Guide to responding to image-based abuse involving AI deepfakes

eSafety Toolkit for Schools

Creating safer online environments

Why has this guide been produced?

This guide provides support and advice to Australian schools to respond confidently and effectively to image-based abuse that involves AI deepfakes. Visit eSafety's web page on **what you can report** for a definition of what constitutes image-based abuse, including the use of deepfakes. Online safety policies and procedures should align with relevant legislation, as well as departmental or sector policies and procedures.

School response

Regardless of when or where an incident has occurred, or whether the incident meets the threshold of 'serious', if a student or staff member is distressed and needs support, their wellbeing, rights and best interests should guide your response.

It is important to provide that support in a timely manner. If a staff member is unsure about what to do, they should seek advice from the school leadership or online safety team.

Schools should have a designated person or team of people responsible for online incidents. All members of staff (including non-teaching staff) need professional learning to recognise, respond to, or refer, serious online safety incidents.

What is deepfake image-based abuse?

Image-based abuse is when someone shares, or threatens to share, an intimate image or video of a person without their consent. This includes nude or sexual deepfakes created using artificial intelligence (AI) tools and apps.

These highly realistic AI deepfake images or videos often appear to show real people in sexually suggestive or explicit situations that did not take place. This means creating and sharing them without consent can be extremely harmful.

Nude or sexual deepfakes and other synthetic media are becoming more common as AI tools and apps become more readily available and easier to use.

Nude or sexual deepfakes can be used for:

- sexual gratification
- peer to peer sexual harassment
- posting on fake social media accounts created to humiliate the victim or damage their reputation
- selling or trading in schools or workplaces
- coercive control within intimate relationships
- posting to online pornography sites and dating apps without the victim's knowledge
- child sexual abuse and exploitation material.

Victims of AI deepfake image-based abuse may not know that the material exists or has been shared or posted.

Important note

It is important for school leaders and staff to act when they are aware of an incident in which AI-assisted or digitally altered images have been used to depict students or staff in intimate or sexual contexts without consent. This type of content may be shared via social media, messaging apps, or across school social and other networks, or it could be held on a device without being shared.

Schools and youth-serving organisations act as frontline responders and can provide access to trusted, trauma-informed support systems while fostering environments that reinforce accountability and promote help-seeking behaviours for image-based harms.

Deepfake laws

Adults

Australia's Criminal Code Amendment (Deepfake Sexual Material) Act 2024 strengthened the Criminal Code Act 1995 by introducing offences for transmitting sexual material relating to an adult without their consent. The offences capture both unaltered material and material created and altered using technology (commonly referred to as 'deepfakes'). A standalone offence for the non-consensual sharing of adult private sexual material carries a maximum prison sentence of six years.

It is irrelevant whether the photos, videos or audio depicting the person are in 'unaltered form' or have been 'created or altered using technology'.

Under 18s

It's important for young people under 18 to understand that the creation and distribution of nude or sexual deepfakes (even as a misguided joke or meme) can be a criminal offence in some states, as it is for adults across the country. Laws around this kind of behaviour are evolving, and there is increasing legal recognition of the serious harm it can cause. For example, legislation in South Australia criminalises the creation and distribution of AI-assisted deepfakes that are humiliating, degrading, invasive, or sexually explicit, even if the content is entirely synthetic, and the offender is under 18.

A range of responses are also being used including diversion programs, restorative approaches, community service or supervised orders and counselling and education, recognising that education and accountability can be more effective than punishment alone.

Is your school prepared for a deepfake incident?

To reduce the risks, schools and education sectors can start by clearly identifying deepfake intimate imagery as a form of abuse and including it in existing policies relating to student and staff behaviour. For example, policies addressing bullying prevention and response, student protection, code of conduct, and ICT acceptable use. Schools and education sectors can take these two steps.

- Clearly specify the responses and actions that will be taken in cases involving the creation, sharing or distribution of nude or sexual deepfakes, including disciplinary measures and appropriate pathways for reporting and support for all those involved.
- Consider communication protocols that include guidance on how the school will communicate with impacted individuals, relevant staff and where appropriate, students' parents/carers, and the broader school community, while respecting privacy and legal considerations.

How to manage a deepfake incident

- The wellbeing of young people and/or the staff member should be the primary concern.
- An organisation's reputation should not be the focus of the response.
- Ensure the incident is managed by a designated member of the school's leadership team, with information shared only on a need-to-know basis to maintain confidentiality and minimise unintended effects.
- Follow the steps on the next page for reporting image-based abuse to police. Also report it to eSafety if it has been shared online or someone is threatening to share it.
- Encourage those responsible to follow eSafety's advice for stopping the image or video spreading: **What to do if you shared someone's intimate image or video.**
- It is essential that the student/s and or staff member/s impacted by a deepfake incident is supported in a way that promotes their sense of agency and involvement in decision-making.
- A student's family should also be kept appropriately informed about the actions being taken by the school to address the situation and ensure their child's safety and wellbeing.
- Ensure your organisation's critical incident response plan includes the likelihood of a deepfake event.
- Consider this guidance alongside your safeguarding and child protection policies relevant to your school or education sector.

For more information **Deepfake nudes and young people**, Thorn and Burson (March 2025), and **Youth perspectives on online safety 2023: One in 10 minors say peers have used AI to generate nudes of other kids**, Thorn and BSG (2024).

Steps to report deepfake image-based abuse

Initial response

- Upon notification of an incident, the school should find out whether the image/s or video/s depict students or staff in intimate or sexual contexts without consent. Prioritise the immediate safety and wellbeing of all individuals involved.
- Inform the principal and determine a designated lead from the school's leadership team to coordinate the response.
- Consult with relevant education sector support (for example, critical incident or wellbeing team) where appropriate.
- Follow school and/or education sector policies and procedures.
- Ensure those affected are meaningfully included in the response process and supported to make informed decisions.
- Collect evidence to provide to police and eSafety, but avoid unnecessary exposure to and storage of explicit material. Make a written description and note where it is located. Read more about collecting evidence and securing devices in **RESPOND 4 Guide to responding to image-based abuse, including sexual extortion**.

Report to local police

- Report to local police first – they have the powers to seize devices and conduct interviews.
- Obtain an event number from local police and then report it to eSafety.

Report to eSafety

- Use the **eSafety Report Abuse portal**.
- You may provide a police event number (if able) to eSafety.
- Report if the images (real, digitally altered or AI-assisted deepfakes) are shared without consent.
- Report if threats have been made to share images or videos.
- Provide collected evidence and details.

Support for affected students and/or staff

It is misleading to assume that synthetically generated content causes no harm or is victimless. This minimises the serious impacts such content can have on those targeted. The harms and impacts experienced by victims may include:

- compromise of dignity and privacy
- feelings of shame, humiliation and self-blame
- fear they will not be believed or what has happened will be discounted or minimised
- fear they will be blamed for what has happened
- feelings of isolation, mistrust, and these create barriers to help-seeking.

A trauma-informed support system prioritises the emotional safety and recovery of affected student/s and/or staff members targeted. These are some steps you can take.

- Provide support for the victim and affected peers and family. Remind them the situation is not their fault. Keep the staff member, young person, and their parents/carers informed about actions being taken.
- Hold the creator of the material accountable in a way that aligns with sector policies, considering trauma-informed and restorative approaches.
- Maintain confidentiality.
- Engage with additional support services to assist with recovery, for example, school counsellor, education sector support, Kids Helpline.

Include AI in consent and respectful relationships education

Integrating information about AI-assisted deepfakes into consent and respectful relationship education helps students develop the awareness, empathy and critical thinking skills needed for safe and respectful digital interactions. Content might include:

- learning that the creation of nude or sexual AI deepfakes of someone without their consent is illegal
- recognising the nature and significance of harm, despite the synthetic nature of the material
- understanding that sharing and on-sharing nude or sexual deepfakes is also image-based abuse
- understanding how nude or sexual deepfakes can be used as a tool of coercion within intimate relationships
- ensuring students know where and how to report deepfake image-based abuse.

eSafety resources

- **Report** to eSafety
- Find more information on **image based abuse** and **deepfakes**.
- Register for **eSafety's professional learning** for educators and youth-serving professionals.
- Access **eSafety's Toolkit for Schools** including the **RESPOND Element** to update policies and procedures.
- Make sure your school is part of the **eSafety Champions Network**.
- Access curriculum-aligned and wellbeing **classroom resources**.
- Share resources for **Young People** with secondary students.
- **Access parent and carer advice.**
- Access **The eSafety Guide** to learn more about apps and platforms, including safety and reporting features.
- For additional information: contact **educationsectors@esafety.gov.au**

Published 06/2025