

GUIDANCE FOR EMPLOYERS AND HR

Managing Allegations Involving Suspected Deepfake Content

A clear, proportionate safeguarding response when digital material may be manipulated or fabricated.



1 Safeguarding first

Take the allegation seriously and prioritise any immediate risk to children.

2 Evidence unverified

Preserve original material and avoid deciding whether content is real or fake yourself.

3 Risk-led response

Consult LADO early, work with relevant agencies and record uncertainty clearly.



If in doubt, consult your LADO early. It is always better to discuss than delay.

Managing Allegations Involving Suspected Deepfake Content

Guidance for Employers and HR

Who this guidance is for

This guidance applies to any organisation working with children that engages individuals in positions of trust, as employees, contractors, commissioned staff, and/or volunteers. This includes, but is not limited to, schools and colleges, early years providers, fostering services, residential settings, school transport providers, health services, youth and voluntary organisations, sports providers, and faith or community groups. It is intended to help employers, managers and HR teams understand what their responsibilities are when an allegation is made and the available evidence includes, or may include, **suspected deepfake or AI-generated image, video or audio material**.

This guidance should be read alongside the organisation's own safeguarding policies and local allegation management arrangements procedures.

1. Purpose of this guidance

Deepfakes are images, audio, or videos that have been created or altered using digital or AI technology to look real. They can be used to humiliate, harass, blackmail, mislead, or make false allegations. Some may be sexual in nature, while others may imitate a person's voice or behaviour. They can cause serious harm, including damage to a person's reputation and safeguarding concerns.

Even when content is not real, the impact can be very real. National guidance is clear that AI-generated sexual images and other deepfake material should always be taken seriously. Safeguarding action must not be delayed while trying to prove whether the content is real or fake.

This guidance supports employers and HR teams to respond appropriately where a deepfake may form part of an allegation against an adult who works with children. It does not introduce a new process or threshold. Instead, it explains how existing safeguarding and allegation management procedures should be applied when digital or AI-generated content is involved.

2. Key principle: respond to potential harm and risk, not confirmed authenticity

The most important principle is that **safeguarding decisions must be based on potential risk and harm, not on proving whether digital material is real or fake**.

Deepfakes can be very convincing, and it is not always possible to quickly or reliably confirm authenticity. Even specialists may need detailed technical analysis to reach a view.

In practice:

- Do not dismiss a concern because someone claims the material is AI generated.
- Do not assume the material is genuine and reach firm conclusions without further consideration.
- Treat the material as unverified digital evidence.
- Focus on any risk to children and seek advice from the LADO and, where appropriate, the police.

Professionals working with children are **not expected to determine authenticity before taking action**. Suspected deepfake or AI-generated abuse material should be treated with the same seriousness and urgency as other safeguarding concerns.

This approach is essential because harm can occur even if the material is fabricated. A child may feel scared, humiliated, or unsafe. A professional may be wrongly accused. Deepfake content may also be used as part of wider patterns of abuse, coercion, or harassment.

The content may be digital, but the impact is real.

3. LADO thresholds still apply in the usual way

The existence of suspected deepfake content does **not** alter the threshold for consulting the LADO. Employers should continue to ask whether the person working or volunteering with children may have:

1. **Behaved in a way that has harmed a child, or may have harmed a child**
2. **Possibly committed a criminal offence against or related to a child**
3. **Behaved towards a child or children in a way that indicates they may pose a risk of harm to children**
4. **Behaved in a way that raises concerns about suitability to work with children**

These threshold questions are the same as in any allegation management case. A suspected deepfake can fall within any of the usual categories, depending on the concern. For example, it may suggest harmful behaviour, possible criminal conduct, or behaviour that raises concerns about an adult's suitability to work with children. It is also important to recognise that creating or sharing false or manipulated material about someone may itself be abusive and require safeguarding action.

Because deepfake material can both suggest harm and be used to make false allegations, decisions must not be reduced to a simple question of whether the video or image is real. The full context and potential risks must always be considered.

If the concern would normally meet the threshold for LADO consultation, a referral should still be made—even if there is reason to believe the material may have been manipulated. Suspecting that content is fabricated should lead to careful handling of evidence and balanced enquiries. It should not be used as a reason to avoid following the allegation management process.

4. Immediate actions for employers and HR

When an allegation is received, employers should take calm and practical steps to **protect children, preserve evidence, and ensure a fair process**. The immediate focus should be on safeguarding risk, not on analysing whether the media is real or fake.

This includes:

- Considering whether any child may be at immediate risk
- Deciding if an individual's duties or contact with children need to change as an immediate precautionary measure.
- Seeking advice from the LADO and, where appropriate, the police without delay
- How best to preserve evidence.

Any digital material should be secured carefully. Wherever possible, keep original files rather than screenshots or forwarded copies. Originals may contain important information about where the material came from and how it has been handled. Early decisions about how evidence is stored and shared may be important later. See section 6.

Employers should also ensure support is available for everyone involved. A child may need safeguarding support. The adult subject to the allegation may also be distressed, particularly if they believe the material is fabricated. A fair process includes clear communication, appropriate confidentiality, and emotional support for all parties.

5. Risk assessment before authenticity is established

A suspicion that material may be a deepfake does not remove the employer's duty to assess and manage risk. In many cases, uncertainty makes this even more important. Employers should consider:

- What is the risk to children if the allegation is true?
- What is the risk if the material is false but part of harassment, intimidation, or coercive behaviour?
- What temporary steps are needed to keep children safe while enquiries continue?

Deepfakes can be used to create false allegations, and genuine material may be dismissed as fake. This means decisions should not rely on whether the content is believed to be real. Instead, they should be based on a balanced assessment of the allegation, the context, and any wider information and evidence available.

Risk management actions may include increased supervision, changes to duties, redeployment, or, in some cases, suspension. The response should match the level of risk. HR advisors and LADO can offer support with these decisions.

All decisions must be proportionate, kept under review, and clearly recorded. Records should explain:

- what is known and not yet known
- what risks were considered
- what actions were taken and why
- who was involved in decision-making

Clear recording is especially important where evidence is uncertain, as decisions may be closely reviewed later.

6. Handling digital evidence carefully

Deepfake allegations require careful handling of digital evidence. What a file shows is only part of the picture. Its value depends on where it came from, how it has been handled, and whether it has been changed.

Chain of custody is important. If files are copied, edited, screenshot, or shared informally, it becomes harder to know if they are original or have been altered.

For this reason, employers should:

- **Keep the original file where possible.** Original images, videos, audio files or messages may contain information that helps identify the source and history of the content. Do not crop, edit, enhance, or annotate the material.
- **Take screenshots of the content and context.** Capture the post, image or video together with relevant details such as the username, date, time and platform, as content may later be edited or removed.
- **Record account details.** Note any available identifiers, such as usernames, profile URLs, user IDs, account handles or linked telephone numbers, to support any subsequent investigation.
- **Maintain an audit trail.** Record when the content was identified, by whom, where it was found and any actions taken to preserve the evidence.

Staff should also not try to decide if it is real or fake based on appearance. Research shows that people are often not good at spotting deepfakes, and personal judgement is not a reliable basis for decision-making.

Early consultation with the police is important where digital evidence may be at risk. This helps coordinate action to preserve evidence—such as mobile phone data—and reduces the risk of material being deleted or altered if individuals become aware of the allegation.

If the material is central to the concern, early advice from the police may be needed—especially where there may be a criminal offence or specialist digital analysis is required. The aim is to ensure that decisions are informed by appropriate expertise, not guesswork.

7. Investigation approach: corroborate, balance and record

A good investigation in a deepfake case should not rely only on the digital material. Employers should look for evidence from other sources to build a clear picture. This may include witness accounts, work records, communication logs, attendance information, timelines, and any existing notes or system records.

Decisions should be based on source, context, and cross-checking information—not on the media alone.

The adult who is the subject of the allegation must be given a fair chance to respond at the right stage. A balanced enquiry should consider both possibilities: that the material is fabricated, and that it reflects real behaviour. Relevant factors may include motive, opportunity, access to technology, relationship history, and whether the content matches verified facts.

The aim is not to decide whether the video or image is convincing. The aim is to establish, as far as possible, what happened and what risk remains.

If the evidence cannot be resolved clearly, this should be stated. Not all cases will lead to a firm conclusion about the material itself. However, the organisation may still need to make safeguarding or employment decisions based on the overall evidence and any remaining risk.

Records should clearly set out:

- what is established
- what is disputed
- what remains uncertain

This ensures decisions are transparent, proportionate, and fair.

8. Safeguarding risks that are particularly relevant in deepfake cases

Employers should be aware of different patterns of risk linked to deepfakes.

One risk is the use of manipulated or AI-generated media to falsely accuse, embarrass, or humiliate someone, including professionals or other children. Another is the use of “nudify” or similar tools to create sexual images of others without consent. This is a growing form of image-based abuse, including among school-age children.

Deepfakes may also be part of wider patterns of harm, such as coercion, blackmail, harassment, or other forms of technology-facilitated abuse. They can be used to impersonate, threaten, or damage someone’s reputation.

Employers should therefore look beyond the content itself and consider:

- who created the material
- why it was created
- how it has been shared
- whether this behaviour is itself abusive

The focus should always be on the wider context and any safeguarding risks, not just on the media in isolation.

9. Partnership working and referral decisions

Deepfake cases are rarely managed well by one agency alone. Involving the LADO helps ensure clear separation between safeguarding thresholds, evidence issues, and employment decisions.

Police involvement may be needed where:

- the material is sexualised
- there are concerns about blackmail, coercion, or harassment
- illegal images may have been created or shared
- potential criminal offences are identified

AI-generated child sexual abuse material is illegal and must be treated in the same way as other child sexual abuse material, including reporting to the police.

Employers also have ongoing responsibilities after the allegation process. Depending on the outcome, this may include:

- internal disciplinary action
- referral to a professional regulator
- referral to the Disclosure and Barring Service (DBS), where the legal criteria are met

These decisions must be based on all available evidence and the level of risk, not simply on whether digital material was proven to be real or fake.

10. Community impact and communication

Deepfake and AI-generated content can affect more than the individuals directly involved. Even where material is later found to be false or manipulated, its circulation can cause significant distress, damage reputations, undermine trust, and create fear or uncertainty within schools, organisations, families and the wider community.

Agencies should be mindful that rumours, speculation and the sharing of unverified information can cause additional harm and may affect safeguarding enquiries, employment processes, and potential criminal investigations. In some cases, the creation or sharing of the material may itself be abusive or criminal.

Where concerns become known more widely, employers should work closely with the LADO, police, HR advisers and legal advisers to consider how information is managed and communicated. A coordinated approach can help:

- protect children and vulnerable individuals;
- preserve evidence and ongoing enquiries;
- maintain appropriate confidentiality;
- avoid unnecessary speculation;
- support those affected; and
- maintain confidence in the fairness and integrity of the process.

Communications should be factual, proportionate and consistent. Agencies should avoid commenting on the authenticity of material until enquiries have been completed and should focus on the actions being taken to safeguard children and manage risk.

11. What good practice looks like

Good practice in deepfake cases does not require employers to be technology experts. It requires applying normal allegation management principles carefully, where digital evidence may be uncertain or disputed.

This includes:

- making a timely referral to the LADO where the threshold is met
- carrying out a clear and proportionate risk assessment
- preserving digital material properly
- seeking evidence from other sources
- involving police or other agencies where needed

- recording any uncertainty clearly and honestly

It also means supporting both children and adults involved and avoiding pressure to reach quick conclusions before the facts are understood.

Good practice includes learning and improvement. The use of deepfakes is increasing and changing quickly. Organisations should use each case—whether proven or not—to strengthen training, raise awareness, and improve local safeguarding procedures.

In conclusion

Where an allegation involves suspected deepfake or AI-generated material, employers must act on potential harm rather than waiting to confirm whether the content is real. The usual LADO thresholds still apply. Digital material should be treated as unverified evidence unless and until its reliability is established. Decisions should be based on:

- the risk to children
- information from other sources
- careful handling of digital evidence
- clear, reasoned decision-making that is properly recorded

With a calm, structured approach grounded in safeguarding principles, employers can manage these cases confidently. Even where technology creates uncertainty, applying consistent processes, seeking the right advice, and focusing on risk will lead to safe, fair, and defensible decisions.

North Yorkshire LADO

For more **information on Managing Allegations** please visit the North Yorkshire Safeguarding Children Partnership (NYSCP) webpage for LADO:

[**NYSCP - LADO WEB PAGE**](#)

[**NYSCP - LADO PROCEDURES**](#)

For all **new notifications** to LADO or to request a consultation please submit a notification form:

[**NYSCP - LADO NOTIFICATION FORM**](#)

o **contact LADO:**



: 01609 798005



: lado@northyorks.gov.uk



References

This guidance has been informed by statutory safeguarding guidance, including *Working Together to Safeguard Children* and *Keeping Children Safe in Education*, together with guidance and resources produced by the National Crime Agency (NCA), CEOP Education, the Internet Watch Foundation (IWF), and other recognised organisations specialising in online safety, digital evidence, AI-generated abuse material and deepfake harms.

Statutory Safeguarding Guidance

Working Together to Safeguard Children
(Department for Education) - [Working together to safeguard children - GOV.UK](#)

Keeping Children Safe in Education (KCSIE)
(Department for Education) - [Keeping children safe in education - GOV.UK](#)

AI-Generated Child Sexual Abuse Material (AI-CSAM)

Internet Watch Foundation (IWF) and National Crime Agency (NCA)

Professional Guides: Understanding and Responding to AI-Generated Child Sexual Abuse Material - [AI CSAM Guides for Professionals | IWF & NCA](#)

Safeguarding Board for Northern Ireland (SBNI) Online Safety Hub

Child Sexual Abuse Material Generated by Artificial Intelligence: Guidance for Professional [Child sexual abuse material generated by artificial intelligence - Safeguarding Board for Northern Ireland](#)

NB: The following visual guides are available to download and print from [NYSCP - LADO WEB PAGE](#)

Deepfakes and Image-Based Abuse

Safeline

Deepfakes and Image-Based Abuse: Information for Professionals - [Deepfakes and Image-Based Abuse - Safeline](#)

Southwest Grid for Learning

Support and Advice for Adults on Synthetic Media and Deepfakes - [Support and Advice for Adults on Synthetic Media & Deepfakes | SWGfL](#)

Further Reading and Resources

[AI Image Manipulation - how will you protect staff and pupils?](#)

[Protecting your setting's images from AI manipulation and abuse](#)

Support for Victims

Report Remove (NSPCC / Internet Watch Foundation)

Helps under-18s report sexual images of themselves, including AI-generated or manipulated images, and seek removal from the internet - [Eliminating Child Sexual Abuse Online | Internet Watch Foundation IWF](#)

Safeline

Provides support for child and adult victims of image-based abuse and synthetic images of abuse. Offers information for professionals, victims, and families about the impact of deepfakes - [Get Help - Safeline](#)

SWGfL Report Harmful Content

Supports individuals experiencing harmful online content, including synthetic media, impersonation, harassment and deepfake abuse. Can assist with reporting content to platforms and seeking removal. [Support and Advice for Adults on Synthetic Media & Deepfakes | SWGfL](#)



MANAGING ALLEGATIONS INVOLVING SUSPECTED DEEPFAKE CONTENT

QUICK GUIDE FOR DSLs, MANAGERS AND HR LEADS



WHY THIS MATTERS

AI and deepfake technology can create realistic images, audio and video that may appear genuine. Safeguarding risks, the welfare of children and the LADO threshold remain the primary considerations. Do not focus solely on whether the material is real or fake.



1 RESPOND IMMEDIATELY



- ✓ Take the allegation seriously.
- ✓ Prioritise the safety and wellbeing of children.
- ✓ Secure any available evidence.
- ✓ Consider whether immediate safeguarding measures are required.
- ✓ Refer to the LADO and/or Police without delay where thresholds are met.
- ✓ Do not wait for confirmation that content is genuine before taking action.

2 APPLY NORMAL LADO THRESHOLDS

Ask whether the allegation suggests the adult has:

- Harmed or may have harmed a child
- Possibly committed a criminal offence
- Posed a risk of harm to children
- Behaved in a way that indicates they may be unsuitable to work with children

→ If YES to any of the above → refer to the LADO
Do not wait to confirm whether content is real.

3 TREAT DIGITAL MATERIAL AS UNVERIFIED EVIDENCE



- ✓ Treat all digital content as unverified.
- ✓ Preserve original files where possible.
- ✓ Record who provided the material, when it was received and how it was obtained.
- ✓ Avoid editing, cropping or widely sharing the content.
- ✓ Do not attempt to determine authenticity yourself.

4 FOCUS ON RISK, NOT PROOF



- ✓ What is the risk if the allegation is true?
- ✓ What is the risk if the allegation is fabricated or malicious?
- ✓ Are precautionary safeguards needed while enquiries are ongoing?

Any action should be: proportionate, based on risk and reviewed regularly.

5 WORK WITH RELEVANT AGENCIES



Consult the LADO for advice and coordination.



Involve Police where criminal offences, sexual content, threats or blackmail may be present.



Involve Children's Social Care where there is potential risk to a child.

6 INVESTIGATE THOROUGHLY

Do not rely solely on digital evidence.



Seek corroboration through:

- Witness accounts
- Records and logs
- Chronologies
- Behaviour history
- Contextual information

Explore both possibilities:



The material may be fabricated.



The material may reflect genuine behaviour.

7 MANAGE UNCERTAINTY



Some cases may remain inconclusive.

Clearly record:

- ✓ What is known.
- ✓ What is disputed.
- ✓ What cannot be determined.

Avoid assuming content is either genuine or fake without evidence.

8 EMPLOYMENT AND HR DECISIONS



Employment decisions should be based on:

- ✓ The overall evidence available.
- ✓ Safeguarding risk.
- ✓ Professional standards and conduct expectations.



Seek HR and LADO advice before making significant employment decisions.

9 RECORD CLEARLY



Ensure records explain:

- ✓ Actions taken.
- ✓ The rationale for decisions.
- ✓ Risk assessments.
- ✓ Agencies consulted.
- ✓ Areas of uncertainty.

Clear recording supports transparency and defensibility.

10 MANAGE COMMUNICATIONS CAREFULLY



- ✓ Avoid speculation.
- ✓ Do not share unverified information unnecessarily.
- ✓ Consider the impact on children, families, staff and the wider community.
- ✓ Agree communications with relevant agencies where appropriate.



Deepfake allegations can cause significant harm even when the content is false.

REMEMBER



Safeguarding comes first



Deepfake or not, the impact on children may be real



Treat digital content as unverified evidence



Do not rely solely on media or digital material



Refer to the LADO, Police and Children's Social Care where thresholds are met



Record uncertainty clearly and defensibly



KEY MESSAGE: Treat suspected deepfake material as unverified evidence. Act on safeguarding risk rather than assumed authenticity, and apply normal safeguarding and LADO procedures throughout.



DIGITAL EVIDENCE PRESERVATION

Quick Guide for Managers, HR, DSLs & Safeguarding Leads



STOP. DON'T TRY TO PROVE IT'S REAL OR FAKE.

Your role is to:



Safeguard children



Preserve evidence



Record actions taken



Seek advice from LADO and/or Police

1 PRESERVE THE ORIGINAL

DO

- ✓ Keep original images, videos, audio files and messages
- ✓ Save emails and online posts
- ✓ Retain links and usernames



DON'T

- ✗ Delete content
- ✗ Crop images
- ✗ Edit or enhance files
- ✗ Annotate evidence

2 CAPTURE THE CONTEXT

Take screenshots showing:

- 👤 Username/account name
- 📅 Date and time
- 🌐 Website or platform
- 💬 Comments and messages
- 🔗 URL or web address



Screenshots preserve what was seen.
Original files remain the best evidence.

3 RECORD WHAT YOU FIND

Create a simple evidence log:

📅	When was it found?
👤	Who found it?
📍	Where was it found?
📋	What action was taken?
👥	Who was informed?



Remember:
If it isn't recorded, it didn't happen.

4 LIMIT SHARING

Share only with:

- ✓ DSL
- ✓ Senior Manager
- ✓ HR
- ✓ LADO
- ✓ Police



Avoid:

- ✗ WhatsApp groups
- ✗ Personal devices
- ✗ Unnecessary email forwarding
- ✗ Staff gossip or speculation

5 THINK SAFEGUARDING

Ask:

- 🛡️ Is any child at immediate risk?
- 👤 Is anyone distressed or vulnerable?
- ➡️ Is the content being shared further?
- ⚠️ Does immediate action need to be taken?



Safeguarding action should not wait for proof of authenticity.

6 DON'T BECOME THE EXPERT

Avoid:

- ✗ Using online deepfake detection tools
- ✗ Trying to analyse metadata
- ✗ Making conclusions based on appearance
- ✗ Conducting your own forensic investigation



Instead:
Seek advice from the LADO and/or Police.

REMEMBER: S . E . C . U . R . E .



S

Safeguard children and adults



E

Evidence preserved



C

Capture context



U

Use need-to-know sharing only



R

Record all actions



E

Escalate when required



GOLDEN RULE:

PRESERVE FIRST. INVESTIGATE LATER.

Keep evidence safe. • Record what you know. • Protect children. • Seek advice.



This guide supports your organisation's safeguarding and allegation management procedures.
When in doubt, speak to the LADO and/or Police.