

Checklist: AI Deepfake Safety

A checklist to spot deepfake media and take simple, practical steps to protect yourself and your organisation. The checklist below gives you the basics - even simple steps can significantly reduce the risk from deepfakes.

Part 1: Factors That Increase Deepfake Risk

AI attacks may well be aimed at disruption, so all non-profits face growing risks, no matter how small or whether they use AI or not. However, some will be at greater risk. Use the checklist below to assess if you are:

- ☐ We work with women and/or children who are more likely to be the targets of deepfake abuse.
- ☐ We are part of a community that may well be targeted for abuse and misinformation, such as our Jewish, Muslim, and LGBTQI+ communities.
- ☐ Our beneficiaries are more likely to be susceptible to deepfake scams, such as younger or older people, or those with a learning disability.
- ☐ Larger organisation with significant amounts of sensitive personal or valuable financial data, such as credit card details may be at more risk to deepfake scams.

Part 2: Proactive Protection (The Prevention Checklist)

The less data available to create a deepfake, the lower the risk of becoming a target. Use these steps to secure your digital presence.

Action Point	Individual Check	Charity/Organisational Check
1. Digital Privacy Settings	Set all social media accounts (Facebook, Instagram, etc.) to "Friends Only" or Private . Check your settings regularly.	Ensure your social media policy does not allow posts with high-resolution images of beneficiaries or identifiable close-ups.
2. Multi-Factor Security	Enable 2-Step Verification (MFA) on your main email, banking apps, and any cloud storage/photo sites.	Mandate MFA for access to charity data, financial systems, and shared email accounts. Create a safe-word or only execute financial transactions using normal procedures – not calls or video decisions.
3. Content Sharing Rules	Avoid posting full-face, front-facing, or emotionally expressive close-up photos of yourself or family members.	Prioritise, non-identifiable photos (landscapes, objects, or blurred images) for general posts. If a person is needed, use wide-shot or low-resolution images. Disable website right-click.

4. Update Software & Devices	Turn on automatic updates for your phone, computer, and all apps.	Ensure the IT policy requires mandatory software updates across all charity devices (laptops, mobiles) to patch known security flaws.
5. Password Strength	Use a unique, strong password (over 12 characters) for every essential account. Consider using a reputable password manager.	Use strong password generation rules and never reuse passwords between personal and work accounts. Always change passwords on suspected compromise.

✓ Part 2: Detection and Verification (The "Spot the Fake" Checklist)

When you encounter an unexpected or suspicious video, image, or audio message, **STOP** and ask these critical questions.

A. Visual Red Flags (The "Look")

Check	What to Look For	Red Flag Indicator (Pause Immediately If...)
Eyes & Blinking	The person's eyes or blinking seem unnatural.	The person rarely blinks , or their eye movements are rigid and not focused on who they are speaking to.
The Skin & Face	The person's face looks too smooth, blurry, or unnaturally lit.	The skin texture is waxy or overly smoothed , or the shadows on their face don't match the background lighting.
The Edges	The area where the face meets the neck, hairline, or glasses.	There is flickering, digital blur, or rough edges around the person's head, or the body seems disjointed from the face.
Distortion	Look closely at their hands, ears, or teeth when they speak.	The hands look distorted (e.g., too many/few fingers) or the teeth look blurry and lack individual definition.

B. Context & Audio Red Flags (The "Sound and Situation")

Check	What to Listen For	Red Flag Indicator (Pause Immediately If...)
Lip-Sync & Sound	The audio quality or pitch sounds inconsistent or robotic.	The words don't sync up perfectly with the mouth movements, or the voice sounds flat, muffled, or artificial.
The Source	Does this feel right to me?	It's an unknown person or a friend/colleague sending an urgent, emotional request that seems out of character.
The Pressure	Are you being put under pressure to act immediately.	It creates fear, urgency, or excitement and/or demands you click a link, send money, or give away a password right now.
Verification	Can you find this content or story from two other reliable sources?	You cannot find the story or video on any reputable news site or the (supposedly) originating organisation's official channels.

Part 3: Response (What to Do Next)

Situation	Action
If you receive a suspicious message/call	DO NOT click the link, reply, or download the attachment. Use a separate, known channel (like a phone number from their website) to call the person/charity back and verify the request . Do not use a number or website link in any message or which a caller gives you – it may be fake.
If you find a harmful deepfake of yourself	Report the content immediately to the platform (Facebook, YouTube, etc.) where it is hosted and request a take-down. Keep a record of the URL and date. Also see sources of help at the bottom.
If you find a deepfake being used to scam	Report the incident to Action Fraud (UK) or your local police. For organisations, inform the relevant trustees and IT support immediately and isolate any potentially compromised accounts.

Part 4 – More Resources & Try the Quiz

This Which Report ([Deepfake phone calls, Mar 25](#)) provides more stats, examples and advice. TED Talks - [How to spot fake AI images](#) (13 mins, Apr 25). Try this [online deepfake quiz](#) by Iproov.com to see how many you can spot.

Part 5 – How to Get Help if You’re a Victim

If you, or someone you know is becomes the victim of deepfake sexual exploitation, there are people who will help you.

- **Revenge Porn Helpline** supports UK adults (18+) who have had intimate images shared without their consent. It offers free, confidential advice and practical help to get content removed and provides emotional support. Use this, if you are outside of the UK.
- **Childline - Report Remove** - to help young people under 18 in the UK to confidentially report sexual images and videos of themselves and remove them from the internet.

If you or someone you know is the victim of a scam. Action Fraud is the UK’s national reporting centre for fraud and cyber-crime where you should report fraud if you have been scammed. If you have made a payment, contact your bank as soon as possible.



We hope you find this useful. Feel free to share it with your network and recommend our training which is open to everybody; thanks.