La minaccia dei deepfake

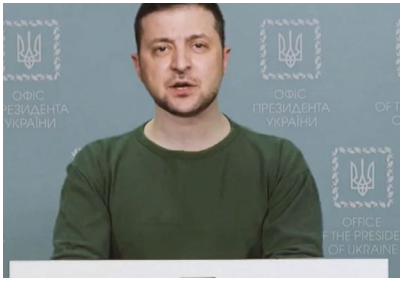
Capire, reagire e proteggersi

La vostra Polizia e la Prevenzione Svizzera della Criminalità (PSC) – un servizio intercantonale della Conferenza delle direttrici e dei direttori dei dipartimenti cantonali di giustizia e polizia (CDDGP) – in collaborazione con l'Istituto di lotta contro la criminalità economica (ILCE) della Scuola universitaria di gestione Arc (Haute école de gestion Arc) di Neuchâtel.

Deepfake – un fenomeno in forte crescita

Un deepfake è un'immagine, un video o una registrazione audio generati o modificati con l'ausilio dell'intelligenza artificiale. Questa tecnica è generalmente utilizzata per imitare in modo realistico la voce o le caratteristiche fisiche di una persona.

Esempi:



2022: video deepfake del presidente ucraino Volodymyr Zelensky che esorta i suoi soldati a deporre le armi.

2023: immagine deepfake di Papa Francesco che indossa un piumino bianco.
(Foto: deepfake generato con Midjourney, software di IA)

I deepfake, utilizzati per ingannare, manipolare o truffare, sono difficili da distinguere dalla realtà e possono diffondersi molto rapidamente sul web e sui social network.

Un deepfake può essere utilizzato, in particolare, per:

- diffondere informazioni false
- manipolare l'opinione pubblica
- usurpare l'identità di una persona
- commettere delle truffe
- creare o diffondere immagini pornografiche alterate
- ledere la reputazione di una persona
- molestare una persona

Possibilità quasi infinite

Gli strumenti dedicati alla creazione di deepfake si stanno moltiplicando e diventano sempre più accessibili. Consentono di realizzare con una certa facilità immagini, video e registrazioni audio che alterano o manipolano la realtà. Questi contenuti, diffusi attraverso canali pubblici e privati (p. es. WhatsApp), possono assumere diverse forme e perseguire vari obiettivi.

- Mettere un personaggio pubblico in una situazione inventata o fargli pronunciare un discorso falso.
- Inserire un personaggio famoso in una pubblicità o in una scena senza il suo consenso.
- Modificare un evento storico o una scena di attualità.
- Creare volti o personaggi fittizi per realizzare dei falsi profili online.
- Alterare qualsiasi dettaglio di un'immagine o di un video.
- Inserire persone in situazioni assurde o ridicole.
- Generare scene fittizie a partire da immagini o video esistenti.
- Simulare la voce di una persona per commettere una truffa.
- Utilizzare il volto di un conoscente per creare contenuti pornografici.
- Ingannare un sistema di riconoscimento vocale o di identificazione video.
- Alimentare una campagna di disinformazione.
- Ecc.

Violazioni della personalità e disinformazione

L'IA viene utilizzata in vari modi per esercitare un'influenza politica. A volte, si tratta di influenzare le opinioni e, di conseguenza, anche le votazioni e le elezioni, diffondendo in modo mirato informazioni false o comunicazioni volutamente unilaterali sui social network. A tale fine si falsificano testi, ma anche immagini e video. In altri casi, l'obiettivo principale è la provocazione. Anche in Svizzera si sono già verificati casi di video falsificati con dichiarazioni non veritiere a scopo diffamatorio.

Esempio 1

Truffe sugli investimenti online

Su una piattaforma online, Paul si imbatte in un riferimento a un articolo del "Blick" in cui Roger Federer svela alcuni suoi segreti e spiega com'è riuscito ad accrescere il proprio patrimonio. Paul sta per andare in pensione, ma riceverà una rendita molto esigua. Sarebbe un'occasione ideale per migliorare la sua situazione finanziaria. Clicca quindi sul link, accede al sito del "Blick" e legge l'intervista. Con sua grande gioia, Roger Federer comunica persino il link alla piattaforma d'investimento. Con soli CHF 250.– può già iniziare ad investire. È credibile?

No! È tutto falso! L'annuncio è stato pubblicato da criminali, la presunta pagina del "Blick" è stata replicata e non ha nulla a che vedere con il quotidiano in questione, mentre i consigli finanziari di Roger Federer sono inventati di sana pianta. La sua foto e anche il suo video sono stati contraffatti con l'IA. E poi, perché mai Roger Federer dovrebbe parlare pubblicamente dei suoi investimenti?
A proposito, ci sono criminali che hanno già anticipato queste domande critiche. Per questo motivo, ci sono ora anche annunci e sedicenti articoli su personaggi famosi che sono stati picchiati perché hanno svelato i loro segreti sugli investimenti. E, l'avrete capito, anche questa è una bufala!

Esempio 2

Truffa dell'amore (Romance Scam)

Dopo una lunga ricerca in Internet, Anita ha finalmente trovato il partner dei suoi sogni, Rolf. Si tratta di un ingegnere che attualmente lavora su una piattaforma petrolifera nell'Atlantico. Rolf si è perdutamente innamorato di Anita e vorrebbe farle visita il prima possibile, ma le circostanze non glielo consentono. Chattano ogni giorno e si scambiano regolarmente delle foto. A un certo punto, Rolf si ammala. Per curarsi deve tornare sulla terraferma. Ma i suoi conti sono bloccati e quindi chiede aiuto ad Anita. Non appena sarà guarito, potrà finalmente andare a trovarla. È credibile?

No! È tutto falso! Rolf è in realtà un criminale dall'aspetto completamente diverso. Parla un'altra lingua, vive in un altro luogo ed è solo interessato ai soldi. Con l'ausilio di assistenti virtuali basati sull'IA, riesce a tradurre i suoi dialoghi in chat nella lingua desiderata. Ha trovato la foto in Internet e l'ha adattata alle situazioni desiderate con l'IA. E, come avrete già intuito, non andrà mai a trovare Anita.

Reagire ai deepfake

I deepfake e altri contenuti manipolati possono apparire molto realistici, anche agli occhi degli specialisti. È quindi fondamentale stare all'erta e dar prova di cautela prima di credere o condividere un'informazione. Occorre procedere con la dovuta prudenza quando si utilizzano strumenti che promettono di identificare con certezza i deepfake. Diventa quindi imprescindibile sviluppare buone pratiche per distinguere i contenuti manipolati presenti nelle informazioni consultate quotidianamente.

Riflettere – verificare – segnalare

1. Sviluppate uno spirito critico

- Diffidate dei contenuti che fanno leva su emozioni forti.
- Non reagite in modo affrettato di fronte a questo tipo di post.
- Chiedetevi chi ha diffuso quel contenuto e per quali motivi.

I deepfake mirano spesso a suscitare paura, rabbia o sorpresa, al fine di provocare reazioni irrazionali.

2. Verificate le fonti (fact-checking)

- Identificate l'autore del contenuto.
- Valutate l'affidabilità della fonte.
- Verificate se l'informazione è riportata da più media autorevoli.

Un'informazione vera viene solitamente divulgata da diverse fonti affidabili.

3. Segnalate il deepfake

- Non condividetelo.
- Segnalatelo alla piattaforma su cui è stato pubblicato.
- Informate il vostro entourage per limitarne la diffusione.

Segnalando ed evitando di condividere un deepfake, contribuite a combattere questo fenomeno.

Cosa fare se siete una vittima?

I deepfake possono prendere di mira direttamente le persone e comportare gravi conseguenze quali violazioni della privacy o danni economici. Ciò si verifica, in particolare, nei casi di usurpazione d'identità, atti persecutori, diffusione di contenuti pornografici o truffe. È quindi di fondamentale importanza limitare la diffusione di tali contenuti e il loro impatto.

- **Non diffondete il deepfake,** nemmeno per avvisare il vostro entourage.
- **Conservate tutte le prove:**
  - schermate
  - link ai contenuti
  - sequenza dei messaggi scambiati
- **Segnalate il deepfake** alla piattaforma per richiederne la rimozione.

Possono entrare in considerazione diversi atti di rilevanza penale, segnatamente:

- Truffa (art. 146 CP)
- Diffamazione (art. 173 CP) e calunnia (art. 174 CP)
- Usurpazione d'identità (art. 179^{decies} CP)
- Atti persecutori (art. 181b CP)
- Pornografia (art. 197 CP), condivisione indebita di contenuti sessuali non pubblici (art. 197a CP) e molestie sessuali (art. 198 CP)

È inoltre possibile intentare azioni civili facendo valere la protezione della personalità (art. 28, 28a et 28b CC).

A seconda della gravità del danno, è possibile sporgere denuncia alla polizia.

Informazioni supplementari e consigli

Se avete dubbi o sospettate la presenza di deepfake, potete consultare il sito dell'Ufficio federale della cibersicurezza (www.ncsc.admin.ch) oppure il sito cybercrimepolice.ch. Entrambi forniscono informazioni aggiornate, consigli per la prevenzione e raccomandazioni pratiche. In questi siti troverete inoltre indicazioni su come sporgere denuncia.

Ulteriori informazioni e consigli sulla cibersicurezza sono pure disponibili sul sito della Prevenzione Svizzera della Criminalità (www.skppsc.ch).

Per reagire ai deepfake

→ sviluppate uno spirito critico
→ verificate le fonti
→ segnalateli

Ogni verifica contribuisce a limitare la diffusione di informazioni false. Non diffondete i deepfake!

Seguiteci sui nostri canali:

Facebook: @Prevenzione Svizzera della Criminalità **Instagram:** @skppsc_svizzera
LinkedIn: @Prevenzione Svizzera della Criminalità **YouTube:** @SKPPSCSCP

Maggiori informazioni: www.skppsc.ch

SKPPSC

Prevenzione Svizzera della Criminalità
Casa dei Cantoni
Speichergasse 6
3001 Berna

www.skppsc.ch

Giugno 2026