CyberWardens.

The rise of deepfake threats

FACTSHEET

Deepfakes decoded: Protecting your business from AI-driven scams

In the last decade, advancements in Artificial Intelligence (AI) have transformed everyday life.

From virtual assistants to personalised recommendations, AI has become a driving force in technology. However, as handy as it can be, this technology has a darker side.

One of the most concerning developments is the rise of deepfake technology. By creating highly realistic fake videos, audio, or images, deepfakes have raised significant privacy and security concerns—and opened the door to a new wave of cyber crime. In this virtual reality, what you see really isn't always what you get.

What are deepfakes and how do they work?

Deepfakes are highly realistic fake videos, audio, or images created using AI. They rely on advanced machine learning techniques, particularly Generative Adversarial Networks (GANs), to replicate a person's appearance, voice, or mannerisms with astonishing accuracy. Think of deepfakes as digital chameleons—they mimic, adapt, and blend so well they're almost impossible to spot.

The process starts with the AI being trained on a large dataset of photos, videos, or audio recordings of the target individual. The AI learns to mimic their unique features, expressions, and voice patterns. It then generates synthetic media, adding details like synchronised lip movements, realistic lighting, and accurate shadows to create a highly convincing portrayal.

While deepfake technology has legitimate uses in fields like entertainment and education, its misuse has become a significant concern.

How deepfakes threaten small businesses

Deepfakes can impact your business in several ways:

- **Fraud:** Cyber criminals can impersonate leaders or persons in a position of authority to authorise fake financial transactions. It's like having a wolf in sheep's clothing running your accounts—except this wolf knows your CFO's voice.
- **Reputational harm:** Fake videos or images can spread false information about your business or employees. A deepfake video can snowball into a PR nightmare, leaving your business scrambling to repair the damage.
- **Social engineering:** Deepfake phishing scams can deceive employees into sharing sensitive data. They are the ultimate bait-and-switch—tricking employees with what looks like a trusted source.
- **Identity theft:** Criminals use deepfakes to bypass biometric security systems, such as facial recognition.

Deepfakes in action

Deepfake technology has enabled cybercriminals to orchestrate increasingly sophisticated attacks, often targeting businesses and individuals with devastating consequences.

- In early 2024, cyber criminals used AI-generated deepfake technology during a video conference to impersonate senior executives at British engineering company Arup. Employees were duped into believing they were attending a legitimate meeting with the company's Chief Financial Officer and other colleagues. Reassured by the familiar faces on the call, a staff member authorised 15 transactions totalling $25 million USD to accounts controlled by the criminals. Every participant in the meeting, except the employee, was a deepfake recreation.
- Deepfakes have also been used to target politicians and public figures, manipulating their images and voices to spread misinformation and defraud the public. Australian politicians like Penny Wong and Katy Gallagher were featured in deepfake videos promoting fake investment schemes. These AI-generated clips were circulated as ads on platforms like Facebook, reaching thousands of Australians before being removed. One manipulated video falsely depicted Gallagher endorsing an "anti-inflation plan" that promised users $36,000 in monthly returns. Another deepfake of former Prime Minister Scott Morrison promoted a similar get-rich-quick scheme.
- Deepfake technology extends beyond businesses and politics to exploit personal relationships. A French woman fell victim to a scam where a deepfake impersonating actor Brad Pitt convinced her she was in a long-term relationship with the celebrity. The impersonator used AI-generated images and fake documents, including a counterfeit passport, to gain her trust. Over 18 months, Anne transferred €830,000 for supposed medical expenses, only to discover the truth when she saw media coverage of Pitt with his actual partner.

How to spot deepfakes

Deepfakes can look overwhelmingly convincing. Here are some signs to spot deepfake content:

- **Visual inconsistencies:** Look for unnatural eye movements, mismatched lighting, or distorted facial features.
- **Audio clues:** Check for robotic-sounding voices or mismatched lip movements in videos.
- **Contextual errors:** Scrutinise requests that seem out of character or unusual for the person supposedly making them. If something feels 'off', trust your instincts and verify independently.
- **Blurring or pixelation:** Pay attention to areas around the mouth, eyes, and neck. In poorly created deepfakes, these regions often exhibit blurring, cropped effects, or pixelation, especially when the subject moves.
- **Skin inconsistencies or discolouration:** Deepfake technology often struggles to replicate the natural texture and tone of human skin. Look for inconsistencies like overly smooth areas, discolouration, or visible patches that don't align with the rest of the image.
- **Glitches or background changes:** Watch for glitches or inconsistencies in the video. Sections of lower quality, changes in lighting, or background shifts within the same clip are red flags that the content might be fake.

Tips to protect your business from deepfakes

- **Implement multi-factor authentication (MFA):** Ensure critical accounts and systems require multiple verification steps.
- **Train employees:** Educate your team on deepfake risks and how to verify requests before taking action.
- **Verify requests:** Use multiple channels (e.g., phone call or in-person confirmation) to authenticate unusual or sensitive demands.
- **Limit your public exposure:** Reduce the availability of personal media (photos, videos, and voice recordings) on public platforms.
- **Invest in detection software:** Emerging technologies can help spot deepfakes before they cause harm. Use AI tools designed to detect deepfake videos or images.

What to do if you're targeted

- **Report the incident:** Contact **ASD's Report Cyber, Scamwatch,** and **IDCARE** for assistance.
- **Secure your accounts:** Change passwords, enable MFA, log out of accounts, and check for unauthorised changes.
- **Alert stakeholders:** Notify your bank, suppliers, and employees about potential threats.

Deepfake technology will continue to evolve, but so will tools to detect and counteract it. Protecting your business starts with awareness, education, and proactive measures.

Take the first step today by enrolling in the free Cyber Wardens training program at **cyberwardens.com.au.**