Beazley Logo

Article

How deepfakes are reshaping child safeguarding risks

July 24, 2026

AI-generated and AI-manipulated images (known as “deepfakes” which come under the broader term “synthetic media”) are changing how safeguarding risks develop in schools. They make it much easier for harm such as child sexual abuse material (CSAM) to be created, spread and escalate quickly. While the technology is recent, the behaviours are not. The technology just shortens the gap between decision and real harm. Distribution then makes it worse through bullying and ruined reputations, with challenges around identifying the source of distribution and how to contain it.

Schools in the US[1] and Australia[2] are already seeing cases and lawsuits where students used publicly available images to generate explicit material featuring their peers at scale. Offences by teachers using the technology to create CSAM have emerged in Mississippi[3] and Texas[4].

Scrutiny is scaling, and so are the risks

Regulation and law are struggling to keep pace with the development and deployment of deepfake technology and there is inconsistency across jurisdictions. This leaves schools without clear guidance on how to police and respond to such incidents in circumstances where parents, regulators and communities still expect a thoughtful, effective and timely response.

And though the threat outpaces regulation, the direction of travel is clear. The Australian government has already banned social media for children under 16 and the UK is now set to follow suit by 2027. Furthermore, the EU[5], UK[6] and US[7] enacted legislation against AI-generated and deepfake CSAM. Cases are fast surfacing globally with around 90 schools said to have been hit[8] and with how widespread and accessible technology has become, this may not represent the full extent of the problem. With this in mind, we can expect more civil claims and stricter duty-of-care expectations. And when regulation does catch up, decisions can be picked apart retrospectively.

How schools can turn duty of care into action

Schools must focus on prevention to protect against such incidents, and preparation to lessen the impact on pupils, their families, wider communities, as well as themselves.

Prevention

Early awareness

Issues arise from everyday behaviours – unmonitored messaging, informal relationships, shared images and unclear boundaries. Proactive safeguarding means noticing and acting decisively on early warning signs, but policies against these behaviours only work when people understand them, apply them consistently, and assume responsibility for reporting concerns.

External access

Vulnerabilities pervade where formal oversight fades and informal roles – volunteers, contractors, visiting staff or informal helpers – carry risks that are easy to miss. Deepfake abuse can happen through a single interaction and liability can arise whether someone is full-time or part time. For that reason, consistency with background checks and safeguarding measures are imperative regardless of an individual's employment status within an organization.

Stronger controls on public-facing content

Every public-facing or casually stored image or video can become material for misuse, with incidents, guidance and warnings already surfacing[9]. Tighter standards and controls on personal information, identifiable media, storage and access helps protect pupils and reduce the school’s liability.

IT and security

On the technical side, schools can close obvious doors by blocking known deepfake tools, flagging risky activity, watermarking official content, conducting social media screening (where appropriate) and tightening verification around sensitive requests that arrive via eerily convincing audio or video.

Education and awareness

Safeguards alone won’t hold unless students are prepared for today’s deepfake reality. Where curiosity, shock or humour can override judgement, media literacy and Personal, Social, Health and Economic (PSHE) education lessons can enable understanding around how easily content can be manipulated, how fast “just sharing” can cause real harm, and the ethical considerations around consent, power and the ensuing psychological harm.

Preparation

Updating policies and procedures

Safeguarding against deepfakes starts with recognising this reality and naming it clearly: synthetic media is not a hypothetical risk; it is a live safeguarding threat. Policies need to say so explicitly, and schools need a clear, rehearsed response for the moment it happens – who the victim goes to, who takes control, how quickly parents are informed and when and how outside authorities are involved.

Incident response

Should the worst happen, the response must be swift and human: surround the child with support as well as offering assurance to others involved such as parents and school boards, preserve evidence and take down what can be taken down. In moments like these, how a school responds not only mitigates litigation and/or resolves an incident, but determines whether students feel protected, believed and safe enough to speak up.

Supporting organisations with synthetic threats

As malicious AI-enabled deepfakes surge and spread, specialist standalone Sexual Molestation Liability (SML) insurance solutions that combine risk management and incident response services, such as Safeguard, can help organisations to stay ahead of the risks before they take hold. Risk management can include best practice guidance, training and dedicated support. But when prevention isn’t enough, the right policies are built to respond with crisis communications, legal services, counselling, financial help, investigatory services and dedicated reputational damage control.

20 years on from writing one of the first standalone SML policies, the need for specialist cover has never been clearer. Beazley Safeguard is grounded in prevention, and as the risk continues to evolve, so does our commitment to helping organisations build safer environments and protecting people and communities.

Read on...

[1] Teens who made deepfakes of classmates were just sentenced. Will it make a difference?

[2] Female pupils have 'AI deepfakes' made of them as parents criticise school

[3] Former middle school teacher sentenced to prison for creating AI-generated porn of students

[4] https://texasscorecard.com/local/austin-isd-teacher-re-arrested-for-using-ai-to-generate-sexual-images-of-students/

[5] https://euroweeklynews.com/2025/06/17/new-eu-law-bans-ai-generated-child-abuse-content/

[6] https://www.legislation.gov.uk/ukpga/2025/18/section/138

[7] https://news.ballotpedia.org/2025/07/22/forty-seven-states-have-enacted-deepfake-legislation-since-2019/

[8] https://www.wired.com/story/deepfake-nudify-schools-global-crisis/

[9] https://www.theguardian.com/technology/2026/may/08/uk-schools-remove-pupils-photos-online-ai-blackmail-threat-grows

Home

©Beazley Group | LLOYD's Underwriters