - cont.

Customized digital entities:

Experimentation is emerging with AI bots that leverage children's interest in creating artificial or bespoke virtual girlfriends⁵ or boyfriends, rather than engaging in or pursuing real-world relationships.

Chatbots:

Certain Al-powered chatbots, such as Character.ai, can simulate human-like conversations and respond to messages in real-time. Perpetrators can use chatbots to automate interactions with multiple children simultaneously, allowing them to scale their efforts and target a larger number of victims at once.

Deepfake apps and technology:

Deepfake technology enables the creation of highly realistic fake videos and images by manipulating existing footage or photos. Perpetrators use deepfake technology to create pornographic material featuring children or to impersonate someone the child knows and trusts.

Voice changers and cloning tools:

Voice changer apps and software can modify the pitch and tone of a person's voice in real-time. Perpetrators may use voice changers to disguise their voices during phone calls or video chats, making it harder for children to recognize them or detect their true intentions.



With older students gaming has led to a couple of threats of sexploitation - students feel ashamed rather than believing themselves to be a victim, it's so difficult to support..."

School

United Kingdom

What schools said:

We asked schools if they were familiar with some of the more common ways Al can be exploited to target children.

	UK	US	AU	NZ
Training Al to act as a child	59.7%	37.1%	47.9%	37.5%
Using Al to create more convincing fake profiles	74.8%	65.7%	66.7%	53.1%
Using Al chatbots to contact more children at once	61.1%	40%	50%	42.2%
Generating and sharing scripts with other predators	45.5%	42.9%	39.6%	26.6%
Not familiar with any of these	20.5%	31.4%	29.2%	40.6%

The findings suggest that more education is needed in schools on what are becoming commonplace tools for those wishing to harm children.



3.1 What is the CSAM Risk for Schools? Continued.

(ii) Child-generated explicit content

What happens when children themselves become the initiators?

There is a growing trend among young people to utilise AI to create and share both real and manipulated explicit content including non-consensual nude images.

As students spend more time online, often unsupervised, and continue to utilise digital devices and tools for learning and creativity, their exposure to explicit material and subsequent participation in peer-to-peer exploitation has escalated. (See also 'Factors driving CSAM sharing by children' - page 25.)

What schools said:

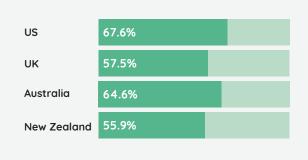
When asked about the prevalence of students possessing, sharing or requesting nude content, a relative low number of respondents said they were experiencing incidents at least once a week. A considerably higher number said they were experiencing an incident at least every 4 months. Further research over time will suggest if this is an upward trend.



It's also happening at a much younger age than people might assume.

When asked about the ages of students possessing, sharing or requesting nude content, respondents shared that this was occurring most commonly with children aged 11-13 years of age.

And 21% of respondents in the UK said they were seeing this behavior in students even as young as 8-10 years old.



Reported incidents of possessing, sharing or requesting nude content in children aged 11-13.





Schools' responses do not state that all instances involved manipulated and/or non-consensual imagery. And it's important to acknowledge that children exploring and expressing their identity, including sharing images online, can be a natural part of adolescent development as they seek validation and social acceptance.

But a growth in nude image sharing can lead to a normalization of exposure to explicit content, the desensitizing impact of which can reduce a child's ability to recognize abusive situations or harmful behaviors.

When asked about their experiences of students using technology to create fake, sexually graphic images, schools hadn't experienced a huge volume of incidents, but it was still significant.

The rise of such incidents may indicate deeper issues around peer dynamics. It may reflect a growing culture of disrespect or a lack of empathy among some students, who view such actions as pranks without recognizing their damaging impact.

The prevalence of these incidents combined with the lack of awareness amongst schools shows the need for more robust visibility of digital activity and educating students on the dangers of creating or distributing harmful content.

What schools said:

A small but still significant number of respondents said they had experienced at least one incident in the last 12 months of students using technology to create a fake sexually graphic image of a classmate.

US 5.9%

UK 9.6%

Australia 10.4%

New Zealand 6.3%

Had seen at least one incident of sexually graphic image creation from students in the last 12 months.



What schools said:

We asked schools what the most common platforms or media were that students were using to request or share nude content.

	UK	US	AU	NZ
Snapchat	49.2%	75.8%	47.9%	21.9%
Instagram	5.6%	24.2%	6.3%	-
Discord	2.3%	9.1%	12.5%	-
Text and Messages	15.6%	45.5%	10.4%	9.4%
Online Games	5.4%	9.1%	10.4%	-
Email	Negligible	18.2%	2.1%	4.7%

Staff education has been high on our agenda as the inevitable use of AI is becoming more of a reality. We know we need to be ahead of the game in terms of how we support and educate our students whilst keeping them, and us, safe."

SchoolUnited Kingdom



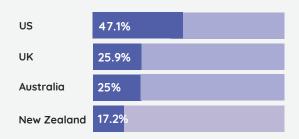
Al tooling used by children

Evidence from industry confirms that AI tooling is becoming increasingly prevalent in young people's lives, with nudification tools, for example, attracting over 24 million visitors to websites globally each month. There is an increasing volume of news stories involving young people and explicit content. In Australia, a recent incident at a school involved a teen being arrested for creating nude images of 50 female classmates utilising Al tools⁶, and similar incidents have been reported in the United States, Spain and elsewhere.

What schools said:

When asked if they had identified any instances in the past year where students specifically used AI apps or tools to create CSAM or nude content, such as deepfakes, many schools said they hadn't heard of it.

This raises important questions. Is AI tool usage not prevalent among younger people? Or is it taking place on kids' personal devices, where schools, and parents without parental controls on their child's devices, can't see it?



 Number of schools who weren't aware if students were using technology to create a fake sexually graphic image of a classmate.

When staff become victims

Some school staff members have also become victims of image-based abuse through the misuse of Al technology.

Reports have surfaced globally of students filming staff or doctoring images without consent, then using AI to manipulate that content to generate deepfakes or fictitious online accounts that are then utilised to bully, defame, or humiliate their targets.

While student can often do this as a joke with no malicious intent, this behavior can, and has had, severe consequences for victims. This includes emotional distress, damage to reputation and severe psychological injuries.

Generating deepfakes has become widely accessible, affordable and utilised by students.

To design a deepfake of their teacher students can simply:

1. Access Al-powered deepfake tools, often available online or through mobile apps, to create manipulated images or videos of school staff.

- 2. Students then freely access publicly available images and videos of school staff. Staff may believe that benign pictures are harmless to post and very low risk, but these are easily doctored. School staff should remain focused and vigilant about privatizing their accounts and personal information.
- 3. Use tools that allow users to swap faces, superimpose images, or create entirely new content that appears realistic
- **4.** Share the manipulated content through social media, messaging apps, or online platforms, inciting jokes and ridicule toward the targeted staff member.

The spread of these deepfakes can be rapid, making it challenging for the victim to address the situation before it circulates out of control and causes significant harm.

It is an important step for schools to recognize this as a developing HR concern. Like all employers, they have a legal duty of care to do all they reasonably can to support the mental health, safety, and wellbeing of their employees. This includes taking steps to identify and mitigate risks that could impact staff mental health.

(See page 31, section 5.2. 'Review and update school policies')





3.1 What is the CSAM Risk for Schools? cont.

(iii) Factors driving CSAM sharing by children

Whether real or Al-generated, the rise of young people sharing explicit content in online environments is a concern.

There are a number of underlying issues that contribute to the misuse of technology and young people's willingness to participate in this type of risk-taking behavior. Understanding these influencers can help schools better understand which students may be more vulnerable.



Isolation & Ioneliness among adolescents

Loneliness and lack of social connection can make young people more likely to engage in online interactions that offer a sense of companionship or belonging. They may also have limited support networks to notice signs of abuse or intervene.



The challenge lies with the vulnerable students who want to feel special; they are easy targets and are very reluctant to share issues. The technology is so advanced that monitoring them is not enough."

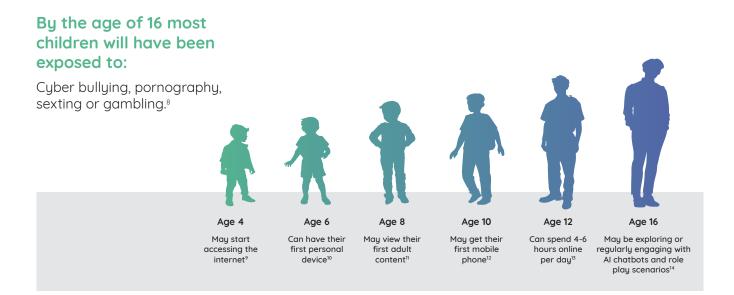
School leaderAustralia School



Pornography

The link between young people viewing pornography and the creation or solicitation of explicit content is a complex issue. Research⁷ suggests that exposure to pornography can negatively influence young people's attitudes and behaviors, potentially leading to desensitization and the normalization of harmful sexual behaviors.

Pornography often portrays unhealthy sexual behaviors, which shape young people's 'sexual scripts', and ideas of what is considered normal. Repeated exposure can lead to a distorted view of consent, boundaries, and healthy sexual interactions leading to acts like sharing nudes as a form of sexual expression.



NMC (Not My Child) Syndrome

'Not My Child' Syndrome refers to the tendency of some parents to believe that their children are not susceptible to the risks and dangers present in the online world. This is despite evidence suggesting otherwise.

70% of U.S. respondents to our survey said one of their most significant challenges in addressing issues of Al, CSAM, and explicit content is "lack of awareness among parents".

The NMC mindset can lead parents to:

- Underestimate the risks their children encounter online.
- Not monitor or supervise their children's online activities adequately.
- Not educate their children about online safety and responsible technology use.
- Dismiss warning signs or concerning online behaviors exhibited by their children.

While engaging parents in online safety discussions can be difficult, schools should persist and aim to innovate when it comes to the engagement and enablement of parents in safeguarding their children.

See 'Education for parents' on page 35 for more.



Section 4

School Response Impediments

We asked our schools about their biggest challenges in addressing the issues of AI, CSAM and the sharing of explicit content.

What schools said:

	UK	US	AU	NZ
Limited staff training, knowledge and time	63.6%	79.4%	31%	71.9%
Difficulty detecting and monitoring harmful behavior	50.3%	73.5%	27.1%	46.9%
Lack of awareness among parents	77.6%	70.6%	22.9%	60.9%
Student resistance to education	30.5%	35.3%	6.3%	18.8%
Limited budget	24.2%	50%	6.3%	25%





4.1 Limited staff knowledge, training and time

Students are more likely to seek help or report incidents if they believe school staff, or parents, will understand and resolve the issue in ways they view as appropriate.

An environment where staff and educators are visibly well-versed in digital trends and technologies significantly impacts open communication and trust.

A 'head in the sand' approach is understandable. But it could compromise some of the gargantuan strides schools have made in student digital safety and wellbeing over the past few years.

Missing the early signs of AI misuse mean that harmful behaviors can go unnoticed, allowing them to proliferate. As a result, students may feel confident in continuing their activities, believing that there are limited or no repercussions.

Recent research in the US¹⁵ has shown that 60% of teenagers either attend a school that has no rules around the use of generative AI or are unsure if their school even has any rules.

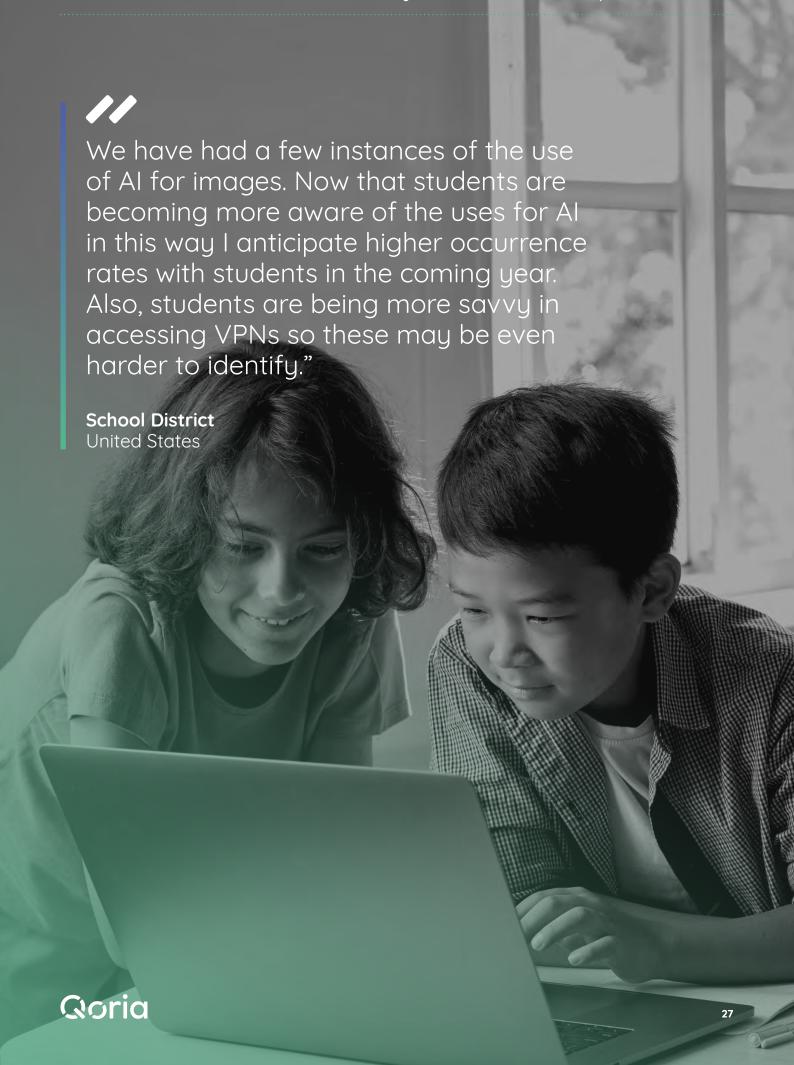
When schools are unable to prioritize educating students on the responsible use of Al and the dangers of explicit content generally, students are left to navigate these complex issues on their own. This can lead to increased experimentation and misuse.



It's a very new area we feel we are sometimes navigating on our own."

School Australia





4.2 Lack of awareness among parents

Without parental involvement in reinforcing online safety, schools may struggle to address the growing threat of CSAM and protect students from exploitation.

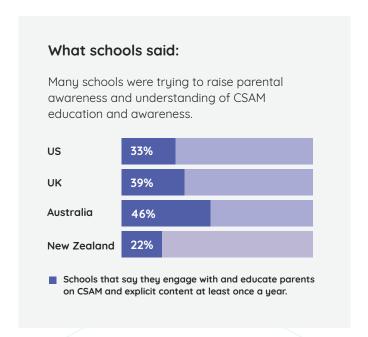
Increasing and/or optimizing parental education can be an ongoing challenge for schools. But it's worthy of the sustained effort, given the additional level of vigilance it provides to children.

Key to this endeavor is understanding the reasons for any disengagement.

- Do parents believe that risks reside only in the physical world?
- Do they mistakenly believe that school filtering systems, government regulations, or standard parental controls are sufficient to protect their children from CSAM, leaving them less vigilant about monitoring online behavior?
- Is the stigma and discomfort of talking about it leading parents to avoid the topic altogether?
- Are they more focused on visible, immediate online dangers like screen time, gaming, and social media addiction, without realizing that CSAM poses a hidden but severe risk to their children's safety?

Addressing these challenges may require schools to rethink their approach to parental communication possibly by using more engaging methods and by addressing misunderstandings and concerns in order to resonate better.

One thing's for sure, digital parenting in a tech-driven and challenging world can be overwhelming, making it harder for schools to address specific threats like CSAM among the many other daily concerns they face.







4.3 Detecting and monitoring harmful behavior

Early detection allows schools to take swift action to safeguard vulnerable students. It also provides vital evidence to support collaboration with law enforcement, ensuring that perpetrators are held accountable and that affected students receive appropriate support.

Schools relying on singular technologies such as filtering only, or monitor using eyes and ears alone, have a more limited view of risk compared to those using a broader range of methods. This means harmful interactions or illegal content may go undetected for prolonged periods.

If technologies are outdated or insufficiently advanced, schools will be similarly hindered and unable to detect code words, slang, or hidden language students use to discuss or share harmful material.

Outdated filters will also overblock, and hinder learning. Or fail to detect harmful or illegal imagery and other content on the school drive. All is cutting edge, so detection of what is happening within these technologies can be difficult for many basic filters.

In reality, help may be closer to hand than some schools know and an audit of their digital safety technology stack can uncover existing but unrealised capabilities, alongside genuine gaps in provision.

(See page 36, section 5.7 'Increase digital risk visibility with technology'.)

What schools said: Schools cited that difficulty in detecting harmful behavior was their most significant challenge when addressing the issues of AI, CSAM and the sharing of explicit content amongst students. US 73.5% 50.3% UK Australia 27.1% 46.3% **New Zealand** Number of schools who said that difficulty detecting harmful behavior was the most significant challenge in addressing the issues.



Section 5

Response Strategies for Schools

Schools should not feel helpless in the face of this changing landscape.

While the challenges posed by Al-enabled CSAM may seem daunting, there are many proactive and proven strategies that can help.

Updated school policies, technological solutions, educational programs, and collaborative efforts, combined, can help schools get ahead and create safer environments for their students.

5.1 Develop an Al working party

Some schools tell us they have or are in the process of setting up an AI working party. This is an important first step in that it can align all stakeholders around common understanding and strategies as well as support resources for staff who are victims of deepfake bullying and defamation, including robust incident follow-up plans.

Regular meetings and discussions will also improve communication, empower staff, enable shared responsibilities, as well as promote professional development in Al, risk management, and response.



Our school is lucky because we have an Al department... a student Al policy, and a staff Al Policy and they're helping us a lot."

College

Australia

5.2 Review and update school policies

Schools should prioritize updating their policies and incident management procedures to address the growing risks associated with Al-based incidents, with a particular focus on victim support.

They should give clear guidelines for safeguarding students and staff, providing timely support to those affected by Al-related issues as well as consequences for perpetrators. Schools can assign this responsibility to an Al working group or manage it separately under broader risk management and emergency response strategies.



We are reviewing our policy and working with our Community Police Officer. She is coming into talk to parents next week."

School

New Zealand



5.3 Staff training

Investing in ongoing professional development and providing the right tools for staff is essential to creating a safer and more effective digital learning environment.

Leaders can take concrete steps by scheduling regular training sessions that cover not only the latest digital trends but also the deeper psychological and social factors driving poor online behavior among students

This ensures that staff are better equipped to recognize, and address such issues early on.

In addition, holding policy workshops on intervention and support strategies - specifically tied to the digital tools they use - can empower staff to respond more effectively when problems arise. Including student leaders in certain aspects of this policy-making can help ensure the student voice/perspective is captured.



Common strategies schools can use to enhance staff knowledge and development

Address budget constraints: Schools can address concerns and challenges around limited budgets or resources, by pooling funds, working in clusters, or even creating inter-school peer-to-peer learning initiatives and groups to share learning and development opportunities, while simultaneously benefitting from increased collegiality and community.

Encourage student-led initiatives: Engage students who are passionate about technology to lead workshops or training sessions for teachers on Al. This not only empowers students but also provides staff with fresh perspectives on technology use, and addresses the important aspect of student voice in education outcomes.

Conduct AI app of the week workshops: Organize workshops where staff can share their expertise or

insights on specific apps or platforms. This fosters collaboration and helps staff learn from each other in a safe and supportive environment.

Leverage open educational resources: Utilize free or low-cost online courses and resources focused on AI and digital literacy. Websites like Coursera, edX, and Khan Academy offer valuable content that can be accessed at little to no cost.

Form partnerships with local universities: Collaborate with local universities or colleges that have strong technology programs. These institutions may offer guest lectures, workshops, or mentorship programs led by students or faculty who are knowledgeable in Al and technology.

Create a technology task force: Form a small group of enthusiastic staff members to explore technology advancements and present their findings to the larger school community. This group can focus on identifying useful tools and strategies for integrating technology into an educational context.



5.4 Education for parents

Measures put in place to support and protect students and staff must be extended to parents and guardians so they are also able to provide support and protection in the home environment where the majority of such concerning behavior is taking place.

Surveys to understand families' concerns and issues are key. They can be used to drive the development of high quality training.

A review of overall needs can be a helpful way to start.

A comprehensive strategy that includes consistent communication, workshops, and educational resources can help to build awareness, resulting in a more informed and vigilant community.

Questions to ask

- Q: How aware are our parents of digital safety risks including those posed by AI and CSAM?
- Q: What are the challenges they face in learning about the issues?
- Q: To what extent are we supporting them with knowledge and education programs? Are they sufficient or frequent enough?
- Q: Is the information we are sharing up to date?
- Q: Do we have the ability to give our parents visibility about their child's digital activity?

Common strategies schools can use to enhance parent knowledge and development

Provide regular resources and communication channels: Resources that provide helpful, 'snackable' content through centrally located online platforms (such as Qoria's Online Safety Hubs for parents and staff) have proven both popular and effective.

Regularly communicating the availability of this central platform is key, along with the establishment of additional open communication channels, which allows for interactivity and advice sharing, as well as provides agency and a sense of ownership for parents over their child's online world.

Update strategies: Use parent year group communication channels to share information directly with parents, from class teachers or leaders. Content such as AI App of the Week overviews, scam alerts, or even incident management tips, can really help reinforce effective and consistent safety messaging, especially if there has been a high-profile incident in the media students may be exposed to.

Family technology nights: Host events where families can come together to learn about technology and its impact on education. These nights can include demonstrations, discussions, and activities designed to engage both parents and students.

Host parent workshops: Organize workshops or information sessions focused on AI, CSAM, avoiding exploitation and how to enhance digital safety. These can also deep dive into cross-over topics such as online privacy, identifying misinformation, and understanding the implications of technology on children's education and wellbeing.

Collaborate with experts: The digital world is evolving quickly, and it is unrealistic to expect teachers to be able to provide all the answers and education. Schools can invite guest speakers or experts in child digital safety, child psychology, or law enforcement to speak to parents about online safety.

These professionals can provide diverse and valuable insights, real-life examples, and actionable advice to help parents understand the risks and equip them with effective strategies to safeguard their children.



5.5 Education for students

It is equally important to educate students on the implications, risks and benefits of AI for several reasons.

Al powers the digital environments students

use daily: From social media platforms to online games, Al-driven algorithms shape much of students' digital lives. Educating them about how these systems work helps students understand how they can proactively influence content exposure, personal data collection, and interaction with others online. This awareness is vital for helping them protect their privacy and make informed choices.

Al tools are being exploited by perpetrators:

As mentioned previously Al technologies, such as deepfakes, nudification tools, and voice cloning, are increasingly being used by cybercriminals to create fake content, manipulate identities, and engage in harmful behavior like online grooming. Educating students about these Al risks in age appropriate ways can help them recognize warning signs and avoid falling victim to exploitation or manipulation.

Understanding AI helps students build digital resilience:

Al literacy is essential for building digital resilience, helping students navigate Al-powered environments safely, from social media to online learning platforms. This education allows for a proactive mindset in students, teaching them how to protect themselves from risks like algorithmic manipulation, data breaches, and unwanted surveillance.

Strategies schools can use to teach Al citizenship

Use real-world case studies: Integrate real-world case studies into the curriculum that highlight incidents of CSAM and exploitation. Discussing these cases can help students understand the seriousness of these issues, and recognize the signs of exploitation in their own lives.

Use pre-existing peer-to-peer education programs:

Integrate the topic of Al-based CSAM and explicit content into pre-existing peer-to-peer or mentoring programs where students can lead appropriate discussions and workshops on CSAM and online safety. Peer-to-peer education can create a more relatable and supportive learning environment, encouraging open conversations about sensitive topics.

Consider art and media projects: Encourage students to create art, videos, or presentations that address the themes of digital safety, consent, exploitation and the impact of CSAM. This creative approach allows students to express their understanding and feelings about the subject while raising awareness across the school community through exhibitions or launch nights.

Gamified learning: Develop educational games or online quizzes that test students' knowledge about digital safety, including recognizing CSAM and understanding exploitation tactics. Gamification can make learning more engaging and memorable.

Questions to ask

- Q: Do we currently provide Al-specific education in our curriculum for students?
- Q: Do we teach students about Al's benefits as well as its risks?
- Q: How are we preparing students to critically assess Al-driven content, such as Al-generated fake news or misleading information?
- Q: Are we able to provide up-to-date education for students on emerging risks and trends relating to Al?
- **Q**: Are we providing dynamic and interesting formats in our curriculum to engage students in Al education?





5.6 Parental control tools

Anecdotally, schools have shared with us experiences of parents monitoring their children's activity at home, if at all, being heavily focused on screen time, rather than what their child is actually accessing.

This may be down to a lack of understanding of risk, or a lack of familiarity with the technology that can give them a deeper visibility.

Given the possibility that young people may encounter explicit content intentionally or unintentionally at home, schools might consider the opportunity to explore safeguarding tools, such as parental control apps, that not only allow parents to manage screen time, but also tailor and filter content outside of school, create reports of their child's online activity, and have the functionality to block age-inappropriate content.

Parental control tools, such as the Qustodio by Qoria app, also allow parents to filter and block inappropriate apps, games, and websites. They help ensure children access age-appropriate, educational, and child-friendly platforms while automatically restricting access to sites that may contain harmful material such as mature content, gambling, or violence.

These tools not only protect children but gradually build awareness for parents about the realities of their child's online activity.

When parents are able to view accurate, data-based insights it can create valuable opportunities for learning and communication between themselves and their children. This helps encourage regular discussions about online safety within the household, while reinforcing the same at school.

Questions to ask

- Q: Do we provide parents with safeguarding options like parental controls that are flexible for a child's individual age, maturity or individual needs?
- Q: Do these tools allow parents to modify or update screen time routines easily if their child needs unexpected access? For example, if a child is at home sick?
- Q: Do these tools allow for flexible reporting options?
- Q: Is the information provided for parents clear and useful?
- Q: Does our school's web filter provider offer these apps?



Parents seem to monitor time spent, not actually what their child is able to view/access."

School

United Kingdom



5.7 Increase digital risk visibility with technology

Schools are realising that eyes and ears are no longer enough to spot students at risk online. This realization, compounded by the fact that perpetrators are now leveraging AI to refine their strategies, necessitates a closer look at technologies capable of improving visibility of risk.

Digital monitoring

Digital risk visibility solutions, such as digital monitoring, are becoming widely used by schools around the world. In the case of UK schools, the requirement to detect digital risks early through some form of monitoring is now a government requirement.

Digital monitoring solutions are often human moderated and provide real-time insights into potential risks, such as exposure to harmful content, inappropriate conversations and online grooming.

They can also provide early markers of potential issues or concerning incidents before they happen. They often include granular risk categorizations as well as contextual based notifications to pastoral staff members within minutes of a risk being detected, enabling a swift and accurate response.

Staff may feel more confident and protected when visibility is increased. Issues prevented from escalation, reduce the need for longer term interventions, saving them time and stress.

Questions to ask

- Q: Would we know if a student said, did, or shared, something on their school device related to CSAM or grooming?
- Q: Could we identify inappropriate conduct or conversations students may be having with perpetrators?
- Q: Would we know this information within minutes of it occurring?
- Q: Does our current monitoring solution minimise false positives and save staff time?
- **Q:** Can our current monitoring systems detect student colloquialisms and coded language used in online grooming?
- Q: Does our monitoring solution look for risk everywhere a students goes in the digital space not just behind the web browser?

Answering no to any of these questions may indicate a gap in provision and a lower level of protection than thought.

Upgrading to more robust and contemporary solutions could significantly enhance schools' ability to safeguard students against exposure to CSAM and grooming.

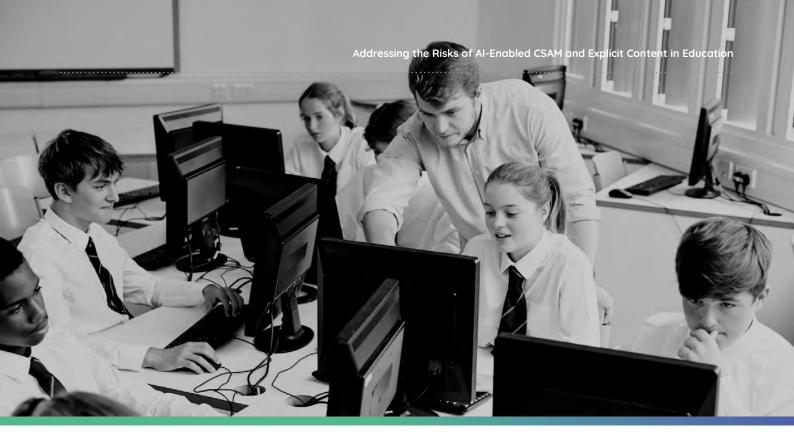


We'd like to be on the front foot...
we also believe that more is likely
happening that we don't know about."

School

Australia





Content filtering

Content filtering solutions are designed to prevent students from accessing online content that could be illegal or harmful to either their physical, mental, or academic wellbeing. Filtering solutions are already integrated into almost every school around the world, and are a mandatory requirement in the US thanks to CIPA (The Children's Internet Protection Act), and in the UK, thanks to KSCIE (Keeping Children Safe in Education).

But not all filters are equal.

Basic content filters, often provided for free, are commonly designed for workplace environments, not education settings. They are usually blunt instruments that block huge swathes of the internet and with it valuable learning resources.

A robust education filter is a flexible tool with granular controls that works in real-time.

Real-time allows for content to be analyzed and appropriately controlled as it is delivered to the user. This eliminates the gap between harmful content going live and the filter's ability to block it - and is critical in the immediate nature of Al-enabled content creation.

Non real-time filters can expose children to harmful content relying as they do, on predefined categorisations of particular domains and URLs. Such is the importance of real-time filtering, in October 2024, the UK's Department for Education required that schools check their filters for real-time capability.

Questions to ask

- Q: Can we adjust our filtering rules to prevent access to problematic Al apps, websites or tools associated with CSAM?
- Q: Can we create rules within our filter that allow us to respond directly to observed behaviors for an individual student that may indicate a risk of exploitation?
- Q: Can our filtering rules delineate between allowing appropriate curriculum exploration (e.g. personal development, health or sex education) and effectively detecting concerning searches or the use of Al apps that enable the development of sexual content?
- **Q**: Are the right reports/alerts reaching the most appropriate staff members in school?
- Q: Can our filtering system effectively block websites or identify coded language often associated with grooming or exploitation?
- Q: Does our filter block harmful content in real-time?

If you answered 'no' to any question above your filter solution may have gaps and may not be protecting students as well as you thought.

Upgrading to a more robust solution could significantly enhance your school's ability to safeguard students against exposure to CSAM and grooming.



Classroom management

Classroom management technology is becoming increasingly common in schools, providing educators with vital tools to monitor student online activities and mitigate risks associated with CSAM.

These technologies empower teachers to observe and manage what students are doing online, reducing digital distractions while enhancing learning.

They are particularly effective in identifying and addressing harmful online behaviors, allowing for timely intervention.

Key features often include screen sharing to promote collaborative learning, controls for teachers to adapt the school's Internet access policy within the classroom and the ability to observe students' screens to give them support or keep them on-task.

By integrating these tools, schools can create a safer digital environment, supporting both educational engagement and the protection of students from potential online threats.

Questions to ask

- **Q:** Do our teachers have the ability to see what students are doing online, both during class and retrospectively after the class has ended?
- Q: Do our teachers have the ability to determine what can and can't be accessed in their lessons when students are online?
- Q: Do they have the ability to tailor or support each individual learner and their specific online needs in the classroom?

If you can't answer yes to every question above your classroom tools may not be providing an adequate level of protection in the classroom.





The rapid advancement of AI makes it difficult for schools to keep up with monitoring and detecting explicit content generated. Schools are therefore investing and training staff to identify and address such issues promptly."

Superintendent United States

Student check-in tools

Students who have been exposed to CSAM or any form of risk often remain silent. This silence can stem from feelings of guilt or fear, a lack of vocabulary to express their experiences, or simply a lack awareness that their situation is problematic.

These barriers can prevent students from seeking help, leaving them vulnerable and isolated in their struggles.

Recognizing this critical issue, many schools are moving away from traditional annual paper surveys to adopt AI-powered check-in tools designed to monitor student wellbeing more frequently, often on a weekly basis.

This shift not only promotes a more proactive approach to mental health and safety but also fosters a culture of openness and support within the school community.

By implementing these innovative tools, schools can gain real-time insights into student emotions and experiences, allowing them to identify at-risk individuals more swiftly and respond appropriately. These AI systems can analyze patterns and trends in student responses, enabling schools to tailor their support strategies and interventions effectively.

The regular check-ins help destigmatize the act of speaking up, as students may feel more comfortable sharing their feelings in a less formal setting. This ongoing dialogue can build trust between students and staff, making it easier for students to voice concerns when they arise. Ultimately, these advancements not only enhance the safety and wellbeing of students but also empower them with the resources and support necessary to navigate their challenges confidently.

Questions to ask

- Q: How frequently do we conduct check-ins, and is the frequency sufficient to catch students who may be at risk?
- Q: Do our check-in processes allow for anonymous feedback or disclosures, giving students a safe space to share their concerns without fear of retribution?
- Q: How do we ensure that staff members are trained to respond effectively to the insights gathered from these check-in tools, especially regarding CSAM or grooming-related risks?
- Q: Are we able to integrate these tools with existing support systems (e.g., counseling services, reporting mechanisms) to provide immediate help when risks are identified?
- Q: How do we communicate the purpose and importance of these check-ins to students, so they understand that these tools are meant to support them?

By asking these questions, schools can ensure that their student check-in tools effectively address the risks of CSAM and exploitation, providing a comprehensive safety net for their students.



Section 6

Final Thoughts

Without doubt, the emergence of AI is bringing greater challenges to keeping children safe and well online. But there is reason for hope.

Schools have always been the lighthouse of support for communities in times of difficulty - a beacon of guidance and enlightenment. Importantly, now, around the world they have come full circle in recognising the vital role and positive outcomes they can influence through the technology solutions and educational strategies they employ.

By leaning into these tools and focusing on a collaborative and strengths-based approach, schools can empower students and their communities with the knowledge and skills they need to thrive in online environments.

And to not just on risk, but on the resilience and potential they have as young learners and responsible digital citizens. With the right support and guidance, children will harness the numerous and meaningful benefits of AI while mitigating the concerns, and, by scaffolding their digital experiences and nurturing their strengths through the promotion of a culture of digital citizenship, schools can build a safer and more positive online environment for future generations.

Together, we can embrace the opportunities presented by AI and pave the way for a brighter, more secure future for our children.





About Qoria

Qoria began life as four parents with nothing more than a slide deck and a determined desire to make the internet a safer place for our kids.

Fast forward a decade and we're an ASX listed, global leader in child digital safety technology and solutions.



5004

Ooria staff

work hand in hand, across 4 continents and multiple disciplines, to keep children safer.



24m

Children

are supported and protected every day by our technologies.



6m

Parents

have greater visibility and control of their child's online behaviours and screen time.



29,000

Schools

give better digital safety and wellbeing support because of our solutions.

We're 500+ people working across 4 continents, looking after 29,000 schools, supporting 6 million parents and helping to keep 24 million children safer than they were before.

Our impact has grown exponentially but our purpose remains the same. Working to keep every child safe and thriving in their digital life is why we exist. But as online risk continues to explode we can't progress our purpose by providing technology alone, as vital as that is.

To build an online environment where children can thrive requires transparency, collaboration, and most importantly, a collective will to demand the changes that truly protect the next generation.

We must facilitate insight sharing between schools, and between schools and parents. We must empower everyone to learn from each other. And we must call out practices in the wider tech world that hinder the efforts and compromise children's safety.

For us, keeping kids safe and thriving in their digital lives is not just a goal, it's a moral imperative. And we're using all that we have - our technologies, our creativity, our voice, our vista - to do something about it.

Our schools, parents and children deserve nothing less.



Further Reading

Useful links

Pornography's Impact on problematic sexual behaviors

https://www.theguardian.com/lifeandstyle/article/2024/sep/02/i-think-its-natural-why-has-sexual-choking-become-so-prevalent-among-young-people

School Preparedness in the Age of AI

https://cdt.org/insights/report-off-task-edtech-threats-to-student-privacy-and-equity-in-the-age-of-ai/https://www.commonsensemedia.org/sites/default/files/research/report/generative-ai-in-k-12-education-whitepaper-updated-aug-2024-final-2.pdf

Supporting Victims of Deepfakes and NCII (Non-Consensual Intimate Images)

https://www.k12dive.com/news/schools-deepfake-images-student-supports/728107/

The Importance of Al-focused Education for Students

https://cdt.org/insights/report-off-task-edtech-threats-to-student-privacy-and-equity-in-the-age-of-ai/.

Addressing Parental Concerns & Providing Support

https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-children/children-and-parents-media-use-and-attitudes-report-2024/



Bibliography

- 1. https://www.internetmatters.org/hub/research/generative-ai-in-education-report/#full-report
- 2. https://childlight.org/sites/default/files/2024-05/executive-summary.pdf
- 3. https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf
- **4.** https://mashable.com/article/online-child-exploitation-dangers
- 5. https://play.google.com/store/apps/eva
- **6.** https://www.theguardian.com/australia-news/article/2024/jun/12/schoolboy-arrested-after-allegedly-post-ing-fake-explicit-images-of-female-students-ntwnfb
- **7.** https://www.abc.net.au/news/2024-06-13/ai-generated-deepfake-pornography-school-students-teachers/103969414
- 8. https://www.theeducationpeople.org/blog
- **9.** https://www.esafety.gov.au/parents/issues-and-advice/are-they-old-enough#:~:text=Determine%20your%20 child%27s%20readiness%20for,by%20the%20age%20of%204.
- 10. https://www.bbc.co.uk/news/technology-68838029
- **11.** https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Policy_and_Legal_Affairs/Onlineageverification/Report
- **12.** https://www.thorn.org/blog/new-thorn-research-examines-youth-experiences-and-attitudes-about-on-line-grooming
- **13**. https://www.sydney.edu.au/arts/news-and-events/news/2023/10/06/new-study-reveals-teenagers-social-media-use-and-safety-concerns.html
- 14. https://mashable.com/article/ai-companion-tee%20ns-safety
- **15.** https://www.commonsensemedia.org/sites/default/files/research/report/2024-the-dawn-of-the-ai-era_final-release-for-web.pdf



Contact Qoria

To find out more about the digital safety solutions Qoria provides for school and their parental communities please reach out to us through our regional touchpoints below. Or visit Qoria.com



North America:

Contact: inquiries@linewize.com Visit: www.linewize.com

Australia:

Contact: enquiries@linewize.io
Visit: www.linewize.io

New Zealand:

Contact: enquiries@linewize.co.nz Visit: www.linewize.co.nz



United Kingdom:

Contact: enquiries@smoothwall.com Visit: www.smoothwall.com



Spain:

Contact: info@qustodio.com Visit: www.qustodio.com



Australia:

Contact: enquiries@ysafe.com.au Visit: www.ysafe.com.au

Qoria

Spain:

Contact: enquiries@qoria.es Visit: www.qoria.es

EMEA:

Contact: **enquiries@qoria.eu** Visit: **www.qoria.eu**



Qoria is an ASX listed, global technology company, dedicated to keeping children safe and well in their digital lives. We harness the power of connection to close the gaps that children fall through, and to seamlessly support them on all sides - at school, at home and everywhere in between.

Find out more www.qoria.com

Qoria







