Rivista di Psicodinamica Criminale
Registro Stampa del Tribunale di Padova n° 2135 del 30 aprile 2008 ISSN 2037-1195

Deepfake: aspetti criminologici

Anno XVI – n. 1 luglio 2023

RIVISTA DI PSICODINAMICA CRIMINALE

Periodico di saggi, criminologia e ricerche

Anno XVI – n. 1 luglio 2023

Direttore scientifico

Laura Baccaro

Redazione amministrazione

Associazione psicologo di strada

Via Armistizio, 281– Padova

rivistapsicodinamica.criminale@gmail.com

Registro Stampa del Tribunale di Padova n° 2135 del 30.04.2008

Sommario

EDITORIALE ... 3

1. DEEPFAKE: BREVE ANALISI E DESCRIZIONE ... 4

1.1 DeepFake: di cosa parliamo ... 4
1.2 Deepfake, cultura di massa e credibilità ... 5
1.3 Deepfake e gli studi intelligenza artificiale ... 6
1.4 Come riconoscere un video DeepFake ... 7
1.5 Il lato social della tecnologia Deepfake ... 7
1.6 Violenza di genere e pornografia non consensuale ... 8
1.7 Politica e deepfake ... 8
1.8 Deepfake e criminalità informatica ... 9

2. TECNICHE FORENSI ... 10

2.1 L'emozione e l'intelligenza artificiale: applicazioni e rischi ... 10
2.2 Testimonianze oculari e deepfake ... 11
2.3 DARPA: programma di semantica forense ... 11
2.4 Deepfake di vittime di crimini veri ... 12
2.5 Cold case e deepfake di Sedar Soares ... 13

3. ASPETTI GIURIDICI E TUTELE ... 13

3.1 DeepFake e Parlamento Europeo ... 14

3.2 LA SITUAZIONE IN ITALIA ... 14

Garante e intelligenza artificiale ... 14
Garante avvia istruttoria su app che falsifica le voci ... 15
Deepfake e privacy in Italia ... 15
Il fenomeno del deepnude ... 16

4. FACING REALITY? LAW ENFORCEMENT AND THE CHALLENGE OF DEEPFAKES ... 16

Introduction ... 16
Understanding deepfakes ... 17
The technology behind deepfakes ... 19
Deep learning ... 19
Generative Adversarial Networks (GAN) ... 20
Deepfake technology's impact on crime ... 21
Disinformation ... 22
Non-consensual pornography ... 23
Document fraud ... 23
Deepfake as a service ... 24
Deepfake technology's impact on law enforcement ... 25
New capacities needed ... 26
Deepfake detection ... 27
Manual detection ... 27
Automated detection ... 27
How are other actors responding to deepfakes? ... 29
European Union ... 30

Conclusion *31*

5. DEEPFAKE E STATI UNITI **33**

MALICIOUS DEEP FAKE PROHIBITION ACT OF 2018 33

IDENTIFYING OUTPUTS OF GENERATIVE ADVERSARIAL NETWORKS ACT 34

6. COME PROTEGGERSI DAI DEEPFAKE **37**

COME PROTEGGERSI IN AZIENDA 37

BIBLIOGRAFIA **39**

VADEMECUM GARANTE PER LA PROTEZIONE DEI DATI PERSONALI **45**

PER GLI AUTORI **48**

Editoriale

L'avvento dei deepfake ha importanti implicazioni per la scienza e la società.

I deepfake sono foto, video e audio creati grazie a software di intelligenza artificiale (AI) che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce. La tecnologia si è dimostrata capace di sovrapporre i volti di due persone diverse per creare un falso profilo o una falsa identità.

Le immagini sintetiche generate sono così reali da ingannare non solo i nostri occhi, ma anche gli algoritmi usati per riconoscerli. Questo alto grado di realismo dei deepfake e la loro indistinguibilità dai video e dalle immagini originali per la mente umana portano alla percezione dei deepfake come una minaccia per la società umana, la democrazia e il discorso pubblico, nonché un potenziale motore di radicalizzazione, polarizzazione e conflitto.

Gli studi attuali si concentrano principalmente sul rilevamento e sui pericoli dei deepfakee, per il potenziale di disinformazione e uso manipolativo possibile, prestando meno attenzione ai benefici di questa tecnologia.

In Italia il Garante per la protezione dei dati personali ha messo a punto una scheda informativa e un vademecum per sensibilizzare gli utenti sui rischi connessi agli usi malevoli di questa nuova tecnologia, sempre più frequenti, anche a causa della diffusione di app e software che rendono possibile realizzare deepfake, anche molto ben elaborati e sofisticati, utilizzando un comune smartphone.

Sono scaricabili dal sito https://www.garanteprivacy.it/home/docweb/-/docweb-display/docwe...

È in allegato a questo numero il vademecum.

Questo numero vuole essere uno strumento di lavoro, raccoglie e mette a disposizione articoli e documentazione utile.

1. Deepfake: breve analisi e descrizione

1.1 DeepFake: di cosa parliamo

"Immagina questo per un secondo: un uomo, con il controllo totale dei dati rubati di miliardi di persone, tutti i loro segreti, le loro vite, il loro futuro... Devo tutto a Spectre. Spectre mi ha mostrato che chiunque controlli i dati, controlla il futuro.

"Mark Zuckerberg" in un video, pubblicato su Instagram, pronuncia queste parole minacciose…

Ma è un falso Mark, era un deepfake. Era giugno del 2019 quando due artisti americani dopo il caso Nancy Pelosi lanciano una provocazione. L'azione degli artisti sembrerebbe una sfida alle stesse policy della società californiana, che qualche settimana prima aveva rifiutato di rimuovere un video modificato in cui si vedeva la speaker della Camera statunitense e dirigente del Partito Democratico Nancy Pelosi poco lucida. Una clip di poco più di trenta secondi in cui la leader democratica sembra ubriaca. Biascicante, fiacca nella dialettica e nei movimenti che compie a fatica, con lo sguardo a tratti perso nel vuoto. Ma il filmato non è fedele alla realtà, bensì artefatto ad hoc: rallentato di circa il 75% rispetto all'originale.

Basta una foto, una sola immagine e il gioco è fatto. Nei social noi vediamo la Gioconda che ci parla e ci strizza l'occhiolino… ecco quanto la tecnologia è in grado di fare.

In sintesi un deepfake è un falso media, di solito è un video, che sembra molto reale. È l'applicazione della tecnologia AI e dell'apprendimento automatico (deep learning) per manipolare video e voce, creando e falsandone contenuti.

L'inizio della tecnologia Deepfakes è solitamente attribuito a un utente anonimo, noto solo come "u/deepfakes", della piattaforma di social media Reddit. L'utente nel novembre del 2017 crea una community di condivisione e discussione di contenuti chiamata r/deepfakes, ed è qui che hanno iniziato a circolare i primi facewap che utilizzano l'algoritmo Deepfakes. "U/deepfakes" ha caricato un pacchetto di file anonimo tramite un servizio di condivisione, rendendo pubblico l'algoritmo Deepfakes, ovvero i codici. Nel post di Reddit che mostrava il codice non sono chiari i motivi della condivisione, si leggeva che il codice era molto semplice e che non aveva senso tenerlo segreto, specificava che ha utilizzato librerie open source (TensorFlow: software di apprendimento automatico open source gratuito di Google), la ricerca di immagini di Google, i siti Web di social media come Instagram, foto stock, e video di YouTube per creare un algoritmo di apprendimento automatico che gli ha permesso di inserire i volti di personaggi famosi su video preesistenti, fotogramma per fotogramma. Così "u/deepfakes", e come lui altri, nella sezione r/deepfakes di Reddit, condividevano i deepfake che creavano, in molti scambiavano i volti di attrici note (Daisy Ridley, Emma Watson, Ariana Grande, Katy Perry, Taylor Swift o Scarlett Johansson) con quelli di protagonisti di video pornografici. Video che furono cancellati poco tempo dopo.

Il fenomeno dei deepfake fu riportato per la prima volta nel dicembre 2017 dal magazine Vice, dove Samantha Cole pubblicò un primo articolo il quale ha attirato la prima attenzione sui deepfake che venivano condivisi nelle comunità online. Sei settimane dopo, Cole scrisse in un articolo

dell'incremento della falsa pornografia assistita dall'intelligenza artificiale. Anche grazie a questi articoli a febbraio 2018 questo subreddit è stato rimosso da Reddit e le regole del sito sono state aggiornate per vietare quella che venne definita "pornografia involontaria", e anche alcuni altri siti web bannarono a loro volta l'uso dei deepfake per pornografia involontaria, incluso Twitter e il sito Pornhub. Altre comunità online che condividevano deepfake raffiguranti celebrità, politici, e altri in scenari non-pornografici rimasero attive.

Nel 2018 "u/deepfakes" rilascia FakeApp, una piattaforma facile da usare per creare media contraffatti a partire da poche immagini, anche da una sola. L'app utilizza una rete neurale artificiale e permette di creare facilmente video con volti scambiabili. Il programma per "apprendere" quali aspetti dell'immagine si debbano cambiare necessita di parecchio materiale visivo raffigurante la persona "vittima", apprende usando un algoritmo di deeplearning basato su sequenze di video e immagini. Il software è libero e gratuito è ha effettivamente democratizzato il potere dei GAN (Generative Adversarial Networks) facendolo uscire dall'ambito accademico e della ricerca: così chiunque abbia accesso a Internet e alle immagini del volto di una persona potrebbe generare il proprio deep fake. "U/deepfakes" già nel 2018 scriveva anche che esiste un intero settore che studia su come generare immagini realistiche. Profetizzava che tra poco tempo (mesi) la metà degli strumenti di Photoshop si baserà sull'apprendimento automatico. Gli animatori avranno uno strumento basato sull'apprendimento automatico per creare animazioni di personaggi naturali. Alcuni addestreranno il modello per rilevare immagini false e altri addestreranno il modello per creare falsi non rilevabili. Insomma lo scambio di faccia non è nulla paragonabile alla creazione di avatar 3D realistici e inseriti poi nella realtà virtuale.

1.2 Deepfake, cultura di massa e credibilità

Un effetto dei deepfake è che non è più possibile distinguere se il contenuto è mirato (ad esempio satira) o genuino. Ovvero se è credibile e autentico. Immaginiamo che un deepfake venga inserito in una videoconferenza per una riunione o assemblea o consigli d'amministrazione. Può essere un infiltrato per raccogliere informazioni importanti, per seminare informazioni errate, per orientare decisioni, per compiere truffe economiche e depistaggi.

Tutti dovrebbero sapere quanto velocemente le cose possono essere falsificate con questa tecnologia, e che il problema non è tecnico, piuttosto è la fiducia nell'informazione e nel giornalismo. Il problema è che non è più possibile determinare se il contenuto di un media corrisponde alla verità. Si deve rivedere la vecchia idea che le immagini non mentono e i processi sociali attraverso i quali arriviamo collettivamente a conoscere le cose e ritenerle vere o false sono minacciati. Insomma il contatto costante con la disinformazione costringe le persone a smettere di fidarsi di ciò che vedono e sentono. In altre parole, le persone possono arrivare a considerare tutto come un inganno. Un'implicazione delle caratteristiche del DeepFake ingannevole è che è in grado di creare nuove verità non essendo più possibile distinguere tra fatti e verità.

1.3 Deepfake e gli studi intelligenza artificiale

Il termine deepfake deriva dalla tecnologia sottostante "deep learning", che è una forma di intelligenza artificiale, cioè deepfake è una tecnica utilizzata per la sintesi dell'immagine umana basata sull'intelligenza artificiale. In pratica viene utilizzata la tecnologia di apprendimento automatico chiamata Generative Adversarial Networks o GAN (rete antagonista generativa) che è in grado di apprendere da sola il modo in cui modificare i contenuti. Essenzialmente, il software viene esposto a tutta una serie di dati e, sfruttando il deep learning, riesce a risolvere problemi scambiando letteralmente i volti di due individui nei contenuti video e digitali per creare supporti falsi dall'aspetto realistico. Vengono realizzare anche dei "cloni vocali" nel caso dei file audio.

L'intelligenza artificiale produce immagini e video apparenntemente realistici che possono avere un modello di riferimento, es. foto di persone reali ma anche foto di persone create ex novo dall'intelligenza artificiale, ovvero che non esistono fisicamente. Il video prodotto è talmente simile che il soggetto che guarda o ascolta non riesce a distinguere se ciò che percepisce è vero oppure no… è qualcosa che accade davanti ai suoi occhi e pertanto assume un significato di verità, di vero.

"U/deepfakes" nei sui scritti in Reddit affermava che il suo contributo era di scarsa importanza nel causare le applicazioni indesiderate dell'apprendimento automatico, perché erano già tante le tecnologie correlate e studiate dal settore. Un primo progetto accademico di riferimento fu il "Video Rewrite", pubblicato nel 1997. Si mostra la modificazione di un video esistente di una persona che parla, intervenendo in modo che quella persona sembra stia pronunciando le parole appartenenti a una traccia audio differente da quella originale. Un altro progetto accademico pubblicato nel 2017 è "Synthesizing Obama", è la modifica di un video dell'ex presidente americano Barack Obama, raffigurandolo mentre pronuncia parole che appartengono a una traccia audio separata. Il programma "Face2Face" è del 2016 e modifica video raffiguranti la faccia di una persona, raffigurandola mentre imita le espressioni facciali di un'altra persona in tempo reale. Nell'agosto 2018, i ricercatori dell'Università della California, Berkeley pubblicarono un articolo introducendo una fake-dancing app, la quale creava abilità di danza false utilizzando l'AI.

È stata creata una grande quantità di software deepfake e possono essere trovate open souce in comunità di sviluppo software. La creazione di deepfake non è vietata poiché il loro utilizzo è anche per scopi di puro intrattenimento.

Vi sono applicazioni che consentono di sincronizzare il proprio labiale con la musica o con un altro video, per animare piccole GIF, o per animare vecchie foto, per modificare i volti, sostituire l'intera testa del soggetto nel video, invecchiare una persona o alterare il movimento delle labbra; cambiare le espressioni facciali presenti nei video, o aggiungere al volto elementi aggiuntivi, come barba e occhiali.

I deepfake più diffusi sono quelli in cui il volto di un soggetto è sostituito con quello di un altro. Famoso è il deepfake che il MIT Center for Advanced Virtuality in cui l'ex presidente Nixon tiene un discorso annunciando che gli astronauti dell'Apollo 11 non sarebbero mai più tornati dalla luna. (https://moondisaster.org/film/). Il progetto è stato pensato per educare l'opinione pubblica al fatto che i deepfake possono essere molto convincenti. Il risultato finale è impressionante e mostra l'importanza di sviluppare approcci di rilevamento della contraffazione che possano operare diversi tipi di manipolazioni facciali, anche mai incontrati finora.

1.4 Come riconoscere un video DeepFake

Sapere come riconoscere un video DeepFake è senza ombra di dubbio utile in modo da evitare di credere a contenuti discutibili.

Un video DeepFake può talvolta avere problemi di riproduzione, nel senso che il volto modificato della persona non è in grado di seguire movimenti realistici, as es. soggetti che non sbattono le palpebre, o che lo fanno comunque in modo innaturale.

Si devono guardare attentamente i dettagli del video in questione, un aiuto potrebbe essere la luminosità del video, talvolta troppo alta o troppo bassa; porre attenzione all'audio e al movimento delle labbra; anomalie per ciò che riguarda la pelle ed i capelli, volti che sembrano essere più sfocati rispetto all'ambiente in cui sono posizionati. Si tratta comunque di errori che i nuovi deepfake stanno progressivamente risolvendo: ciò rappresenta senz'altro un problema per il futuro, non consentendo più di distinguere, quantomeno a occhio nudo, la differenza tra il contenuto originale e il contenuto generato tramite l'intelligenza artificiale.

Attualmente si sta sperimentando l'uso di caratteristiche legate al modo in cui un soggetto esprime le proprie emozioni in quanto numerosi studi che hanno mostrato la capacità di sistemi di intelligenza artificiale di riconoscere le emozioni nei soggetti .

Sarà sempre più importante considerare il contesto in cui il video appare e quindi analizzare non solo il video stesso, ma tutte le fonti multimediali ad esso collegate: testo, voce, immagini, informazioni accessibili in database e lavorare sull'autorevolezza delle fonti per comprendere quali siano i contenuti falsi.

1.5 Il lato social della tecnologia Deepfake

Per cominciare, la tecnologia deepfake è forse l'ultima forma di ingegneria sociale . Usa i nostri istinti profondi, come la fiducia, per indurci a credere che qualcosa sia reale. La manipolazione della fiducia è una delle ragioni del successo dei deepfake in quanto i falsi si basano sull'utilizzo di voci reali o video di persone.

La capacità di scambiare automaticamente i volti e clonare le voci per creare video sintetici credibili e realistici ha alcune interessanti applicazioni apparentemente non lesive o illegali come ad esempio nel cinema, oppure deepfake usati per l'umorismo e la satira.

Questa tecnologia può rivelarsi molto pericolosa se applicata per creare fake news, bufale e truffe, per compiere atti di cyberbullismo o altri crimini informatici di varia natura. Il rischio effettivo per tutti parte dal fatto che miliardi di persone hanno caricato selfie o altre "impronte digitali" su Google, Facebook o Linkedin. Con i video Deepfake chiunque e tutti potrebbero avere le loro sembianze inserite in uno scenario pornografico.

Ma ci sono anche aspetti positivi e utili.

Nel 2018, l'Illinois Holocaust Museum and Education Center ha creato un nuovo display rivoluzionario utilizzando ologrammi in modo che i visitatori potessero parlare e interagire con i sopravvissuti all'Olocausto, contribuendo a dare vita alle loro storie in modo ancora più efficace rispetto ai video tradizionali. Una società ha accuratamente ricercato e caricato migliaia di clip audio della voce di JFK per creare un deepfake di lui che

legge il discorso di Dallas che avrebbe dovuto pronunciare il giorno del suo assassinio e calcolano che un miliardo di persone l'ha ascoltato e si è confrontato con la storia in un modo completamente nuovo. Questa tecnologia è stata adottata in ambito medico da aziende per restituire la voce ai malati di SLA (sclerosi laterale amiotrofica).

1.6 Violenza di genere e pornografia non consensuale

È stata usata per creare falsi video pornografici ritraenti celebrità e per il revenge porn e rappresenta una grave minaccia per le donne, le quali possono vedere la propria reputazione danneggiata a causa della pornografia involontaria nei deepfake o dal revenge porn e rappresentano un nuovo e degradante mezzo di umiliazione, molestia e abuso.

Sono in aumento anche i casi di sextortion con utilizzo di un video deepfake, truffa economica che consiste nella minaccia di mostrare un video sexy del destinatario. Video deepfake nel quale si vede il destinatario come attore di scene erotiche. Anche se la vittima sa che il video non è reale, potrebbe ritenere di non avere altra scelta che pagare, poiché a volte i video possono essere estremamente realistici. Anche questo non è un problema da poco. Nel gennaio 2021, Avast ha ricercato, identificato e bloccato oltre 500.000 casi di sextortion contro i propri clienti, in tutto il mondo

La maggior parte dei post in r/deepfakes finora sono porno, ma alcuni utenti stanno anche creando video che mostrano le implicazioni di vasta portata di una tecnologia che consente a chiunque disponga di materiale grezzo sufficiente con cui lavorare per posizionare in modo convincente qualsiasi volto in qualsiasi video.

Il caso

Rana Ayyub[1], una giornalista investigativa in India, è stata avvertita da una fonte di un video di sesso deepfake che mostrava il suo volto sul corpo di una giovane donna. Il video si stava diffondendo a migliaia su Facebook, Twitter e WhatsApp, a volte allegato a minacce di stupro o accanto al suo indirizzo di casa.

Ayyub, 34 anni, ha dichiarato di aver subito molestie online per anni. Ma il deepfake sembrava diverso: unicamente viscerale, invasivo e crudele. Ha vomitato quando l'ha visto, ha pianto per giorni e si è precipitata in ospedale, sopraffatta dall'ansia. In una stazione di polizia, ha detto, gli agenti si sono rifiutati di sporgere denuncia e li ha visti sorridere mentre guardavano il falso.

"È riuscito a spezzarmi. È stato travolgente. Tutto quello a cui riuscivo a pensare era il mio personaggio: è questo ciò che la gente penserà di me? lei disse. "Questo è molto più intimidatorio di una minaccia fisica. Questo ha un impatto duraturo sulla tua mente. E non c'è niente che possa impedire che accada di nuovo a me.

1.7 Politica e deepfake

[1] Rana Ayyub, "In India, Journalists Face Slut-Shaming and Rape Threats," New York Times, May 22, 2018, https://www.nytimes.com/2018/05/22/opinion/india-journalists-slut-shaming-rape.html.

Il video deepfake è stato utilizzato anche in politica al fine di destabilizzare leader politici oppure governi o manipolare le elezioni. Un utente di Reddit ha combinato filmati di Hitler con il presidente argentino Mauricio Macri. Il regista Jordan Peele e BuzzFeed hanno rilasciato un deepfake di Barack Obama nel quale definisce Trump un "totale e completo idiota" con l'obbiettivo di aumentare la consapevolezza su come i media sintetici generati dall'IA potrebbero essere usati per distorcere e manipolare la realtà .Gli esperti di politica internazionale si stanno preparando per una futura ondata di fake news basate su video deepfake realizzati molto bene in quanto questa tecnologia consente, dai propagandisti autorizzati dallo stato ai troll, di distorcere le informazioni, manipolare le credenze e di orientare ideologicamente online le comunità.

1.8 Deepfake e criminalità informatica

Le minacce deepfake rientrano in quattro categorie principali: sociali (che alimentano disordini sociali e polarizzazione politica); legale (falsificazione di prove elettroniche); personale (molestie e bullismo, pornografia non consensuale e sfruttamento minorile online); e la sicurezza informatica tradizionale (estorsione, frode e manipolazione dei mercati finanziari).

La tecnologia Deepfake può essere utilizzata per creare nuove identità e rubare le identità di persone reali. Gli aggressori utilizzano la tecnologia per creare documenti falsi o falsificare la voce della loro vittima, il che consente loro di creare account o acquistare prodotti fingendo di essere quella persona.

I passaporti contraffatti con una fotografia deepfake saranno difficili da rilevare. Questi potrebbero quindi essere utilizzati per facilitare molti altri crimini, dal furto di identità e traffico di esseri umani all'immigrazione clandestina e ai viaggi terroristici.

Frode fantasma

La frode fantasma si verifica quando un criminale ruba i dati di una persona deceduta e li impersona a scopo di lucro. L'identità rubata potrebbe essere utilizzata per ottenere l'accesso a servizi e account online o per richiedere cose come carte di credito e prestiti.

Frode dell'account nuovo

Definita anche frode dell'applicazione, la frode di nuovi account comporta l'utilizzo di identità rubate o false allo scopo di aprire nuovi conti bancari. Una volta che un criminale ha aperto un conto, può causare gravi danni finanziari esaurendo le carte di credito o contrarre prestiti che non ha intenzione di rimborsare.

Frode di identità sintetica

La frode di identità sintetica è un metodo di frode più complesso che in genere è più difficile da individuare. Piuttosto che sfruttare l'identità rubata di una singola persona, i criminali cercano informazioni e identità di più persone per creare una "persona" che in realtà non esiste. Questa identità prodotta viene quindi utilizzata per transazioni di grandi dimensioni o nuove richieste di credito.

Truffe

Frodi, identità false e truffe sono facilitate proprio dai deepfake poiché utilizzati da malintenzionati per fornire informazioni fuorvianti sulla criptovaluta, nonché offerte di valute ed investimenti, utilizzando profili di celebrità.

Nel Regno Unito è stato utilizzato un deepfake per interagire con i dipendenti e ingannarli al fine di compiere investimenti finanziari. Il CEO ha ricevuto un'e-mail di spear-phishing presumibilmente dall'amministratore delegato della società madre tedesca dell'azienda. Ciò richiedeva un pagamento di £ 243.000 (circa $ 335.000) da effettuare a un fornitore ungherese entro un'ora. È stata subito seguita da una telefonata dell'amministratore delegato, che ha confermato che il pagamento era urgente e doveva essere effettuato immediatamente. La vittima dice di aver riconosciuto non solo la voce del suo capo e il leggero accento tedesco, ma anche la cadenza e l'attenta enunciazione. Quindi ha felicemente effettuato il pagamento.

In una campagna di phishing sui social media, un video falso di una celebrità potrebbe essere utilizzato per estorcere denaro a vittime inconsapevoli e il deepfake renderebbe più difficile da rilevare come truffa.

I criminali informatici minano e destabilizzano le organizzazioni. Ad esempio, un utente malintenzionato potrebbe creare un falso video di un alto dirigente che ammette attività criminali, come reati finanziari, o fa affermazioni false sull'attività dell'organizzazione.

L'identità digitale che richiede il riconoscimento facciale sarà una delle strade percorse dai truffatori che utilizeranno i deepfake per agire indisturbati le loro le truffe tramite creazione di account falsi.

2. Tecniche forensi

Gli esperti di video forensi sono stati colti di sorpresa da Deepfakes tanto che la tecnologia forense per identificare i falsi digitali sta attualmente lavorando a nuovi metodi di rilevamento per contrastare la diffusione dei deep fake. Con l'intensificarsi della minaccia dei deep fake, aumentano anche gli sforzi per produrre nuovi metodi di rilevamento. Una delle recenti scoperte si è concentrata sui piccolissimi cambiamenti di colore che si verificano nel viso dovuti alla circolazione sanguigna. Il segnale è così minuto che il software di apprendimento automatico non è in grado di rilevarlo, almeno per ora.

I ricercatori stanno studiando come i video falsi potessero essere identificati dalla mancanza di ammiccamento nei soggetti sintetici. Facebook si è anche impegnata a sviluppare modelli di machine learning per rilevare deep fake.

2.1 L'emozione e l'intelligenza artificiale: applicazioni e rischi

Oggi lo sviluppo tecnologico grazie all'intelligenza artificiale, sistemi di machine learning, ma soprattutto all'affective computing, emotion IA che si avvalgono in buona parte di meccanismi di deep learning, consente di varcare la soglia delle nostre emozioni rendendole intelligibili e facendo così cadere la parete antincendio che divideva corporeità e la sfera intima dell'individuo.

La capacità degli algoritmi di riconoscere e manipolare stati d'animo, nuova frontiera dell'intelligenza artificiale, solleva questioni che vanno dal grado di attendibilità, agli eventuali esiti discriminatori,

sino a scenari distopici di un loro uso nella correzione di stati emotivi non funzionali a un dato ambiente.

2.2 Testimonianze oculari e deepfake

La domanda se le prove fabbricate possono indurre false testimonianze oculari diventa fondamentale in ambito criminologico.

Sappiamo dalla psicologia giuridica che le informazioni false possono influenzare le credenze e i ricordi delle persone. La letteratura sui falsi ricordi mostra che le informazioni false o le prove fabbricate possono essere fortemente suggestive; può alterare le convinzioni degli individui e coltivare ricchi falsi ricordi su eventi sia pubblici che personali. Sulla base di studi sulla falsa memoria, Loftus (2003) ha sostenuto che false suggestioni creano una "falsa memoria" che potrebbe indurre le persone a riportare informazioni errate su qualcosa o un evento a cui hanno assistito, o qualcosa che hanno fatto, in risposta a un suggerimento fuorviante. È importante sottolineare che non ci sono conseguenze per la segnalazione di informazioni inesatte.

Ma le immagini fabbricate o le riprese video manipolate possono indurre le persone ad accusare un'altra persona di un fatto-reato che non ha mai fatto? L'esperimento condotto ha confermato che le persone accuseranno falsamente un'altra persona di aver commesso un reato anche quando sanno che la loro testimonianza sarà usata contro l'accusato in un processo penale. I video e le immagini possono essere straordinariamente convincenti e difficili da rilevare, e gli esperti forensi sono sempre più chiamati nei casi penali e civili per determinare se le prove digitali sono state alterate create con la metodologia deepfake.

2.3 DARPA: programma di semantica forense

Il gruppo di ricercatori Darpa (Defence Advanced Research Projects Agency) ha creato il programma Semantic Forensics (SemaFor).

La ricerca del gruppo continua per sviluppare strumenti automatizzati che aiutino gli analisti mentre affrontano l'incombente aumento della manipolazione automatizzata dei media multimodali.

SemaFor cerca di dare agli analisti il sopravvento nella lotta tra rivelatori e manipolatori sviluppando tecnologie in grado di automatizzare il rilevamento, l'attribuzione e la caratterizzazione di risorse multimediali falsificate.

Dal punto di vista della difesa, SemaFor si concentra sullo sfruttamento di una debolezza critica nei generatori di media automatizzati. Attualmente, è molto difficile per un algoritmo di generazione automatizzata ottenere tutta la semantica corretta. Garantire che tutto sia allineato dal testo di una notizia, all'immagine di accompagnamento, agli elementi all'interno dell'immagine stessa è un compito molto arduo. Attraverso questo programma si esplorano le modalità in cui le attuali tecniche per sintetizzare i media falliscono.

Sfruttando parte della ricerca di un altro programma DARPA, il programma Media Forensics (MediFor), gli algoritmi di rilevamento semantico cercheranno di determinare se un asset multimediale è stato generato o manipolato. Gli algoritmi di attribuzione mireranno ad automatizzare l'analisi del fatto che i media provengano da dove dichiarano di provenire, e gli algoritmi di caratterizzazione cercheranno di scoprire l'intento dietro la falsificazione del contenuto.

In realtà, a parere di molti studiosi, affidarsi solo al rilevamento forense informatico per combattere i falsi profondi è difficilmente praticabile a causa della velocità con cui le tecniche di apprendimento automatico possono aggirarli.

2.4 Deepfake di vittime di crimini veri

I video generati dall'intelligenza artificiale, basati su immagini di bambini che hanno subito abusi, stanno diventando popolari su TikTok. È notizia di maggio 2023 che account TikTok hanno pubblicato clip generate dall'intelligenza artificiale di vittime di omicidio e abusi, per lo più bambini, che descrivono la loro scomparsa[2].

Con voce infantile a bambina con giganteschi occhi azzurri e una fascia floreale nel video di TikTok dice "La nonna mi ha chiuso in un forno a 230 gradi quando avevo solo 21 mesi". In sottofondo la melodia lamentosa di "Love Is Gone" di Dylan Mathew. La bambina si presenta come Rody Marie Floyd, una bambina che viveva con la madre e la nonna nel Mississippi. Racconta che un giorno aveva fame e non smetteva di piangere, la nonna la mette nel forno, portandola alla morte. "Per favore, seguimi in modo che più persone conoscano la mia vera storia", dice la bambina alla fine del video.

La bambina nel video, ovviamente, non è reale: è un deepfake AI di vera vittima di crimini. È una creazione generata dall'intelligenza artificiale pubblicata su *@truestorynow*, un account con quasi 50.000 follower che pubblica video di vittime di reati, realmente accaduti nella vita reale, che raccontano le loro storie. La raccapricciante storia che sta raccontando è vera, anche se fino a un certo punto. Il nome del bambino non era Rody Marie, ma Royalty Marie, ed è stata trovata pugnalata a morte e bruciata in un forno nella casa di sua nonna nel Mississippi nel 2018; la nonna, la 48enne Carolyn Jones, è stata condannata per omicidio. Ma Royalty aveva 20 mesi quando è morta, non 21, e a differenza della bambina nel video di TikTok, era nera, non bianca.

Tali imprecisioni sono la norma nel mondo di TikTok. Anche in *nostalgia narratives (*@nostalg1anarratives) - TikTok, un account che pubblica video di intelligenza artificiale delle vittime di veri crimini con 175.000 follower si legge un disclaimer in cui si afferma che il video non utilizza foto reali delle vittime, come un modo per "rispettare la famiglia".

L'account *nostalgia narratives (*@nostalg1anarratives) non racconta solo le storie di famose vittime di omicidio di bambini come Elisa Izquierdo, una bambina di sei anni assassinata dalla madre violenta nel 1995, e Star Hobson , un bambino di un anno assassinato dalla fidanzata di sua madre nel 2020 , ma anche vittime di omicidi adulti come George Floyd e JFK .

La proliferazione di questi video di vittime di crimini reali su TikTok è l'ultima questione etica sollevata dall'immensa popolarità del genere del vero crimine in generale.

Sembrano progettati per innescare forti reazioni emotive, perché è il modo più sicuro per ottenere clic e Mi piace.

In realtà si evidenziano almeno due problemi principali: problemi di privacy e può causare disagio emotivo a coloro che conoscevano la vittima. Questa preoccupazione vale doppiamente per i video con immagini "reali" riviste e modificate da intelligenza artificiale, che raccontano la storia di una vittima dal loro punto di vista e usano il loro nome, presumibilmente senza il consenso della

[2] https://www.rollingstone.com/culture/culture-features/true-crime-tiktok-ai-deepfake-victims-children-1234743895/

famiglia, con effetti incredibilmente inquietanti. Queste situazione ha un reale potenziale di rivittimizzare le persone che sono state vittime prima o i loro famigliari, che ritrovano on line un'immagine del figlio morto che racconta dettaglia cruenti di cosa gli è successo.

2.5 Cold case e deepfake di Sedar Soares

La polizia olandese sta descrivendo il suo uso della tecnologia "deepfake" nel freddo caso dell'omicidio di Sedar Soares nel 2003, una "prima mondiale".

Sedar Soares di Rotterdam nel febbraio del 2003 aveva 13 anni e lanciava palle di neve con i suoi amici in un parcheggio ferroviario quando fu ucciso a colpi di arma da fuoco. Nonostante l'omicidio sia avvenuto in un'area pubblica, nessuno è mai stato catturato per l'omicidio di Sedar. La polizia ha annunciato che il giovane non è stato ucciso, come si era ipotizzato per anni, da un automobilista arrabbiato perché aveva lanciato una palla di neve contro la sua auto. Dopo la denuncia inaspettata di un testimone alla polizia nel 2020, si è rivelato uno scenario completamente diverso. Gli investigatori del National Investigation Communication Team ora credono che una banda di criminale organizzata abbia operato vicino alla stazione della metropolitana e che Sedar sia stato "vittima della violenza della malavita, per pura sfortuna"" e che si trovava semplicemente "nel posto sbagliato al momento sbagliato". La polizia olandese crede che sia stato colpito da un proiettile vagante mentre una banda criminale ne derubava un'altra.

Ora, la polizia vuole rilanciare le indagini nella speranza di rendere giustizia alla famiglia di Sedar e spera che l'utilizzo di un video "deepfake" per ricreare l'immagine del ragazzo possa finalmente aiutare a risolvere il "cold case".

Nel 2022 è stato creato un filmato che sembra riportare in vita l'adolescente.Un attore ha interpretato la parte di Sedar per il corpo, mentre il suo volto è stato successivamente modificato utilizzando deepfake.

L'idea era quella di 'ridare vita' alla vittima, affinché potesse lanciare un appello ad aiutare le indagini. Le immagini della vittima, Sedar Soares, che all'epoca aveva appena 13 anni sono state manipolate con l'ausilio dell'animazione computerizzata. Nel corto su un campo da calcio, vestito con una tuta da ginnastica, si vede Soares, camminare attraverso una fila d'onore composta da familiari, amici e allenatori. Poi, Sedar, i suoi parenti e amici, prima che la sua immagine scompaia dal campo e il video fornisca i dettagli di contatto della polizia, dicono: «Sapete di più? Allora parlate».

Come con la maggior parte dei video deepfaked, i movimenti del ragazzo rianimato sono inquietanti.

Prima che il ragazzo si fermi e lasci la palla a terra, una voce dice: «Qualcuno deve sapere chi ha ucciso il mio caro fratello. Ecco perché è stato riportato in vita per questo film».

Per il momento, la polizia di Rotterdam ha riferito di essere stata contattata da una decina di testimoni, senza però rivelare il contenuto delle dichiarazioni.

3. Aspetti giuridici e tutele

3.1 DeepFake e Parlamento Europeo

In ambito comunitario, sono diverse le preoccupazioni che vedono il DeepFake come una vera e propria minaccia.

Nel 20 gennaio 2021, il Parlamento Europeo ha trattato delle "questioni relative all'in¬terpretazione e applicazione del diritto inter¬nazionale nella misura in cui l'UE è interessata relativamente agli impieghi civili e militari e all'autorità dello Stato al di fuori dell'ambito della giustizia penale".

In questo senso, si parla di minacce ai diritti umani fonda¬mentali, che vedono lo Stato sovrano sull'uso delle tecnologie AI soltanto per scopi di sorveglian¬za. Di conseguenza, il Parlamento Europeo ha chiesto il divieto dell'utilizzo di applicazioni invasive. Tra queste, è stata citata appunto la tecnologia del DeepFake, in quanto può creare incomprensioni e minacce tra i vari Paesi. Basti pensare a un video in cui si ritrae un politico che diffama il Governo di un altro Paese con cui ci sono, o ci sono state, delle questioni. I creatori dovrebbero essere, quindi, obbligati a inserire la dicitura "non originale" sotto al contenuto. Inoltre, sono stati considerati ad alto rischio anche determinati sistemi di identificazione. Capita spesso, infatti, che un app o una piattaforma richiedano di identificarsi tramite il proprio volto. Queste informazioni possono essere talvolta utilizzate in maniera impropria, con conseguente violazione della privacy.

3.2 La situazione in Italia

Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia

I deepfake sono foto, video e audio creati grazie a software di intelligenza artificiale (AI) che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce.

Il Garante per la protezione dei dati personali ha messo a punto una scheda informativa per sensibilizzare gli utenti sui rischi connessi agli usi malevoli di questa nuova tecnologia, sempre più frequenti, anche a causa della diffusione di app e software che rendono possibile realizzare deepfake, anche molto ben elaborati e sofisticati, utilizzando un comune smartphone.

Garante e intelligenza artificiale

https://www.garanteprivacy.it/temi/intelligenza-artificiale. *La sezione contiene una selezione di contenuti in costante aggiornamento*

"Le parole dell'AI" - Una serie di video per raccontare le principali tematiche legate all'intelligenza artificiale e il loro rapporto con la protezione dei dati.

Etica e Intelligenza artificiale nelle parole di Pasquale Stanzione*, Presidente del Garante per la protezione dei dati personali

Deepfake e Deepnude nelle parole di Ginevra Cerrina Feroni, Vice Presidente del Garante per la protezione dei dati personali

Gli assistenti digitali nelle parole di Agostino Ghiglia*, Componente del Garante per la protezione dei dati personali

Riconoscimento facciale e sorveglianza di massa nelle parole di Guido Scorza*, Componente del Garante per la protezione dei dati personali
[CLICCANDO I LINK, SI POSSONO VISIONARE I VIDEO SUL CANALE YOUTUBE DEL GARANTE https://www.youtube.com/GARAntedatipersonaliGP]

Garante avvia istruttoria su app che falsifica le voci

Il Garante per la protezione dei dati personali ha aperto un'istruttoria nei confronti della società che fornisce la app Fakeyou che, da notizie di stampa, consentirebbe di riprodurre file di testo mediante voci false, ma realistiche, di personaggi noti, anche italiani (*12 ottobre 2022).*

Le preoccupazioni del Garante si indirizzano verso i potenziali rischi che potrebbero determinarsi da un uso improprio di un dato personale, quale è appunto la voce.

L'Autorità ha dunque chiesto alla società "The Storyteller Company - Fakeyou" di trasmettere con urgenza ogni possibile elemento utile a chiarire l'iniziativa.

La società dovrà, tra l'altro, fornire le modalità di "costruzione" della voce dei personaggi famosi, il tipo di dati personali trattati, nonché le finalità del trattamento dei dati riferiti ai personaggi noti e agli utenti che utilizzano l'app.

La società, inoltre, dovrà indicare l'ubicazione dei data center che archiviano i dati personali, sia con riferimento agli utenti registrati dall'Italia, sia ai personaggi noti, e le misure tecniche ed organizzative adottate per garantire un livello di sicurezza adeguato al rischio.

Deepfake e privacy in Italia

A ottobre 2022, il Garante Privacy ha aperto un'istruttoria nei confronti della società fornitrice della app Fakeyou, che genera, a partire da articoli di stampa, realistici file vocali a partire dalle voci di personaggi pubblici noti. Ciò che l'Autorità teme, nel caso di specie, è che l'utilizzo di detto sistema possa ingenerare elevati rischi, conseguenti a un uso improprio di un dato personale come la voce.

Sempre per utilizzo abusivo del deepfake, nel maggio 2021 in Corea sono stati arrestati 94 ragazzi tra i 18 e i 20 anni con l'accusa di aver compiuto diversi reati legati alla diffusione di video pornografici che ritraevano oltre 100 vittime di giovanissima età (tra i 10 e i 20 anni) nell'atto di compiere atti osceni. L'evento di cronaca è stato descritto come un vero e proprio furto di identità realizzato mediante il deepfake.

Sulla tematica, già nel 2020 il Garante Privacy affermava proprio che "Quella realizzata con i deepfake è una forma particolarmente grave di furto di identità. Le persone che compaiono in un deepfake a loro insaputa non solo subiscono una perdita di controllo sulla loro immagine, ma sono private anche del controllo sulle loro idee e sui loro pensieri, che possono essere travisati in base ai discorsi e ai comportamenti falsi che esprimono nei video. Le persone presenti nei deepfake potrebbero inoltre essere rappresentate in luoghi o contesti o con persone che non hanno mai frequentato o che non frequenterebbero mai, oppure in situazioni che potrebbero apparire compromettenti. In sostanza, quindi, un deepfake può ricostruire contesti e situazioni mai effettivamente avvenuti e, se ciò non è voluto dai diretti interessati, può rappresentare una grave minaccia per la riservatezza e la dignità delle persone."

Il fenomeno del deepnude

"In particolari tipologie di deepfake, dette deepnude, persone ignare possono essere rappresentate nude, in pose discinte, situazioni compromettenti (ad esempio, a letto con presunti amanti) o addirittura in contesti pornografici. Con la tecnologia del deepnude, infatti, i visi delle persone (compresi soggetti minori) possono essere "innestati", utilizzando appositi software, sui corpi di altri soggetti, nudi o impegnati in pose o atti di natura esplicitamente sessuale. È anche possibile prendere immagini di corpi vestiti e "spogliarli", ricostruendo l'aspetto che avrebbe il corpo sotto gli indumenti e creando immagini altamente realistiche".

Tant'è che proprio nel 2021, preoccupato dall'attività illecita a sfondo pornografico legata al deepfake il Garante aveva aperto un'istruttoria nei confronti di Telegram, per la elevata diffusione sulla piattaforma di video di deepnude, che esponevano le vittime a gravi lesioni della dignità e delle privacy, "considerati anche il rischio che tali immagini vengano usate a fini estorsivi o di revenge porn e tenuto conto dei danni irreparabili a cui potrebbe portare una incontrollata circolazione delle immagini, fino a forme di vera e propria viralizzazione. La facilità d'uso di questo programma rende, peraltro, potenzialmente vittime di deepfake chiunque abbia una foto sul web."

I sistemi di deepfake, infatti, sintetizzando i dati biometrici dei soggetti designati, svolgono un trattamento di dati assolutamente sensibile, che porta il sistema a poter ricollegare una persona specifica a un volto perfettamente ricostruito nella sua tridimensionalità, o a una voce.

Ne consegue la possibilità di utilizzare il deepfake anche per attività illecite mediante il c.d. spoofing (ossia il furto di informazioni mediante la falsificazione di identità di persone o dispositivo, in modo da ingannare altre persone o dispositivi e ottenere la trasmissione di dati).

4. Facing reality? Law enforcement and the challenge of deepfakes

An Observatory Report from the Europol Innovation Lab. Publications Office of the European Union, Luxembourg[3]. Publications Office of the European Union, 2022

Introduction

Today, threat actors are using disinformation campaigns and deepfake content to misinform the public about events, to influence politics and elections, to contribute to fraud, and to manipulate shareholders in a corporate context. Many organisations have now begun to see deepfakes as an even bigger potential risk than identity theft (for which deepfakes can also be used), especially now

[3] PDF | ISBN 978-92-95220-40-9 | ISSN 2600-5182 | DOI: 10.2813/08370 | QL-AS-22-001-EN-N Neither the European Union Agency for Law Enforcement Cooperation nor any person acting on behalf of the agency is responsible for the use that might be made of the following information. © European Union Agency for Law Enforcement Cooperation, 2022

that most interactions have moved online since the COVID-19 pandemic. This concern is echoed by a recent report by University College London (UCL) that ranks deepfake technology as one of the biggest threats faced by society today[4].

This poses a risk to EU citizens. Europol, as the criminal information hub for law enforcement organisations, will continue to play its part in supporting law enforcement authorities in the EU Member States to counter this threat.

This report presents the first published analysis of the Europol Innovation Lab's Observatory function, focusing on deepfakes, the technology behind them and their potential impact on law enforcement and EU citizens. Deepfake technology uses Artificial Intelligence to audio and audio-visual content. Deepfake technology can produce content that convincingly shows people saying or doing things they never did, or create personas that never existed in the first place.

To date, the Europol Innovation Lab has organised three strategic foresight activities with EU Member State law enforcement agencies and other experts. During strategic foresight activities conducted by the Europol Innovation Lab, over 80 law enforcement experts identified and analysed the trends and technologies they believed would impact their work until 2030. These sessions showed that one of the most worrying technological trends is the evolution and detection of deepfakes, as well as the need to address disinformation more generally. The findings in this report are the result of extensive desk research supported by research provided by partner organisations, expert consultation, and the strategic foresight activities.

Those workshops provided the initial input for this report. Furthermore, the findings are the result of extensive desk research supported by research provided by partner organisations, expert consultation and the strategic foresight activities conducted by the Europol Innovation Lab.

Strategic foresight and scenario methods offer a way to understand and prepare for the potential impact of new technologies on law enforcement. The Europol Innovation Lab's Observatory function monitors technological developments that are relevant for law enforcement and reports on the risks, threats and opportunities of these emerging technologies.

Understanding deepfakes

Disinformation is being spread with the intention to deceive. Tools of disinformation campaigns can include deepfakes, falsified photos, counterfeit websites and other information taken out of context to deceive the audience[5].

In the original, strict sense, deepfakes are mostly disseminated with malicious intent, although they are now often used for positive applications too.[6] Experts estimate that as much as 90 %[7] of online content may be synthetically generated by 2026. Synthetic media refers to media generated or manipulated using artificial intelligence (AI). In most cases, synthetic media is generated for gaming, to improve services or to improve the quality of life, but the increase in synthetic media and improved technology has given rise to disinformation possibilities, including deepfakes.

[4] UCL – London's Global University, ''Deepfakes' ranked as most serious AI crime threat', https://www.ucl.ac.uk/news/2020/aug/deepfakes-ranked-most-serious-ai-crime-threat.

[5] Die Bundesregierung, 'What is disinformation?', accessed 15 March 2022, https://www. bundesregierung.de/breg-de/themen/umgang-mit-desinformation/disinformation- definition-1911048 .

[6] ENLETS, 'SYNTHETIC REALIT Y & DEEP FAKES: IMPACT ON POLICE WORK', 2021, accessed on 15 March 2022, https://enlets.eu/wp-content/uploads/2021/11/Final-Synthetic-Reality- Deep-fakes-Impact-on-Police-Work-04.11.21.pdf.

[7] Schick, Nina, Deepfakes: The Coming Infocalypse: What You Urgently Need To Know, Twelve, Hachette UK, 2020.

Deepfakes were examined and discussed at great length in one of the Europol Innovation Lab's strategic foresight activities. Law enforcement experts who participated in these activities expressed concern about the consequences of disinformation, fake news and social media on political and social discourse. These trends are expected to become more pronounced as the supporting technologies, such as deepfakes, are becoming more sophisticated. Their impact on privacy and personal security will doubtless result in new categories of crime that will have to be policed. Participants were especially concerned about the weaponisation of social media and the impact of misinformation on public discourse and social cohesion.

On a daily basis, people trust their own perception to guide them and tell them what is real and what is not. This applies not only to people in their private lives, but also law enforcement officers trying to do their jobs. First-hand accounts are valued higher than second-hand versions of an event. Auditory and visual recordings of an event are often treated as a truthful account of an event. Photographs and videos are important intelligence for police work and evidence in court. But what if these media can be generated artificially, adapted to show events that never took place, to misrepresent events, or to distort the truth?

For instance, prior to the invasion of Ukraine by Russia in 2022, the United States revealed a Russian plot to use deepfake video to justify an invasion of Ukraine.[8] After the invasion happened, officials of the Ukrainian government warned that Russia might spread deepfakes that will show the Ukrainian president Volodymyr Zelenskyy surrendering.[9] This fear appears to have become reality after hackers made a Ukrainian news website show a video of president Zelenskyy telling his soldiers to surrender.[10] At the time of writing much is still unclear about the video and it has not been verified to be a real deepfake or another fake, but it does show how the use of (deep)fakes are being used for disinformation purposes.

Examples like the one above show that this type of disinformation can be dangerous. Its aim is to intensify existing conflicts and debates, undermine trust in state-run institutions and stir up anger and emotions in general. The erosion of trust is likely to make the business of policing harder.

This challenge to policing is coupled with a public that seems relatively uninformed about the dangers of deepfakes. Despite their increasing prevalence at the time, research in 2019 showed almost 72% of people in a UK survey to be unaware of deepfakes and their impact.[11] This is particularly worrying as people might be unable to identify deepfakes (videos, photos, audios) since they are not aware of the existence of such virtual forgeries or how they work. The lack of understanding of the basics of this technology presents various challenges, some of which are relevant for law enforcement (such as disinformation and document fraud). Even more worrying results from recent experiments have shown that increasing awareness of deepfakes may not improve the chances for

[8] CBS News, 'U.S. reveals Russian plot to use fake video as pretense for Ukraine invasion', 2022, accessed on 10 March 2022, https://www.cbsnews.com/news/russia-disinformation-video- ukraine-invasion-united-states/.

[9] Metro, 'Ukraine warns Russia may deploy deepfakes of Volodmyr Zelensky surrendering', 2021, accessed on 15 March 2022, https://metro.co.uk/2022/03/04/ukraine-warns-russia- may-deploy-deepfakes-of-zelensky-surrendering-16217350.

[10] National Public Radio, 'Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn', 2021, accessed on 17 March 2022, https://www.npr. org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation- ukraine-russia.

[11] iProov, 'Almost Three-Quarters of UK Public Unaware of Deepfake Threat, New Research', 2019, accessed 15 March 2022, https://www.iproov.com/press/uk-public-deepfake-threat.

people to detect them.[12] Researchers are therefore expecting criminals to increase their use of deepfakes in the coming years.[13] This shows it is vital to understand the deepfake threat and prepare ourselves.

The technology behind deepfakes

Deepfake technology uses the power of deep learning technology to audio and audio-visual content. Employed properly, these models can produce content that convincingly shows people saying or doing things they never did, or create people that never existed in the first place. The rise of the application of AI to generate deepfakes is already having, and will have, further implications for the way people treat recorded media. Here we discuss two core advancements behind deepfake technology, namely deep learning and generative adversarial networks, and how 5G technology may further enable the use of deepfakes.

Deep learning

Deep learning is a kind of machine learning where a computer analyses datasets to look for patterns with the help of neural networks. Machine learning is an application of AI where computers automatically improve through the use of data. Deep learning is a kind of machine learning that applies neural networks. These neural networks mimic the way our brains work to more effectively learn from the data provided. Deep learning technology, paired with the availability of large databases with material to train the generative models on, has allowed for rapid improvement of deepfake technology.

Deep learning algorithms use neural networks that mimic the brain's processes to find patterns in data.[14] Therefore, the availability of data is essential for a good deepfake system; it needs examples to learn what the result has to look like. It will try to discover patterns in the available data and thus extract what features are important and how these relate to each other. That will allow it to construct a complete and convincing picture. Depending on the quality of the available data and the factors the algorithm uses, the result may be more or less realistic.

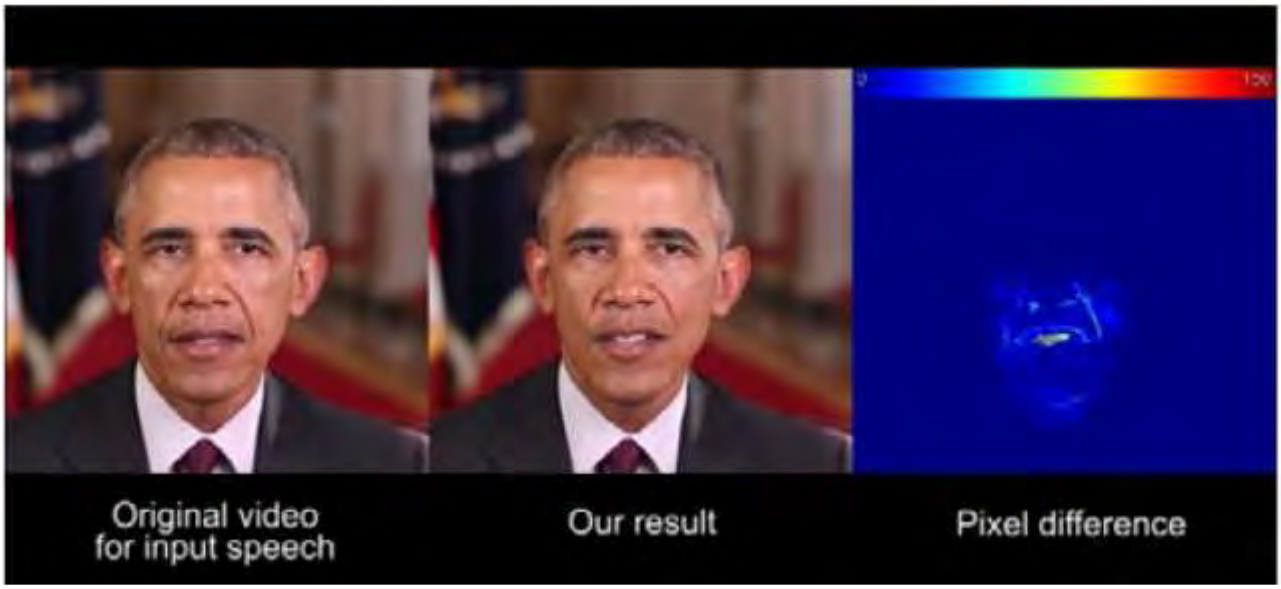
Today, large datasets with labelled visual material are becoming freely available on the internet. These datasets are essential for the training of the machine learning algorithms needed to produce deepfakes. Creators of deepfakes can use these freely available datasets on the internet and avoid the time-consuming work of creating datasets themselves.

In one example from 2018, filmmaker Jordan Peele and BuzzFeed CEO Jordan Peretti created a deepfake video to warn the public about disinformation, specifically regarding the public's perception of political leaders.

Peele and Peretti used free tools with the help of editing experts to overlay Peele's voice and mouth over a pre-existing video of Barack Obama. In the video, Obama allegedly said, "We are entering an

[12] Köbis, N.C. et al., 'Fooled twice: People cannot detect deepfakes but think they can', iScience, 24(11), 2021, accessed 15 March 2022, https://doi.org/10.1016/j.isci.2021.103364.
[13] Recorded Future, Insikt Group, 'The Business of Fraud: Deepfakes, Fraud's Next Frontier', 2021.
[14] Code Academy, 'What Is Deep Learning?', 2021, accessed on 10 March 2022, https://www.codecademy.com/resources/blog/what-is-deep-learning/.

era in which our enemies can make it look like anyone is saying anything, at any point in time. Even if they would never say those things."[15]



Source: Suwajanakorn, S. et al., 2017, 'Synthesizing Obama: learning lip sync from audio', ACM Transactions on Graphics, 36(4), accessed on 15 March 2022, https://dl.acm.org/doi/10.1145/3072959.3073640.

Generative Adversarial Networks (GAN)

A great leap in the quality and accessibility of deepfake technology was made by the adaptation of generative adversarial networks (GANs) as proposed in 2014 by Ian Goodfellow et al.[16] A GAN works with two competing models: a generative and a discriminating model. The generative model creates content based on the available training data, trying to capture the data as closely as possible, to create content that most closely mimics the examples in the training data. A discriminative model then tests the results of the generative model by assessing the probability the tested sample comes from the dataset rather than the generative model.

With the results from these tests, the models continuously improve until the generated content is just as likely to come from the generative model as the training data. This powerful method both simplifies the learning process, making it more accessible, and also improves the outcome by incorporating a mechanism designed to minimise the chance its product would be discriminated from authentic content.

When a new feature that may help discriminate between synthetic and authentic content is discovered, it allows for an easy incorporation of that feature. For example, people's eyes would not blink in early deepfake videos, making them relatively easy to detect.[17] Even though the training data for deepfake models included many pictures of people, these people generally did not blink in pictures. Adding more videos with people blinking to the database allowed both models to work together to

[15] Ars Electronica, 'Obama Deep Fake', 2018, accessed on 10 March 2022, https://ars.electronica.art/center/en/obama-deep-fake/.

[16] Goodfellow, I. et al, (2014), Generative Adversarial Nets (PDF). Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp.2672–2680.

[17] GIZMODO, 'Most Deepfake Videos Have One Glaring Flaw', 2018, accessed on 10 March 2022, https://gizmodo.com/most-deepfake-videos-have-one-glaring-flaw-1826869949.

produce people with blinking eyes, making the result more realistic and consequently harder to differentiate from authentic content.

Training data to create deepfakes may be applied in various ways for video and image deepfakes:

Face swap

Transfer the face of one person for that of the person in the video;

Attribute editing

Change characteristics of the person in the video, e.g. style or colour of the hair;

Face re-enactment

Transferring the facial expressions from the face of one person onto the person in the target video;

Fully synthetic material

Real material is used to train what people look like, but the resulting picture is entirely made up.

See for example https://www.thispersondoesnotexist.com and https://generated.photos

Optimising these factors will improve the outcome. The more extensive the database and the more complex the algorithm becomes, the more computing power is necessary. Generating quality data requires a large volume and diversity of data with enough examples of similar but slightly different representations of the same characteristics to work. For example, if a database mostly contains pictures of white men with black hair, it will not perform too well on creating Asian women with blonde hair. As an increasing number and volume of databases are available, the quality and quantity of training data increases. This has allowed the models generating deepfakes to increase in sophistication.

Participants in the Innovation Lab's foresight activities noted how the roll-out of 5G would enhance connectivity and communication within law enforcement agencies (LEAs) and would strengthen

the privacy and security of organisations and individuals alike. However, they noted that those same benefits would be leveraged by criminals to perpetrate their crimes. The additional bandwidth offered by new communication technologies, such as 5G, enables users to utilise the power of cloud computing to manipulate video streams in real time. Deepfake technologies can therefore be applied in videoconferencing settings, live-streaming video services and television.

Deepfake technology's impact on crime

Participants of the foresight activities cited several trends that European LEAs should be sensitive to. Of note is crime as a service (CaaS), with criminals selling access to the tools, technologies and knowledge to facilitate cyber and cyber-enabled crime. CaaS is expected to evolve in parallel with current technologies, resulting in the automation of crimes such as hacking and adversarial machine learning and deepfakes. Indeed, participants flagged the tendency of criminal actors to become early adopters of new technologies.

As a result, they are always one step ahead of law enforcement in their implementation, use and adaptation of these technologies.

The growing availability of disinformation and deepfakes will have a profound impact on the way people perceive authority and information media. With the increasing volume of deepfakes, trust

in authorities and official facts is undermined. Experts fear this may lead to a situation where citizens no longer have a shared reality, or could create societal confusion about which information sources are reliable; a situation sometimes referred to as 'information apocalypse' or 'reality apathy'.[18]

This makes it essential to be aware of this manipulation and be prepared to deal with the phenomenon, so as to distinguish between benign and malicious use of this technology.

The 'Malicious Uses and Abuses of Artificial Intelligence' report by Europol, TrendMicro and UNICRI[19] included a case study on this topic.

The report also shows that deepfake technology can facilitate various criminal activities, including:

- harassing or humiliating individuals online;
- perpetrating extortion and fraud;
- facilitating document fraud;
- falsifying online identities and fooling 'know your customer' mechanisms[20];
- non-consensual pornography;
- online child sexual exploitation;
- falsifying or manipulating electronic evidence for criminal justice investigations;
- disrupting financial markets;
- distributing disinformation and manipulating public opinion;
- supporting the narratives of extremist or terrorist groups;
- stoking social unrest and political polarisation.

Disinformation

Disinformation campaigns are operations to deliberately spread false information in order to deceive.[21] One major concern about this use is the ease of creating a fake emergency alert that warns of an impending attack. Another concern is the disruption of elections or other aspects of politics by releasing a fake audio or video recording of a candidate or other political figure. To illustrate this, the BBC created a video for the 2019 general election in the UK in which the candidates Boris Johnson and Jeremy Corbyn endorsed each other.[22] If this kind of manipulation successfully deceives a large enough part of the populace, this could have a serious impact on the outcome of an election.

Businesses are also at risk of being targets of disinformation, as deepfakes can be used to generate false information that could fool the public. For example, a threat actor could create a deepfake that makes it appear that a company's executive engaged in a controversial or illegal act. Certain deepfakes could be used for false advertising and disinformation, which could lead to bad publicity for a targeted company. Such applications of deepfakes could impact areas like stock market and

[18] The Guardian, 2018, accessed on 10 March 2022, 'An information apocalypse is coming. How can we protect ourselves?', https://www.theguardian.com/commentisfree/2018/mar/16/an-information-apocalypse-is-coming-how-can-we-protect-ourselves.

[19] Europol, 'Malicious Uses and Abuses of Artificial Intelligence', 2020, accessed on 10 March 2022, https://www.europol.europa.eu/publications-events/publications/malicious-uses-and-abuses-of-artificial-intelligence.

[20] KYC stands for Know Your Customer and refers to the processes for identity verification and fraud risk assessment used by institutions.

[21] Marriam-Webster, 'Disinformation', accessed on 10 March 2022, https://www.merriam-webster.com/dictionary/disinformation

[22] BBC News, 'The fake video where Johnson and Corbyn endorse each other', 2019, accessed on 10 March 2022, https://www.bbc.com/news/av/technology-50381728.

company value as the public (stakeholders and shareholders, as well as consumers) may believe the deepfake and start selling their stocks or boycotting the company.

One example that shows the potential for criminal activities supported by deepfakes is the case where criminals used deepfake audio to impersonate the CEO of a company to make an employee transfer USD 35 million.[23] In this chapter/section of the report, we will look closer at four of the criminal uses of deepfakes that participants in the foresight activities identified.

Non-consensual pornography

In a December 2020 study, Sensity, an Amsterdam-based company that detects and tracks deepfakes online, found 85 047 deepfake videos on popular streaming websites, with the number doubling every 6 months.[24] In a previous September 2019 study, Sensity discovered that 96 % of the fake videos involved non-consensual pornography. To create this, one will overlay a victim's face onto the body of a pornography actor, making it appear that the victim is engaging in the act. In many situations, the victims of pornographic deepfakes are celebrities or high-profile individuals.

These videos are popular, having received approximately 134 million views at the time[25], and there are several pornographic sites that specifically produce pornographic celebrity deepfakes. Perpetrators often act anonymously, making crime attribution more difficult.

Document fraud

Passports are becoming increasingly hard to forge with modern fraud prevention measures. Synthetic media and digitally manipulated facial images present a new approach for document fraud. Using different methods and tools, it is possible to combine, or morph, the faces of the person the passport actually belongs to and the person(s) wanting to obtain a passport illegally. This method may increase the chance that the photo in a forged document passes any identity checks including those using automated means (facial recognition systems).[26]

[23] Forbes, 'Fraudsters Cloned Company Director's Voice In $35 Million Bank Heist, Police Find', 2021, accessed on 16 March 2022, https://www.forbes.com/sites/ thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions.

[24] Sensity, 'How to Detect a Deepfake Online: Image Forensics and Analysis of Deepfake Videos', 2021, accessed on 10 March 2022, https://sensity.ai/blog/deepfake-detection/how- to-detect-a-deepfake/.

[25] Government Technology, 'Deepfakes Are on the Rise — How Should Government Respond?', 2020, accessed on 10 March 2022, https://www.govtech.com/policy/deepfakes-are-on-the- rise-how-should-government-respond.html.

[26] Robertson, D.J., Mungall, A., Watson, D.G. et al, 'Detecting morphed passport photos: a training and individual differences approach,' Cogn. Research 3, 27, 2018, accessed on 16 August 2021, https://doi.org/10.1186/s41235-018-0113-8. MIT Technology Review, 'The hack that could make face recognition think someone else is you', 2020, accessed on 10 March 2022, https://www.technologyreview. com/2020/08/05/1006008/ai-face-recognition-hack-misidentifies-person.Pikoulis, E.-V. et al., 'Face Morphing, a Modern Threat to Border Security: Recent Advances and Open Challenges', Applied Sciences, 2021, accessed 17 February 2022, at https://www. mdpi.com/2076-3417/11/7/3207.

The face in the middle of the image above is an example of a digitally manipulated facial image made using this 'morphing' method from the other two images. The images on the left and right are from The SiblingsDB, which contains different datasets depicting images of individuals related by sibling relationships. The subjects are voluntary students and employees of the Politecnico di Torino and their siblings, in the age range between 13 and 50.[27]

This kind of approach to fraud can be applied to any other type of digital identity check that requires visual authentication. It greatly undermines identity verification procedures since there is no reliable way to detect this kind of attack.[28]

Document fraud is a facilitator of other crimes like illegal immigration, trafficking in human beings, selling of various illegal goods, and terrorism, as perpetrators often use fake IDs to travel to their target locations. Deepfake technology might amplify the risk for advanced document fraud by organised crime groups.

In practice, the robustness of any identification process will depend on the process as a whole, and not only its visual step(s). However, a higher quality synthetic image will make a forged document more likely to pass the check of a visual identification step in the process. In general, the prospect of a successful document fraud attempt depends on quality and context of the deepfake used. The quality of the deepfake is largely dependent on available data and processing power, which is beyond the control of the identification process. The context in which the deepfake is applied is partially determined by the process however, providing opportunities to limit the success of just using a good deepfake.

Deepfake as a service

Just like many other new technologies, deepfakes are still used mainly by proficient engineers and research parties. However, deepfake capabilities are becoming more accessible for the masses

[27] T.F. Vieira, A. Bottino, A. Laurentini, M. De Simone, 'Detecting Siblings in Image Pairs', The Visual Computer, 2014, vol 30, issue 12, p. 1333-1345, doi: 10.1007/s00371-013-0884-3

[28] University of Lincoln ScienceDaily, 'Two fraudsters, one passport: Computers more accurate than humans at detecting fraudulent identity photos,' 2019, accessed on 20 July, 2020, at www.sciencedaily.com/releases/2019/08/190801104038.htm. Naser Damer, PhD. (n.d.). Fraunhofer IGD, 'Face morphing: a new threat?' accessed on 20 July 2020, at https://www.igd.fraunhofer.de/en/press/annual-reports/2018/face-morphing- a-new-threat. David J. Robertson, et al. 'Detecting morphed passport photos: a training and individual differences approach,' Springer Nature, 2018, accessed on 20 July 2020, at https:// cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-018-0113-8 .
Robin S.S. Kramer, et al., 'Face morphing attacks: Investigating detection with humans and Computers, Springer Nature, 2019, accessed on 20 July 2020, at https://link.springer.com/ article/10.1186/s41235-019-0181-4.

through deepfake apps and websites. There are special marketplaces on which users or potential buyers can post requests for deepfake videos (for example, requests for non-consensual pornography). The increased demand for deepfakes has also led to the creation of several companies that deliver deepfakes as a product or even online service. Recorded Future has reported a threat actor's willingness to pay USD 16 000 for this kind of service.[29]

Since deepfakes are based on advanced AI and machine learning technologies, a high level of expertise is required to put the technology together. Accordingly, there are not as many threat actors with the skillset to develop them on their own as there are who would be interested in deepfakes as a service. Those who know how to leverage sophisticated AI can perform the service for others, enabling threat actors to manipulate a person's face and/or voice without understanding the intricacies behind how it works. Then they can conduct advanced social engineering attacks on unsuspecting victims, with the aim to make a sizable profit. Platforms offering these kinds of services have already started to emerge.[30]

Deepfake technology's impact on law enforcement

Law enforcement agencies will be adversely impacted by the rise of synthetic media and deepfakes. While they may provide some opportunities to benefit society, this report focuses on the malicious use of deepfakes. Adverse effects not only include the criminal uses described in the previous chapter, but also the more general impact of deepfakes on society. During foresight activities conducted by Europol, participants discussed how certain technologies could impact law enforcement. In relation to deepfakes, law enforcement agencies may even be forced into action, possibly the wrong action, by misinformation.

Impact on police work

Altered material on social media about events such as demonstrations may lead to police coming into action where it is not necessary, or in the wrong place. In police investigations, law enforcement may chase the wrong suspect of a crime when a deepfake version of the suspect fleeing a crime scene goes viral on social media, thereby giving the suspect the opportunity to get away.

Using deepfakes, people could falsely portray police officers committing transgressions in order to discredit the police or even incite violence against officers. In a time where distrust in authorities is growing, deepfakes and manipulated footage may be used to negatively affect public opinion. The impact of such images and footage is not to be underestimated, especially when this is combined with doxxing (exposing the identity of) the officers supposedly involved.

Impact on the legal process

In court, audio-visual evidence is usually trusted to be an authentic representation of events. Whether the file is extracted from the phone of a suspect, downloaded from social media, or received from the CCTV system of a shop near the crime scene, the authenticity of the scene depicted is not usually questioned.

[29] Biometric update, 'Dark news from dark web: deepfakers are getting their act together', 2021, accessed on 16 March 2022, https://www.biometricupdate.com/202105/dark-news-from- dark-web-deepfakers-are-getting-their-act-together.
[30] Europol, 'Malicious Uses and Abuses of Artificial Intelligence', 2020, accessed on 10 March 2022, https://www.europol.europa.eu/publications-events/publications/malicious-uses-and-abuses-of-artificial-intelligence.

With the rise of deepfakes, it will become increasingly important to scrutinise such content and verify if it is real or somehow artificially manipulated or generated.

Cross-checking footage will become even more important. It calls for a thorough vetting of digital evidence with specific attention to show it can be trusted to be authentic. A consistent and transparent chain of custody of digital evidence to prove no one could have doctored the evidence in the investigation is essential. For instance, as part of a child custody case, the mother of a child tried to convince the court that her husband behaved violently.

She manipulated an audio recording of the man to make it look like he was making threats. Although this was not a real deepfake, it raises questions and concerns.[31] What if the manipulated footage remained unproven as fake?

With lighter-weight neural network structures and advances in hardware, training and generating time will be significantly reduced. In the near future, deepfake software will likely be able to generate full body deepfakes, real-time impersonations, and the seamless removal of elements within videos. The most recent algorithms can deliver increasingly higher levels of realism and run in near real time.

New capacities needed

Claims as to the use of deepfake material will require further law enforcement assessment, leading to new cases and new types of work. This will result in an increased workload and a push for

law enforcement officers to develop new skills. Fake evidence has always existed and law enforcement agencies have procedures in place to assess the value of evidence. These procedures are developed for the types of forgeries already known and will have to be updated continuously with the rise of deepfakes. Law enforcement agencies will need to not only upskill their workforce to detect deepfakes, but also invest in their technical capabilities in order to address the upcoming challenges effectively while respecting fundamental rights.

Law enforcement agencies must consider this issue from multiple perspectives, when creating, storing, protecting and analysing audio-visual material. Specifically, they should:

- make use of tested and proven methods when making audio- visual recordings, e.g. certify a certain set-up for use in court and;
- employ technical and organisational safeguards against tampering, in order to be able to prove the authenticity of the footage.

Looking beyond law enforcement, general prevention strategies may be considered to make it harder to use deepfake technology on audio-visual material. For example, technical solutions could be implemented to make deepfakes easier to spot or to increase markers of authenticity. The Content Authenticity Initiative [32] is an example of efforts to provide a standard for content authenticity and provenance.

Participants of the Innovation Lab's foresight activities anticipated new forms of crime, together with the resulting challenges in terms of data collection, criminal attribution and the heightened anonymity of the perpetrators, such as in creating deepfakes for criminal purposes. Criminals are

[31] European Parliamentary Research Service, 'Tackling deepfakes in European policy', 2021, accessed 15 March 2022, https://www.europarl.europa.eu/RegData/etudes/ STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf
[32] Content Authenticy Initiative, accessed on 10 March 2022, https://contentauthenticity.org.

likely to adopt new modus operandi that LEAs will be unable to identify or counter. The failure to legislate for these technologies will further stymie the investigative abilities of LEAs.

Mitigating these risks requires greater research and funding. Law enforcement professionals will need to anticipate possible crime scenarios such as those discussed in this report, and build out their investigative abilities accordingly. Furthermore, they should work with relevant stakeholders to ensure that the appropriate legislation is in place. Greater awareness building and transparency vis-a-vis the public is also needed to ensure the roll-out of these technologies is not hamstrung by concerns over privacy and data protection.

Deepfake detection

Law enforcement has always had to deal with fake evidence and therefore is in a good position to adapt to the presence of deepfakes. In order to handle the material LEAs encounter appropriately, it is important to account for the possibility of synthetic content with malicious intent. Here we discuss some of the ways this synthetic content can be uncovered, and preventative measures that can be taken against this threat.

Manual detection

It is still possible for the vast majority of deepfake content to be manually detected by looking for inconsistencies. This is a labour intensive task, which can only be done for a very limited number of files, and requires appropriate training to be familiar with all the relevant signs. Moreover, this process is further complicated by the human predisposition to believe audio-visual content and work from a truth default perspective.[33] That introduces the possibility of mistakes, both with selecting the files that need to be inspected as well as the inspection itself.

The models generating deepfakes might produce believable images, but these may still contain imperfections upon closer examination.

A few examples include:

- blurring around the edges of the face;
- lack of blinking;
- light reflection in the eyes;
- inconsistencies in the hair, vein patterns, scars etc.;
- inconsistencies in the background, in subject as well as focus, depth etc.[34]

Automated detection

Ideally, a system would scan any digital content and automatically report on its authenticity. Such a system will most likely never be perfect, but with increased sophistication of deepfake technology, a high degree of certainty from such a system could be worth more than the manual inspection. There have already been efforts to create this kind of software from organisations such as Facebook[35] and

[33] Levine, T.R., 'Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection' Journal of Language and Social Psychology, 2014, pp. 378-392., https://www. researchgate.net/publication/273593306_Truth-Default_Theory_TDT_ A _Theory_of_ Human_Deception_and_Deception_Detection.

[34] Venema, A. E., & Geradts, Z. J., 'Digital Forensics Deepfakes and the Legal Process,' 2020,TheSciTechLawyer, 16(4), pp. 14-23.

[35] Michigan State University, MSU, 'Facebook develop research model to fight deepfakes', 2021,accessed on 10 March 2022, https://msutoday.msu.edu/news/2021/deepfake-detection.

security firm McAfee.[36] Detection software will look for signs of manipulation and help the reviewer decide on the authenticity with an explainable AI report on these signs.

As deepfake creation tools need training data to know what a real person looks like, most deepfake detection models are trained using databases of deepfake images. The learned signs of manipulation are thus based on data of known deepfakes, making it difficult to know how successful it will be at detecting deepfakes generated by unknown or updated models. Moreover, a deepfake GAN can be updated to account for the signs detected by known detection models in order to force the results to avoid producing these signs and henceforth go undetected.

Some examples[37] of detection technologies that have been developed in recent years are:

Biological signals

This approach tries to detect deepfakes based on imperfections in the natural changes in skin colour that arise from the flow of blood through the face.[38]

Phoneme-viseme mismatches

For some words the dynamics of the mouth, viseme, are inconsistent with the pronunciation of a phoneme. Deepfake models may not correctly combine viseme and phoneme in these cases.[39]

Facial movements

This approach uses correlations between facial movements and head movements to extract a characteristic movement of an individual to distinguish between real and manipulated or impersonated content.[40]

Recurrent Convolutional Models

Videos consist of frames which are really just a set of images. This approach looks for inconsistencies between these frames with deep learning models.

However, there are also challenges facing deepfake detection technology.

- Detection algorithms are trained on specific datasets. A slight alteration of the method used to generate the deepfake may therefore prevent detection.
- An update to the discriminative model of a GAN for specific artefacts detected by these systems will fool the detection software.
- Videos may be compressed or reduced in size, which causes problems with the reduction in pixels and artefacts, making it harder to detect the inconsistencies the system looks for.
- It has been shown that databases may be manipulated to misclassify images with certain identifiers by adding an identifier to a small part of the dataset (e.g. applying a trigger to 5% of the images resulted in the misclassification of fake images with the trigger as real).[41]

---

[36] McAfee, 'The Deepfakes Lab: Detecting & Defending Against Deepfakes with Advanced AI',2020, accessed on 10 March 2022, https://www.mcafee.com/blogs/enterprise/security- operations/the-deepfakes-lab-detecting-defending-against-deepfakes-with-advanced-ai.

[37] AIM, 'Top AI-Based Tools & Techniques For Deepfake Detection', 2020, accessed on 24 September 2021, https://analyticsindiamag.com/top-ai-based-tools-techniques-for- deepfake-detection.

[38] U. A. Ciftci, I. Demir and L. Yin, "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2020.3009287.

[39] Agarwal, S. et al., 'Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches', 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, accessed on 10 March 2022, https://www.ohadf.com/papers/AgarwalFaridFriedAgrawala_ CVPRW2020.pdf.

[40] Agarwal, S. et al., 'Protecting world leaders against deep fakes', Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 38-45, 2019, accessed on 10 March 2022, http://www.hao-li.com/publications/papers/cvpr2019workshopsPWLADF.pdf.

- Increased image forensics and deepfake detection capabilities drive the increased quality of deepfake videos. GANs can catch up relatively easily; by updating the discriminator to evade the detector, the learning capacity based on feedback loops of those GANs will work to produce a deepfake that can fool the detector.[42]

Preventive measures

Organisations that rely on some kind of authorisation by face or voice biometrics should assess the authorisation process as a whole. Increasing the robustness of this process is currently considered as a better course of action than solely implementing specific deepfake detection systems. Common checks are:

• using audio-visual authorisation rather than just audio;
• demanding live video connection;
• requiring random complicated acts to be performed live in front of the camera, e.g. move hands across the face.

How are other actors responding to deepfakes?

In order to address the challenges posed by deepfake technology, it is important to look into what kind of action other actors, including the online platforms where most deepfakes can and might be shared, are addressing this threat. This is also influenced by the current legislative framework, which can ask for mandatory or voluntary measures. In this section, this report will show some examples of key online service providers and companies and their anti-deepfake measures. This chapter will then examine the EU regulatory framework in this area.

Technology companies

Early in 2020, Meta (formerly Facebook) announced a new policy banning deepfakes from their platforms.[43] Meta said it would remove AI-edited content that would likely mislead people, but made it clear that satire or parodies using the same technology would still be permissible on the platforms. In order for law enforcement to assess and address the impact of deepfakes on its work, it needs to be aware of the policies technology companies have put in place, as it is likely that potential evidence or malicious content will be shared via these platforms. How technology companies such as Twitter and Meta regulate deepfake technology will have an extensive impact on how people will engage with and react to deepfakes.

Examples of company policies:

- Meta (which owns Facebook and Instagram) aims to remove deepfakes, or otherwise edited media, where "manipulation isn't apparent and could mislead, particularly in the case of video content."[44]

[41] Cao, X. and Gong, N.Z., 'Understanding the Security of Deepfake Detection' ArXiv, 2021, accessed on 18 October 2021, https://arxiv.org/abs/2107.02045.
[42] Wired, 'Deepfakes Aren't Very Good. Nor Are the Tools to Detect Them', 2020, accessed on 15 March 2022, https://www.wired.com/story/deepfakes-not-very-good-nor-tools-detect.
[43] Becoming Human: Artificial Intelligence Magazine, 'A Look at Deepfakes in 2020', 2020, accessed on 15 March 2022, https://becominghuman.ai/a-look-at-deepfakes-in-2020-13d3fe2b6ef7.
[44] Meta, 'Manipulated media', accessed on 10 March 2022, https://transparency.fb.com/en-gb/policies/community-standards/manipulated-media/.

- TikTok bans "Digital Forgeries (Synthetic Media or Manipulated Media) that mislead users by distorting the truth of events and cause significant harm to the subject of the video, other persons, or society."[45]
- Reddit "does not allow content that impersonates individuals or entities in a misleading or deceptive manner." This explicitly includes deepfakes "presented to mislead, or falsely attributed to an individual or entity." [46]
- Youtube has an existing ban for manipulated media under the spam, deceptive practices and scam policies of their community guidelines.[47]

Many of the policies use 'intent' as their barometer for deciding whether or not to remove a deepfake. However, defining 'intent' might prove challenging and highly subjective, since it is based on the assessment of individual actors. Nonetheless, it seems that online platforms could play a pivotal role in helping victims of deepfake technology to identify the perpetrator, but how this looks in practice remains to be seen. Moreover, technology providers also have responsibilities in safeguarding positive and legal use of their technologies and cooperating with law enforcement.

In addition to the policies, various technology companies are working on deepfake detection technologies. Developing detection technologies became a priority during the COVID-19 pandemic, and has gained new attention during the current conflict between Russia and Ukraine.

- Meta said it had developed an AI tool that detects deepfakes by reverse engineering a single AI-generated image to track its origin.[48]
- Google has released a large dataset of visual deepfakes that has been incorporated into the FaceForensics benchmark.[49]
- Microsoft has launched the Microsoft Video Authenticator, which can analyse a still photo or video to provide a percentage chance of whether the media has been artificially manipulated.[50]

European Union

Regarding legal trends, participants of the foresight activities noted that at both the national and regional level, European law is struggling to keep pace with the evolution of technology and the changing definitions of crime. Participants flagged the need to establish new regulatory frameworks. These should be sensitive to contemporary law enforcement challenges (particularly in the digital realm), as well as to changing ethical norms. Some participants anticipated greater regulation of the digital sphere in the coming decade.

[45] TikTok, 'Community Guidelines', accessed on 10 March 2022, https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok.

[46] Reddit, 'Updates to Our Policy Around Impersonation', 2020, accessed on 10 March 2022, https://www.reddit.com/r/redditsecurity/comments/emd7yx/updates_to_our_policy_ around_impersonation.

[47] Google Support, 'Misinformation policies', accessed on 10 March 2022, https://support.google.com/youtube/answer/10834785.

[48] Politico, 'POLITICO AI: Decoded: Big Tech on the AI Act — AI inventors — Deepfakes', 2021, accessed on 10 March 2022, https://www.politico.eu/newsletter/ai-decoded/politico-ai- decoded-big-tech-on-the-ai-act-ai-inventors-deepfakes.

[49] Google AI Blog, 'Contributing Data to Deepfake Detection Research', 2019, accessed on 10 March 2022, https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection. html.

[50] Microsoft, 'New Steps to Combat Disinformation', 2020, accessed on 10 March 2022, https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes- newsguard-video-authenticator.

The COVID-19 crisis brought more discussion around regulation of disinformation and deepfake detection tools, but also an increased use of video conferencing tools with adjustable backgrounds and other filters bringing manipulated digital realities into our daily lives. The European Parliament report, 'Tackling Deepfakes in European Policy', explains this and shows that the regulatory landscape in the European Union related to deepfakes "comprises a complex web of constitutional norms, as well as hard and soft regulations on both the EU and the Member State level".[51]

The most relevant regulatory framework for law enforcement in the area of deepfakes will be the AI regulatory framework – which is still at proposal level and not applicable yet - proposed by the

European Commission. The framework takes a risk-based approach to the regulation of AI and its applications. Deepfakes are explicitly covered by the passage about "AI systems used to generate or manipulate image, audio or video content", and have to adhere to certain minimum requirements. Minimum requirements include marking content as deepfake to make clear that users are dealing with manipulated footage."[52]

Deepfake detection software used by law enforcement authorities falls in the category of 'high-risk', as it is considered to pose a threat to the rights and freedoms of individuals. Detection software used by law enforcement under the AI regulatory framework would only be permitted under strict safeguards, such as the employment of risk-management systems and appropriate data governance and management practices.[53]

Conclusion

As this report shows, in order to effectively address the threats posed by deepfake technology, legislation and regulation need to take into account law enforcement needs. Within the regulatory framework, law enforcement, online service providers and

other organisations need to develop their policies and invest in detection as well as prevention technology. Policymakers and law enforcement agencies need to evaluate their current policies

and practices, and adapt them to be prepared for the new reality of deepfakes.

The strategic foresight activities conducted by the Europol Innovation Lab identified a series of challenges that LEAs will have to contend with in the decade ahead. In particular, they identified risks associated with digital transformation, the adoption and deployment of new technologies, the abuse of emerging technology

by criminals, accommodating new ways of working and maintaining trust in the face of an increase of disinformation.

In the months and years ahead, it is highly likely that threat actors will make increasing use of deepfake technology to facilitate various criminal acts and conduct disinformation campaigns to influence or distort public opinion. Advances in machine learning and artificial intelligence will continue enhancing the capabilities of the software used to create deepfakes. According to experts, GANs, availability

[51] European Parliament Research Service, 'Tackling deepfakes in European policy', 2021, accessed on 10 March 2022, https://www.europarl.europa.eu/RegData/etudes/ STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf.
[52] European Parliament Research Service, 'Tackling deepfakes in European policy', 2021, accessed on 10 March 2022, https://www.europarl.europa.eu/RegData/etudes/ STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf.
[53] European Parliament Research Service, 'Tackling deepfakes in European policy', 2021, accessed on 10 March 2022, https://www.europarl.europa.eu/RegData/etudes/ STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf.

of public datasets and increased computing power will be the main drivers of deepfake development in the future and make them more difficult to distinguish from authentic content.

The increase in use of deepfakes will require legislation to set guidelines and enforce compliance. Additionally, social networks and other online service providers should play a greater role in identifying and removing deepfake content from their platforms. As the public becomes more educated on deepfakes, there will be increasing concern worldwide about their impact on individuals, communities, and democracies.

In the EU there are various policies and regulatory attempts to address deepfakes. However, law enforcement's use of technology to detect deepfakes is considered as 'high-risk', according to some proposals. Therefore, it will be very important to clarify which practices should be prohibited under the AI regulatory framework. In order to address the challenges faced with deepfakes, law enforcement agencies need to prepare and train for deepfake detection and ensure e-evidence integrity, developing their capacities as described in this report. The regulatory framework should also support law enforcement preparedness efforts.

The Europol Innovation Lab is continuously monitoring the development of disruptive technologies such as deepfakes.

5. Deepfake e Stati Uniti

La potenziale minaccia dei deepfake è stata riconosciuta dal governo degli Stati Uniti. Il Malicious Deep Fake Prohibition Act del 2018 e l' Identification Outputs of Generative Adversarial Networks Act o IOGAN Act sono stati creati in risposta diretta alle minacce poste dai deepfake.

Malicious Deep Fake Prohibition Act of 2018

A BILL

To amend title 18, United States Code, to prohibit certain fraudulent audiovisual records, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the ``Malicious Deep Fake Prohibition Act of 2018''.

SEC. 2. FRAUD IN CONNECTION WITH AUDIOVISUAL RECORDS.

(a) Amendment.

Chapter 47 of title 18, United States Code, is amended by adding at the end the following:

``Sec. 1041. Fraud in connection with audiovisual records

``(a) Definitions.

In this section--

``(1) the term `audiovisual record'--

``(A) means any audio or visual media in an electronic format; and ``(B) includes any photograph, motion-picture film, video recording, electronic image, or sound recording;

``(2) the term `deep fake' means an audiovisual record created or altered in a manner that the record would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of an individual; and``(3) the terms `interactive computer service' and `information content provider' have the same meaning given the terms in section 230 of the Communications Act of 1934 (47 U.S.C. 230).

``(b) Offense.

It shall be unlawful to, using any means or facility of interstate or foreign commerce--

``(1) create, with the intent to distribute, a deep fake with the intent that the distribution of the deep fake would facilitate criminal or tortious conduct under Federal, State, local, or Tribal law; or

``(2) distribute an audiovisual record with--``(A) actual knowledge that the audiovisual record is a deep fake; and``(B) the intent that the distribution of the audiovisual record would facilitate criminal or tortious conduct under Federal, State, local, or Tribal law.

``(c) Penalty.

Any person who violates subsection (b) shall be--``(1) fined under this title, imprisoned for not more than 2 years, or both; or``(2) fined under this title, imprisoned for not more than 10 years, or both, in the case of a violation in which the creation, reproduction, or distribution of the deep fake could be reasonably expected to--``(A) affect the conduct of any administrative, legislative, or judicial

proceeding of a Federal, State, local, or Tribal government agency, including the administration of an election or the conduct of foreign relations; or ``(B) facilitate violence.

``(d) Limitations.

``(1) In general.--For purposes of this section, a provider of an interactive computer service shall not be held liable on account of--

``(A) any action voluntarily taken in good faith to restrict access to or availability of deep fakes; or

``(B) any action taken to enable or make available to information content providers or other persons the technical means to restrict access to deep fakes.

``(2) First amendment protection.--No person shall be held liable under this section for any activity protected by the First Amendment to the Constitution of the United States.".

(b) Technical and Conforming Amendment.--The table of sections for chapter 47 is amended by inserting after the item relating to section 1040 the following:

``1041. Fraud in connection with audiovisual records.".

Identification Outputs of Generative Adversarial Networks Act

Identifying outputs of generative adversarial networks act

PUBLIC LAW 116–258—DEC. 23, 2020

Public Law 116–258 116th Congress

An Act

To direct the Director of the National Science Foundation to support research on the outputs that may be generated by generative adversarial networks, other- wise known as deepfakes, and other comparable techniques that may be developed in the future, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the "Identifying Outputs of Generative Adversarial Networks Act" or the "IOGAN Act".

SEC. 2. FINDINGS.

Congress finds the following:

(1) Gaps currently exist on the underlying research needed to develop tools that detect videos, audio files, or photos that have manipulated or synthesized content, including those generated by generative adversarial networks. Research on digital forensics is also needed to identify, preserve, recover, and analyze the provenance of digital artifacts.

(2) The National Science Foundation's focus to support research in artificial intelligence through computer and information science and engineering, cognitive science and psychology, economics and game theory, control theory, linguistics, mathematics, and philosophy, is building a better understanding of how new technologies are shaping the society and economy of the United States.

(3) The National Science Foundation has identified the "10 Big Ideas for NSF Future Investment" including "Harnessing the Data Revolution" and the "Future of Work at the Human-Technology Frontier", with artificial intelligence is a critical component.

(4) The outputs generated by generative adversarial networks should be included under the umbrella of research described in paragraph (3) given the grave national security and societal impact potential of such networks.

(5) Generative adversarial networks are not likely to be utilized as the sole technique of artificial intelligence or machine learning capable of creating credible deepfakes. Other techniques may be developed in the future to produce similar outputs.

SEC. 3. NSF SUPPORT OF RESEARCH ON MANIPULATED OR SYN- THESIZED CONTENT AND INFORMATION SECURITY.

The Director of the National Science Foundation, in consulta- tion with other relevant Federal agencies, shall support merit- reviewed and competitively awarded research on manipulated or synthesized content and information authenticity, which may include—

(1) fundamental research on digital forensic tools or other of manipulated or synthesized content, including con- tent generated by generative adversarial networks;

(2) fundamental research on technical tools for identifying manipulated or synthesized content, such as watermarking sys- tems for generated media;

(3) social and behavioral research related to manipulated or synthesized content, including human engagement with the content;

(4) research on public understanding and awareness of manipulated and synthesized content, including research on best practices for educating the public to discern authenticity of digital content; and

(5) research awards coordinated with other federal agencies and programs, including the Defense Advanced Research Projects Agency and the Intelligence Advanced Research Projects Agency, with coordination enabled by the Networking and Information Technology Research and Development Pro- gram.

SEC. 4. NIST SUPPORT FOR RESEARCH AND STANDARDS ON GENERA- TIVE ADVERSARIAL NETWORKS.

(a) IN GENERAL.—The Director of the National Institute of Standards and Technology shall support research for the develop- ment of measurements and standards necessary to accelerate the development of the technological tools to examine the function and outputs of generative adversarial networks or other technologies that synthesize or manipulate content.

(b) OUTREACH.—The Director of the National Institute of Stand- ards and Technology shall conduct outreach—

(1) to receive input from private, public, and academic stakeholders on fundamental measurements and standards research necessary to examine the function and outputs of generative adversarial networks; and

(2) to consider the feasibility of an ongoing public and private sector engagement to develop voluntary standards for the function and outputs of generative adversarial networks or other technologies that synthesize or manipulate content.

SEC. 5. REPORT ON FEASIBILITY OF PUBLIC-PRIVATE PARTNERSHIP TO DETECT MANIPULATED OR SYNTHESIZED CONTENT.

Not later than 1 year after the date of enactment of this Act, the Director of the National Science Foundation and the Director of the National Institute of Standards and Technology shall jointly submit to the Committee on Science, Space, and Tech- nology of the House of Representatives, the Subcommittee on Com- merce, Justice, Science, and Related Agencies of the Committee on

Appropriations of the House of Representatives, the Committee on Commerce, Science, and Transportation of the Senate, and the Subcommittee on Commerce, Justice, Science, and Related Agencies of the Committee on Appropriations of the Senate a report con- taining—

(1) the Directors' findings with respect to the feasibility for research opportunities with the private sector, including digital media companies to detect the function and outputs of generative adversarial networks or other technologies that synthesize or manipulate content; and

(2) any policy recommendations of the Directors that could facilitate and improve communication and coordination between the private sector, the National Science Foundation, and rel- evant Federal agencies through the implementation of innova- tive approaches to detect digital content produced by generative adversarial networks or other technologies that synthesize or manipulate content.

SEC. 6. GENERATIVE ADVERSARIAL NETWORK DEFINED.

In this Act, the term "generative adversarial network" means, with respect to artificial intelligence, the machine learning process of attempting to cause a generator artificial neural network (referred to in this paragraph as the "generator" and a discriminator artificial neural network (referred to in this paragraph as a "discriminator") to compete against each other to become more accurate in their function and outputs, through which the generator and discriminator create a feedback loop, causing the generator to produce increasingly higher-quality artificial outputs and the discriminator to increasingly improve in detecting such artificial outputs.

6. Come proteggersi dai deepfake

Tratto dal sito del garante: **www.gpdp.it**

Le grandi imprese del digitale (piattaforme social media, motori di ricerca, ecc.) stanno già studiando e applicando delle metodologie per il contrasto al fenomeno, come algoritmi di intelligenza artificiale capaci di individuare i deepfake o sistemi per le segnalazioni da parte degli utenti, e stanno formando team specializzati nel monitoraggio e contrasto al deepfake. E le Autorità di protezione dei dati personali possono intervenire per prevenire e sanzionare le violazioni della normativa in materia di protezione dati.

Tuttavia, il primo e più efficace strumento di difesa è rappresentato sempre dalla responsabilità e dall'attenzione degli utenti. Ecco allora alcuni suggerimenti:

- Evitare di diffondere in modo incontrollato immagini personali o dei propri cari. In particolare, se si postano immagini sui social media, è bene ricordare che le stesse potrebbero rimanere online per sempre o che, anche nel caso in cui si decida poi di cancellarle, qualcuno potrebbe già essersene appropriato.
- Anche se non è semplice, si può imparare a riconoscere un deepfake. Ci sono elementi che aiutano: l'immagine può appare pixellata (cioè un pò "sgranata" o sfocata); gli occhi delle persone possono muoversi a volte in modo innaturale; la bocca può apparire deformata o troppo grande mentre la persona dice alcune cose; la luce e le ombre sul viso possono apparire anormali.
- Se si ha il dubbio che un video o un audio siano un deepfake realizzato all'insaputa dell'interessato, occorre assolutamente evitare di condividerlo (per non moltiplicare il danno alle persone con la sua diffusione incontrollata). E si può magari decidere di segnalarlo come possibile falso alla piattaforma che lo ospita (ad esempio, un social media).
- Se si ritiene che il deepfake sia stato utilizzato in modo da compiere un reato o una violazione della privacy, ci si può rivolgere, a seconda dei casi, alle autorità di polizia (ad esempio, alla Polizia postale) o al Garante per la protezione dei dati personali.

Come proteggersi in azienda

- Le aziende devono aggiungere discussioni sui deepfake alla loro formazione sulla consapevolezza della sicurezza informatica. La formazione sulla cyber-consapevolezza dovrebbe far parte dell'induzione di un nuovo avviamento e dovrebbe essere ripetuta periodicamente per tutto il personale.
- Finora, gli attacchi osservati sono versioni raffinate di attacchi di phishing e spear-phishing. Semplici procedure possono aiutare a riconoscere e bloccare molti di questi.
- Nessun trasferimento di finanze dovrebbe essere effettuato esclusivamente al ricevimento di un'e-mail.
- Una telefonata di follow-up dovrebbe essere effettuata dal destinatario dell'e-mail al mittente, non dal mittente al destinatario.
- È possibile incorporare frasi di controllo concordate o parole-chiave che un utente malintenzionato esterno non conoscerebbe.

- Confrontare e ricontrollare tutto ciò che è fuori dall'ordinario.

Bibliografia

"Synthetic Media and Potential Safeguards," Carnegie Endowment for International Peace, https://carnegieendowment.org/siliconvalley/synthetic-media.

James Vincent, "Why We Need a Better Definition of 'Deepfake'," The Verge, May 22, 2018, https://www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news.

Timothy B. Lee, "I Created My Own Deepfake—It Took Two Weeks and Cost $552," Ars Technica, December 16, 2019, https://arstechnica.com/science/2019/12/how-i-created-a-deepfake-of-mark-zuckerberg-and-star-treks-data/.

Samantha Cole, "Deepfakes Were Created As a Way to Own Women's Bodies—We Can't Forget That," Vice, June 18, 2018, https://www.vice.com/en_us/article/nekqmd/deepfake-porn-origins-sexism-reddit-v25n2.

Martin Giles, "The GANfather: The Man Who's Given Machines the Gift of Imagination," MIT Technology Review, February 21, 2018, https://www.technologyreview.com/2018/02/21/145289/the-ganfather-the-man-whos-given-machines-the-gift-of-imagination/.

Charlotte Stanton, "November 2018 Convening Mapping Synthetic Media's Problem and Solution Space," Carnegie Endowment for International Peace, November 16, 2018, https://carnegieendowment.org/2018/11/16/november-2018-convening-mapping-synthetic-media-s-problem-and-solution-space-pub-79892.

Lee, "I Created My Own Deepfake—It Took Two Weeks and Cost $552."

"Combating Spoofed Robocalls With Caller ID Authentication," Federal Communications Commission, https://www.fcc.gov/call-authentication.

Federal Trade Commission, "Consumer Sentinel Network Data Book 2019," January 2020, https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-2019/consumer_sentinel_network_data_book_2019.pdf.

Identity Theft Resource Center, "2018 Annual Report," 2019, https://www.idtheftcenter.org/wp-content/uploads/2019/02/ITRC_ANNUAL-IMPACT-REPORT-2018_web.pdf.

"Thirty-six Defendants Indicted for Alleged Roles in Transnational Criminal Organization Responsible for More Than $530 Million in Losses From Cybercrimes," press release, Department of Justice, February 7, 2018, https://www.justice.gov/opa/pr/thirty-six-defendants-indicted-alleged-roles-transnational-criminal-organization-responsible.

McAfee, "The Hidden Data Economy," 2017, https://www.mcafee.com/enterprise/en-us/assets/reports/rp-hidden-data-economy.pdf.

Catherine Stupp, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case," Wall Street Journal, updated August 30, 2019, https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.

"Deepfakes: The Threat to Financial Services," iProov, January 29, 2020, https://www.iproov.com/newsroom/blog/deepfakes-the-threat-to-financial-services.

Samantha Cole, "A Site Faking Jordan Peterson's Voice Shuts Down After Peterson Decries Deepfakes," Motherboard, August 26, 2019, https://www.vice.com/en_us/article/43kwgb/not-jordan-peterson-voice-generator-shut-down-deepfakes.

"How Lyrebird Uses AI to Find Its (Artificial) Voice)," Wired, October 15, 2018, https://www.wired.com/brandlab/2018/10/lyrebird-uses-ai-find-artificial-voice/; and Natasha Lomas, "Lyrebird Is a Voice Mimic for the Fake News Era," TechCrunch, April 25, 2017, https://techcrunch.com/2017/04/25/lyrebird-is-a-voice-mimic-for-the-fake-news-era/.

Samantha Cole, "'Deep Voice' Software Can Clone Anyone's Voice With Just 3.7 Seconds of Audio," Motherboard, https://www.vice.com/en_us/article/3k7mgn/baidu-deep-voice-software-can-clone-anyones-voice-with-just-37-seconds-of-audio; and Reina Qi Wan, "Clone a Voice in Five Seconds With This AI Toolbox," Synced, September 9, 2019, https://syncedreview.com/2019/09/03/clone-a-voice-in-five-seconds-with-this-ai-toolbox/.

Ellie Rushing, "A Philly Lawyer Nearly Wired $9,000 to a Stranger Impersonating His Son's Voice, Showing Just How Smart Scammers Are Getting," Philadelphia Inquirer, updated March 9, 2020, https://www.inquirer.com/news/voice-scam-impersonation-fraud-bail-bond-artificial-intelligence-20200309.html.

Julie Hirschfeld Davis, "Prankster Calls the President, and the White House Puts Him Right Through," New York Times, June 29, 2018, https://www.nytimes.com/2018/06/29/us/politics/prank-call-donald-trump-stuttering-john.html; and Zachary Evans, "Graham Tricked By Russian Prank Call, Called Kurds a 'Problem' for Turkey," National Review, October 10, 2019, https://www.nationalreview.com/news/lindsey-graham-tricked-by-russian-prank-call-called-kurds-a-problem-for-turkey/.

Sean Gallagher, "The New Spam: Interactive Robo-calls From the Cloud as Cheap as E-mail," Ars Technica, April 15, 2015, https://arstechnica.com/information-technology/2015/04/the-new-spam-interactive-robo-calls-from-the-cloud-as-cheap-as-e-mail/.

Federal Trade Commission, "Consumer Sentinel Network Data Book 2019."

Federal Trade Commission, "Consumer Sentinel Network Data Book 2019; and "Watch Out for 'Grandparent' Scams," Federal Communications Commission, May 22, 2018, https://www.fcc.gov/watch-out-grandparent-scams.

"Imposter Scams," Office of the Minnesota Attorney General, https://www.ag.state.mn.us/Consumer/Publications/ImposterScams.asp.

Federal Trade Commission, "Consumer Sentinel Network Data Book 2019."

"We Are VocaliD, Your Voice AI Company," VocaliD, https://vocalid.ai/about-us/ .

Rushing, "A Philly Lawyer Nearly Wired $9,000 to a Stranger Impersonating His Son's Voice, Showing Just How Smart Scammers Are Getting."

Sarah Krouse, "Robocall Scams Exist Because They Work—One Woman's Story Shows How," Wall Street Journal, November 21, 2019, https://www.wsj.com/articles/robocall-scams-exist-because-they-workone-womans-story-shows-how-11574351204; Federal Trade Commission, "Consumer Sentinel Network Data Book 2019"; and Federal Bureau of Investigation Internet Crime Complaint Center, "2019 Internet Crime Report," https://pdf.ic3.gov/2019_IC3Report.pdf.

"What Is Sextortion?" Federal Bureau of Investigation, https://www.fbi.gov/video-repository/newss-what-is-sextortion/view.

"The Revival and Rise of Email Extortion Scams," Symantec, July 30, 2019, https://symantec-blogs.broadcom.com/blogs/threat-intelligence/email-extortion-scams.

Kate Fazzini, "Email Sextortion Scams Are on the Rise and They're Scary—Here's What to Do If You Get One," CNBC, June 17, 2019, https://www.cnbc.com/2019/06/17/email-sextortion-scams-on-the-rise-says-fbi.html.

"The Revival and Rise of Email Extortion Scams."

Federal Bureau of Investigation Internet Crime Complaint Center, "2019 Internet Crime Report."

Cole, "Deepfakes Were Created As a Way to Own Women's Bodies."

Joseph Cox, "Most Deepfakes Are Used for Creating Non-Consensual Porn, Not Fake News," Vice, October 7, 2019, https://www.vice.com/en_us/article/7x57v9/most-deepfakes-are-porn-harassment-not-fake-news.

James Vincent, "New AI Deepfake App Creates Nude Images of Women in Seconds," The Verge, June 27, 2019, https://www.theverge.com/2019/6/27/18760896/deepfake-nude-ai-app-women-deepnude-non-consensual-pornography.

Jon Porter, "Another Convincing Deepfake App Goes Viral Prompting Immediate Privacy Backlash," The Verge, September 2, 2019, https://www.theverge.com/2019/9/2/20844338/zao-deepfake-app-movie-tv-show-face-replace-privacy-policy-concerns.

Fazzini, "Email Sextortion Scams Are on the Rise and They're Scary."

Gautam S. Mengle, "Law Enforcers Worried as Deep Nude Makes a Return," Hindu, April 13, 2020, https://www.thehindu.com/news/national/law-enforcers-worried-as-deep-nude-makes-a-return/article31334415.ece.

Fazzini, "Email Sextortion Scams Are on the Rise and They're Scary."

43 Federal Bureau of Investigation Internet Crime Complaint Center, "Business E-mail Compromise," public service announcement, January 22, 2015, https://www.ic3.gov/media/2015/150122.aspx; and Federal Bureau of Investigation Internet Crime Complaint Center, "Business Email Compromise: Gift Cards," public service announcement, October 24, 2018, https://www.ic3.gov/media/2018/181024.aspx.

Federal Bureau of Investigation Internet Crime Complaint Center, "2019 Internet Crime Report."

Federal Bureau of Investigation Internet Crime Complaint Center, "Business E-Mail Compromise"; and "Recognizing and Avoiding Business Email Compromise Attacks," Proofpoint, 2019, https://www.proofpoint.com/sites/default/files/pfpt-us-ig-bec.pdf.

Federal Bureau of Investigation Internet Crime Complaint Center, "2019 Internet Crime Report."

Stupp, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case."

Alessandro Cauduro, "Live Deep Fakes—You Can Now Change Your Face to Someone Else's in Real Time Video Applications," Medium, April 4, 2018, https://medium.com/huia/live-deep-fakes-you-can-now-change-your-face-to-someone-elses-in-real-time-video-applications-a4727e06612f.

Lawrence Delevingne, "Short & Distort? The Ugly War Between CEOs and Activist Critics," Reuters, March 21, 2019, https://www.reuters.com/article/us-usa-stocks-shorts-insight/short-distort-the-ugly-war-between-ceos-and-activist-critics-idUSKCN1R20AW; and "SEC Fires Warning Shot Against 'Short and Distort' Schemes," DLA Piper, October 18, 2018, https://www.dlapiper.com/en/us/insights/publications/2018/10/sec-fires-warning-shot-against/.

Joshua Mitts, "Short and Distort," Columbia Law and Economics Working Paper No. 592, February 20, 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3198384; and Thomas Renault, "Pump-and-dump or News? Stock Market Manipulation on Social Media," European Financial Management Association, 2017,

https://efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2017-Athens/papers/EFMA2017_0387_fullpaper.pdf.
Will Kenton, "Poop and Scoop," Investopedia, May 11, 2018, https://www.investopedia.com/terms/p/poopandscoop.asp.
Claire Atkinson, "Fake News Can Cause 'Irreversible Damage' to Companies—and Sink Their Stock Price," NBC News, April 25, 2019, https://www.nbcnews.com/business/business-news/fake-news-can-cause-irreversible-damage-companies-sink-their-stock-n995436.
Leroy Terrelonge and Jim Hempstead, "Deepfakes Can Threaten Companies' Financial Health," Moody's Investors Service, August 1, 2019, https://www.moodys.com/research/Moodys-Deepfakes-can-threaten-companies-financial-health--PBC_1188117?showPdf=true.
Darla Mercado, "Fisher Withdrawals Top $3 Billion as Texas Retirement Plan Exits," CNBC, October 25, 2019, https://www.cnbc.com/2019/10/25/fisher-withdrawals-top-3-billion-as-texas-retirement-plan-exits.html.
Ramona Shelburne, "When the Donald Sterling Saga Rocked the NBA—and Changed It Forever," August 20, 2019, https://www.espn.com/nba/story/_/id/27414482/when-donald-sterling-saga-rocked-nba-changed-forever.
Drew Harwell, "Doctored Images Have Become a Fact of Life for Political Campaigns. When They're Disproved, Believers 'Just Don't Care.'" Washington Post, January 14, 2020, https://www.washingtonpost.com/technology/2020/01/14/doctored-political-images.
59 Megan Metzger, "Effectiveness of Responses to Synthetic and Manipulated Media on Social Media Platforms," Carnegie Endowment for International Peace, November 15, 2019, https://carnegieendowment.org/2019/11/15/legal-ethical-and-efficacy-dimensions-of-managing-synthetic-and-manipulated-media-pub-80439#effectiveness.
Stefano Cresci, Fabrizio Lillo, Daniele Regoli, Serena Tardelli, Maurizio Tesconi, "Cashtag Piggybacking: Uncovering Spam and Bot Activity in Stock Microblogs on Twitter," ACM Transactions on the Web 13 (2018), https://arxiv.org/pdf/1804.04406.pdf.
Rui Fan, Oleksandr Talavera, and Vu Tran, "Social Media Bots and Stock Markets," Swansea University School of Management, Working Paper No. 30, 2018, https://rahwebdav.swan.ac.uk/repec/pdf/WP2018-30.pdf.
"Platform Manipulation and Spam Policy," Twitter, September 2019, https://help.twitter.com/en/rules-and-policies/platform-manipulation; and "How to Spot a Twitter Bot," Symantec, October 26, 2018, https://symantec-blogs.broadcom.com/blogs/election-security/spot-twitter-bot.
Jurgen Knauth, "Language-Agnostic Twitter Bot Detection," Proceedings of Recent Advances in Natural Language Processing, September 2–4, 2019, https://www.aclweb.org/anthology/R19-1065.pdf; John Gramlich, "Q&A: How Pew Research Center Identified Bots on Twitter," Pew Research Center, April 19, 2018, https://www.pewresearch.org/fact-tank/2018/04/19/qa-how-pew-research-center-identified-bots-on-twitter/; and Alyssa Newcomb, "Twitter Is Purging Millions of Fake Accounts—and Investors Are Spooked," NBC News, July 9, 2018, https://www.nbcnews.com/tech/tech-news/twitter-purging-millions-fake-accounts-investors-are-spooked-n889941.
Yoel Roth and Nick Pickles, "Bot or Not? The Facts About Platform Manipulation on Twitter," Twitter Blog, May 18, 2020, https://blog.twitter.com/en_us/topics/company/2020/bot-or-not.html.

Gramlich, "Q&A: How Pew Research Center Identified Bots on Twitter"; and Michael Kreil, "The Army That Never Existed: The Failure of Social Bots Research," Github, November 2, 2019, https://michaelkreil.github.io/openbots.

Newcomb, "Twitter Is Purging Millions of Fake Accounts"; and Luis Sanchez, "Conservatives Say They've Lost Thousands of Followers on Twitter," The Hill, February 21, 2018, https://thehill.com/policy/technology/374842-conservatives-say-theyve-lost-thousands-of-followers-on-twitter.

Paris Martineau, "Facebook Removes Accounts With AI-generated Profile Photos," Ars Technica, December 23, 2019, https://arstechnica.com/tech-policy/2019/12/facebook-removes-accounts-with-ai-generated-profile-photos.

Davey Alba, "Facebook Discovers Fakes That Show Evolution of Disinformation," New York Times, December 20, 2019, https://www.nytimes.com/2019/12/20/business/facebook-ai-generated-profiles.html; and Raphael Satter, "Experts: Spy Used AI-generated Face to Connect With Targets," Associated Press, June 13, 2019, https://apnews.com/bc2f19097a4c4fffaa00de6770b8a60d.

James Vincent, "OpenAI Has Published the Text-generating AI It Said Was Too Dangerous to Share," The Verge, November 7, 2019, https://www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters.

Elana Lyn Gross, "Why Peloton Stock Dropped More Than 10% After 'Sexist' Ad Backlash," Forbes, December 5, 2019, https://www.forbes.com/sites/elanagross/2019/12/05/peloton-stock-is-down-more-than-10-following-backlash-about-sexist-ad/.

This Person Does Not Exist, https://www.thispersondoesnotexist.com/. License: https://nvlabs.github.io/stylegan2/license.html.

"GauGAN Beta," NVIDIA, http://nvidia-research-mingyuliu.com/gaugan/. License: http://nvidia-research-mingyuliu.com/gaugan/term.txt.

The first sentence of the mock Tweet was a human-written prompt, which enabled the algorithm to write the remainder. The algorithm was run several times, and the most convincing output was selected for this mock-up. The mock Twitter bio was completely AI-generated on the first try, using the prompt "About me:". Although this AI algorithm requires a human prompt, prompts could be algorithmically generated in the future.

"20th Century English Name Generator," Fantasy Name Generators, https://www.fantasynamegenerators.com/20th-century-english-names.php.

77 Satter, "Experts: Spy Used AI-generated Face to Connect With Targets"; and Ravie Lakshmanan, "This AI Tool Is Smart Enough to Spot AI-generated Articles and Tweets," The Next Web, July 29, 2019, https://thenextweb.com/artificial-intelligence/2019/07/29/this-ai-tool-is-smart-enough-to-spot-ai-generated-articles-and-tweets/.

78 Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi, "The Paradigm-shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race," Proceedings of the 26th International Conference on World Wide Web Companion (2017), https://arxiv.org/pdf/1701.03017.pdf.

Cresci, et al., "The Paradigm-shift of Social Spambots."

"Applying Social Sentiment to Your Stock Research," Fidelity, 2017, https://www.fidelity.com/learning-center/tools-demos/research-tools/social-sentiment-research-video.

"The Day Social Media Schooled Wall Street," Atlantic Re:think, https://www.theatlantic.com/sponsored/etrade-social-stocks/the-day-social-media-schooled-wall-street/327.

"Investor Bulletin: Social Sentiment Investing Tools—Think Twice Before Trading Based on Social Media," Securities Exchange Commission, April 3, 2019, https://www.sec.gov/oiea/investor-alerts-and-bulletins/ib_sentimentinvesting.

Ben Chapman, "Metro Bank Says Customers' Money Safe After WhatsApp Rumour Sparks Panic," Independent, May 13, 2019, https://www.independent.co.uk/news/business/news/metro-bank-whatsapp-money-account-safe-deposit-box-a8911296.html.

Issaku Harada, "Online Rumors Spark Runs on Smaller Chinese Banks," Nikkei Asian Review, December 19, 2019, https://asia.nikkei.com/Economy/Online-rumors-spark-runs-on-smaller-Chinese-banks.

Matthias Williams and Tsvetelia Tsolova, "Accusations Fly in Bulgaria's Murky Bank Run," Reuters, July 4, 2014, https://www.reuters.com/article/us-bulgaria-banking-insight/accusations-fly-in-bulgarias-murky-bank-run-idUSKBN0F90SG20140704.

Nadine Ajaka, Elyse Samuels, and Glenn Kessler, "Seeing Isn't Believing: The Fact Checker's Guide to Manipulated Video," Washington Post, 2019, https://www.washingtonpost.com/graphics/2019/politics/fact-checker/manipulated-video-guide/.

Shawn Langlois, "This Day in History: Hacked AP Tweet About White House Explosions Triggers Panic," MarketWatch, April 23, 2018, https://www.marketwatch.com/story/this-day-in-history-hacked-ap-tweet-about-white-house-explosions-triggers-panic-2018-04-23.

Metzger, "Effectiveness of Responses to Synthetic and Manipulated Media on Social Media Platforms."

"How High-frequency Trading Hit a Speed Bump," Financial Times, January 1, 2018, https://www.ft.com/content/d81f96ea-d43c-11e7-a303-9060cb1e5f44.

Jean-Philippe Serbera, "Flash Crashes: If Reforms Aren't Ramped Up, the Next One Could Spell Global Disaster," The Conversation, January 7, 2019, https://theconversation.com/flash-crashes-if-reforms-arent-ramped-up-the-next-one-could-spell-global-disaster-109362.

Ding Yi, "Chinese Central Bank Denies Digital Currency Issuance Rumors," CX Tech, November 14, 2019, https://www.caixinglobal.com/2019-11-14/chinese-central-bank-denies-digital-currency-issuance-rumors-101483430.html.

Suvashree Ghosh, "India Central Bank Denies Rumors of Bank Closures," Bloomberg, September 25, 2019, https://www.bloomberg.com/news/articles/2019-09-25/frayed-nerves-force-india-watchdog-to-douse-bank-closure-rumors; and "Myanmar Central Bank Refutes Rumors About Bank Closure," Xinhua, August 19, 2015, http://www.globaltimes.cn/content/937860.shtml.

Venkatesan Vembu, "Buzz of $430 Billion Loss; Top Banker Defection Freaks Out China," DNA India, September 1, 2010, https://www.dnaindia.com/business/report-buzz-of-430-billion-loss-top-banker-defection-freaks-out-china-1431780.

Beth Piskora, "Fed Head Is Not Dead—Rumor False; Markets Drop on Earnings, Rate Hike Fears," June 23, 2000, https://nypost.com/2000/06/23/fed-head-is-not-dead-rumor-false-markets-drop-on-earnings-rate-hike-fears.

Sarah Foster, "The Fed Wants to Be Easier to Understand—and It May Be Risky to the Markets," Bankrate, June 4, 2019, https://www.bankrate.com/banking/federal-reserve/fed-simple-communication-may-be-confusing-markets; and Alister Bull, "Fed Slammed for Poor Communication

by Its Own Advisory Council," Reuters, October 4, 2013, https://www.reuters.com/article/us-usa-fed-communication-idUSBRE99313220131004.
BBC, "Call for Bank of England Executive to Quit Over Security Breach," December 19, 2019, https://www.bbc.com/news/business-50849479.
U.S. Senate Permanent Subcommittee on Investigations, "Abuses of the Federal Notice-and-Comment Rulemaking Process," October 24, 2019, https://www.hsgac.senate.gov/imo/media/doc/2019-10-24%20PSI%20Staff%20Report%20-%20Abuses%20of%20the%20Federal%20Notice-and-Comment%20Rulemaking%20Process.pdf?mod=article_inline.
Jeremy Singer-Vine and Kevin Collier, "Political Operatives Are Faking Voter Outrage With Millions Of Made-Up Comments To Benefit The Rich And Powerful," Buzzfeed News, October 3, 2019, https://www.buzzfeednews.com/article/jsvine/net-neutrality-fcc-fake-comments-impersonation.
James V. Grimaldi and Paul Overberg, "Millions of People Post Comments on Federal Regulations. Many Are Fake," Wall Street Journal, December 12, 2017, https://www.wsj.com/articles/millions-of-people-post-comments-on-federal-regulations-many-are-fake-1513099188.
Vincent, "OpenAI Has Published the Text-generating AI It Said Was Too Dangerous to Share."
King, "Talk to Transformer," https://talktotransformer.com/. This output was produced on the first try. The only human intervention was to separate paragraphs.
Tom McKay, "Turns Out Elon Musk–Backed OpenAI's Text Generator Is More Funny Than Dangerous, For Now," Gizmodo, November 7, 2019, https://gizmodo.com/turns-out-elon-musk-backed-openais-text-generator-is-mo-1839705114.
Max Weiss, "Deepfake Bot Submissions to Federal Public Comment Websites Cannot Be Distinguished From Human Submissions," Technology Science, December 18, 2019, https://techscience.org/a/2019121801/.
Clea Simon, "How I Hacked the Government (It Was Easier Than You May Think)," Harvard Gazette, February 6, 2020, https://news.harvard.edu/gazette/story/2020/02/why-an-undergrad-flooded-government-websites-with-bot-comments/; and Weiss, "Deepfake Bot Submissions to Federal Public Comment Websites Cannot Be Distinguished from Human Submissions."
Lakshmanan, "This AI Tool Is Smart Enough to Spot AI-generated Articles and Tweets."
Karen Hao, "An AI for Generating Fake News Could Also Help Detect It," MIT Technology Review, March 12, 2019, https://www.technologyreview.com/2019/03/12/136668/an-ai-for-generating-fake-news-could-also-help-detect-it/.
U.S. Senate Permanent Subcommittee on Investigations, "Abuses of the Federal Notice-and-Comment Rulemaking Process," October 24, 2019, https://www.hsgac.senate.gov/imo/media/doc/2019-10-24%20PSI%20Staff%20Report%20-%20Abuses%20of%20the%20Federal%20Notice-and-Comment%20Rulemaking%20Process.pdf?mod=article_inline.
Weiss, "Deepfake Bot Submissions to Federal Public Comment Websites Cannot Be Distinguished from Human Submissions."
Beth Simone Noveck, "Astroturfing Is Bad But It's Not the Whole Problem," Nextgov, February 6, 2020, https://www.nextgov.com/ideas/2020/02/astroturfing-bad-its-not-whole-problem/162932/.
James V. Grimaldi, "Federal Agencies Found to Be Lax in Halting Fake Comments on Proposed Rules," Wall Street Journal, October 24, 2019, https://www.wsj.com/articles/federal-agencies-found-to-be-lax-in-halting-fake-comments-on-proposed-rules-11571909402.
Ajaka, Samuels, and Kessler, "Seeing Isn't Believing: The Fact Checker's Guide to Manipulated Video."

"CBS News Admits Bush Documents Can't Be Verified," Associated Press, September 21, 2004, http://www.nbcnews.com/id/6055248/ns/politics/t/cbs-news-admits-bush-documents-cant-be-verified/#.Xp45-FNKh0s; Hugh Schofield, "The Fake French Minister in a Silicone Mask Who Stole Millions," BBC, June 20, 2019, https://www.bbc.com/news/world-europe-48510027; and Ashley Feinberg, "The Pee Tape Is Real, but It's Fake," Slate, September 15, 2019, https://slate.com/comments/news-and-politics/2019/09/inside-the-convincing-fake-trump-pee-tape.html.

Emma Fletcher, "New Twist to Grandparent Scam: Mail Cash," Federal Trade Commission, December 3, 2018, https://www.ftc.gov/news-events/blogs/data-spotlight/2018/12/new-twist-grandparent-scam-mail-cash.

Thomas E. Kadri, "The Legal Implications of Synthetic and Manipulated Media," Carnegie Endowment for International Peace, November 15, 2019, https://carnegieendowment.org/2019/11/15/legal-ethical-and-efficacy-dimensions-of-managing-synthetic-and-manipulated-media-pub-80439#legal.

David Danks and Jack Parker, "The Un/Ethical Status of Synthetic Media," Carnegie Endowment for International Peace, November 15, 2019, https://carnegieendowment.org/2019/11/15/legal-ethical-and-efficacy-dimensions-of-managing-synthetic-and-manipulated-media-pub-80439#ethics; David Danks and Jack Parker, "Ethical Analysis of Responses to Synthetic and Manipulated Media," Carnegie Endowment for International Peace, November 15, 2019, https://carnegieendowment.org/2019/11/15/legal-ethical-and-efficacy-dimensions-of-managing-synthetic-and-manipulated-media-pub-80439#analysis; and Charlotte Stanton, "June 2019 Convening on Defining Inappropriate Synthetic/Manipulated Media Ahead of the U.S. 2020 Election," June 19, 2019, Carnegie Endowment for International Peace, https://carnegieendowment.org/2019/06/19/june-2019-convening-on-defining-inappropriate-synthetic-manipulated-media-ahead-of-u.s.-2020-election-pub-79661.

Amber Frankland and Lindsay Gorman, "Combating the Latest Technological Threat to Democracy: A Comparison of Facebook and Twitter's Deepfake Policies," January 13, 2020, Alliance for Securing Democracy, https://securingdemocracy.gmfus.org/combating-the-latest-technological-threat-to-democracy-a-comparison-of-facebooks-and-twitters-deepfake-policies/; and Kate Cox, "Twitter Wants Your Feedback on Its Proposed Deepfakes Policy," Ars Technica, November 11, 2019, https://arstechnica.com/tech-policy/2019/11/twitter-wants-your-feedback-on-its-proposed-deepfakes-policy/.

Tim Bradshaw, ""Tech Debt': Why Badly Written Code Can Haunt Companies for Decades," Financial Times, November 27, 2019, https://www.ft.com/content/d6822eb0-0fe0-11ea-a7e6-62bf4f9e548a.

Huyskes D., 01.06.2021, Riprendiamoci le nostre emozioni dal controllo dell'intelligenza artificiale, Italian.Tech, https://www.italian.tech/blog/diritti-digitali/2021/06/01/news/l_automazione_delle_nostre_emozioni-303761227/

Lupis M., 10.06.2021, Orwel in Cina è il capufficio. La dura vita dei dipendenti cinesi, sorvegliati 24h, Huffington Post, https://www.huffingtonpost.it/entry/orwell-in-cina-e-il-capufficio-la-dura-vita-de-dipendenti-cinesi-sorvegliati-h24_it_60c21df1e4b0af343e9cf77f

Mattone M., 31.05.2021, L'equazione della felicità giocando con lo smartphone, Italian.Tech, https://www.italian.tech/2021/05/31/news/l_equazione_della_felicita_-303577559/

Redazione Diritto dell'informatica.it, 12.06.2020, L'Affective Computing: implicazioni giuridiche dell'Algoritmo emozionale, Diritto dell'informatica.it, http://www.dirittodellinformatica.it/privacy-e-sicurezza/laffective-computing-implicazioni-giuridiche-dellalgoritmo-emozionale.html

Renda S., 16.04.2021, Può l'intelligenza artificiale insegnarci a essere più umani, Huffington Post, https://www.huffingtonpost.it/entry/puo-un-algoritmo-insegnarci-a-essere-piu-umani_it_60783301e4b020e576c1af82

Verdoliva L., "Media Forensics and Deepfakes: an overview" IEEE Journal of Selected Topics in Signal Processing, 2020.

Agarwal and H. Farid, "Protecting World Leaders Against Deep Fakes", IEEE CVPR Workshops 2019.

Agarwal et al. "Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches", IEEE CVPR Workshops 2020.

Cozzolino et al., "ID-Reveal: Identity-aware DeepFake Video Detection", arXiv preprint arXiv:2012.02512, 2020.

Mittal et al. "Emotions Don't Lie: An Audio-Visual Deepfake Detection Method using Affective Cues", ACM Multimedia 2020.

GPDP | GARANTE PER LA PROTEZIONE DEI DATI PERSONALI

Deepfake Il falso che ti «ruba» la faccia (e la privacy)

Cosa è un deepfake?

I **deepfake** sono foto, video e audio creati grazie a software di intelligenza artificiale (AI) che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce.

La parola deepfake è un neologismo nato dalla fusione dei termini "fake" (falso) e "deep learning", una particolare tecnologia AI. Le tecniche usate dai deepfake sono simili a quelle delle varie app con cui ci si può divertire a modificare la morfologia del volto, a invecchiarlo, a fargli cambiare sesso, ecc.
La materia di partenza sono sempre i veri volti, i veri corpi e le vere voci delle persone, trasformati però in **"falsi" digitali**.

Le tecnologie deepfake, sviluppate come ausilio agli effetti speciali cinematografici, erano inizialmente molto costose e poco diffuse. Ma negli ultimi tempi hanno iniziato a diffondersi app e software che rendono possibile realizzare deepfake, anche molto ben elaborati e sofisticati, utilizzando un comune smartphone.
La diffusione dei deepfake è di conseguenza notevolmente aumentata, e con essa i rischi connessi.

Il deepfake e il furto di identità

Quella realizzata con i deepfake è una forma particolarmente grave di **furto di identità**.

Le persone che compaiono in un deepfake a loro insaputa non solo subiscono una perdita di controllo sulla loro immagine, ma sono private anche del controllo sulle loro **idee e sui loro pensieri**, che possono essere **travisati** in base ai discorsi e ai comportamenti falsi che esprimono nei video.

Le persone presenti nei deepfake potrebbero inoltre essere **rappresentate in luoghi o contesti o con persone che non hanno mai frequentato o che non frequenterebbero mai**, oppure in situazioni che potrebbero apparire compromettenti.

In sostanza, quindi, un deepfake può ricostruire contesti e situazioni mai effettivamente avvenuti e, se ciò non è voluto dai diretti interessati, può rappresentare una grave minaccia per la riservatezza e la dignità delle persone.

I gravissimi rischi del deepnude

In particolari tipologie di deepfake, dette **deepnude**, persone ignare possono essere rappresentate nude, in pose discinte, situazioni compromettenti (ad esempio, a letto con presunti amanti) o addirittura in contesti pornografici. Con la tecnologia del deepnude, infatti, i visi delle persone (compresi soggetti minori) possono essere "innestati", utilizzando appositi software, sui corpi di altri soggetti, nudi o impegnati in pose o atti di natura esplicitamente sessuale.
E' anche possibile prendere immagini di corpi vestiti e "spogliarli", ricostruendo l'aspetto che avrebbe il corpo sotto gli indumenti e creando immagini altamente realistiche.

Inizialmente il fenomeno ha coinvolto personaggi famosi allo scopo di screditarli o ricattarli. Ma negli ultimi tempi, con la sempre maggiore diffusione di software che utilizzano questa tecnologia, il rischio coinvolge anche persone comuni, le quali possono diventare oggetto di azioni psicologicamente e socialmente molto dannose. Come, ad esempio, il **"revenge porn"**, cioè la condivisione online - a scopo di ricatto, denigrazione o vendetta, da parte di ex partner, amanti o spasimanti respinti - di foto e video a contenuto sessuale o addirittura pornografico, che, nel caso del deepnude, sono ovviamente falsi.

Video deepnude possono essere utilizzati, a totale insaputa dei soggetti rappresentati nelle immagini, anche per alimentare la pratica del **sexting** (cioè lo scambio e diffusione di immagini di nudo, che a volte coinvolge anche soggetti minori), la **pornografia illegale** e, purtroppo, anche reati gravissimi come la **pedopornografia**.

Deepfake e cyberbullismo

I video deepfake possono essere creati ad hoc per realizzare veri e propri atti di cyberbullismo, che hanno come vittime soprattutto giovani.
Un deepfake può essere realizzato per denigrare, irridere e screditare le persone coinvolte, o addirittura per ricattarle, chiedendo soldi o altro in cambio della mancata diffusione del video oppure per la sua cancellazione se è già stato diffuso.

Deepfake e fake news

I deepfake possono riguardare politici o opinion leader, con lo scopo di influenzare l'opinione pubblica. Video deepfake possono ad esempio essere mostrati o inviati agli elettori che simpatizzano per un determinato personaggio politico, rappresentandolo mentre compie azioni poco lecite o mentre si trova in situazioni sconvenienti, allo scopo di screditarlo ed **influenzare le opinioni o il voto**. In questo modo, i deepfake possono purtroppo contribuire alla diffusione di fake news e alla disinformazione.

Il deepfake può quindi arrivare a privare le persone della cosiddetta "autodeterminazione informativa" ("ciò che voglio far sapere di me lo decido io"), come pure ad incidere sulla loro libertà decisionale ("quello che penso e faccio è una scelta su cui gli altri non possono interferire").

Deepfake e cybercrime

Il deepfake può essere utilizzato per attività telematiche illecite, come lo spoofing (il furto di informazioni che avviene attraverso la falsificazione di identità di persone o dispositivo, in modo da ingannare altre persone o dispositivi e ottenere la trasmissione di dati), il phishing e il ransomware.

Volti e voci artefatti possono essere utilizzati per ingannare i sistemi di sicurezza basati su dati biometrici vocali e facciali.
Ad esempio, video o audio-messaggi deepfake creati da malintenzionati possono essere inviati ai nostri colleghi, amici o parenti **per invitarli a cliccare su link o aprire allegati a messaggi** che espongono pc, smartphone o altri dispositivi e sistemi a pericolose intrusioni, oppure per convincerli a fornire, ingannando la loro fiducia, informazioni e dati sensibili. Inoltre oggi molti sistemi digitali (domotica, assistenti vocali, smartphone, nonché alcuni sistemi bancari o sanitari) ricorrono a **dati biometrici vocali e facciali come sistema di autenticazione per l'accesso**. Video e audio-messaggi deepfake potrebbero essere utilizzati per **ingannare** tali sistemi.

Anche se al momento il livello avanzato delle tecnologie di sicurezza e la ancora relativa imprecisione dei deepfake stanno limitando questi fenomeni, l'attenzione deve essere comunque alta.

Come proteggersi dai deepfake

Le grandi imprese del digitale (piattaforme social media, motori di ricerca, ecc.) stanno già studiando e applicando delle **metodologie per il contrasto al fenomeno**, come algoritmi di intelligenza artificiale capaci di individuare i deepfake o sistemi per le segnalazioni da parte degli utenti, e stanno formando team specializzati nel monitoraggio e contrasto al deepfake. E le **Autorità di protezione dei dati personali** possono intervenire per prevenire e sanzionare le violazioni della normativa in materia di protezione dati.

Tuttavia, **il primo e più efficace strumento di difesa è rappresentato sempre dalla responsabilità e dall'attenzione** degli utenti. **Ecco allora alcuni suggerimenti**:

- **Evitare di diffondere in modo incontrollato immagini personali o dei propri cari**. In particolare, se si postano immagini sui social media, è bene ricordare che le stesse potrebbero rimanere online per sempre o che, anche nel caso in cui si decida poi di cancellarle, qualcuno potrebbe già essersene appropriato.

- Anche se non è semplice, **si può imparare a riconoscere un deepfake**. Ci sono elementi che aiutano: l'immagine può appare pixellata (cioè un pò "sgranata" o sfocata); gli occhi delle persone possono muoversi a volte in modo innaturale; la bocca può apparire deformata o troppo grande mentre la persona dice alcune cose; la luce e le ombre sul viso possono apparire anormali.

- **Se si ha il dubbio che un video o un audio siano un deepfake realizzato all'insaputa dell'interessato, occorre assolutamente evitare di condividerlo** (per non moltiplicare il danno alle persone con la sua diffusione incontrollata). E si può magari decidere di segnalarlo come possibile falso alla piattaforma che lo ospita (ad esempio, un social media).

- **Se si ritiene che il deepfake sia stato utilizzato in modo da compiere un reato o una violazione della privacy,** ci si può rivolgere, a seconda dei casi, alle autorità di polizia (ad esempio, alla Polizia postale) o al **Garante per la protezione dei dati personali**.

Per gli autori

La collaborazione è aperta agli studiosi ed esperti di ogni indirizzo. Sulla pubblicazione di scritti e contributi decide il Comitato Scientifico entro 60 giorni dal ricevimento dopo aver verificato che la proposta sia conforme alle norme redazionali e che il manoscritto non sia stato già pubblicato in altra sede. I materiali inviati non verranno restituiti.
La Rivista pubblica anche recensioni di libri.
La Rivista si ispira alla Dichiarazione di Berlino per l'accesso aperto alla letteratura scientifica pertanto l'autore o gli autori devono singolarmente allegare la dichiarazione all'autorizzazione alla pubblicazione in open access(allegato finale). Le firme digitali sono accettate.

Norme redazionali

1. Cosa spedire alla redazione

Articolo deve essere inviato in formato Word, non utilizzando in nessun caso programmi di impaginazione grafica. Non formattare il testo in alcun modo (evitare stili, bordi, ombreggiature …). Se i contributi sono più d'uno, devono essere divisi in diversi file, in modo che a ciascuna unità di testo corrisponda un diverso file. I nomi dei file devono essere contraddistinti dal cognome dell'autore. Nel caso di più contributi di uno stesso autore si apporrà un numero progressivo (es.: baccaro.doc, baccaro1.doc, ecc.).
Si tenga presente che i singoli articoli sono raggiungibili in rete attraverso i motori di ricerca. Suggeriamo dunque di utilizzare titoli che sintetizzino con chiarezza i contenuti del testo e che contengano parole chiave a questi riferiti.

Allegare al file dell'articolo completo:

- un abstract (max 1000 caratteri) in italiano, inglese ed eventualmente anche in spagnolo.
- una breve nota biografica dell'autore/trice. A tale scopo dovranno essere comunicati i titoli accademici ed eventuale indirizzo di posta elettronica e/o eventuale Ente di appartenenza.
- le singole tabelle e le immagini a corredo dei contenuti, devono essere in file separati dal testo, numerati per inserirli correttamente nel testo stesso e accompagnate da didascalia e citazione della fonte.
- inserire il materiale (abstract, cenno biografico, indice, testo dell'articolo, bibliografia, siti consigliati) in un unico file, lasciando a parte solo le immagini e le tabelle.
- la bibliografia deve essere collocata in fondo all'articolo.

2. Norme per la stesura dell'articolo

Nel caso in cui l'articolo superi le due cartelle è preferibile suddividere lo scritto in paragrafi titolati, o in sezioni, evidenziati in un indice all'inizio dell'articolo.

Il testo deve avere una formattazione standard, possibilmente con le seguenti caratteristiche:
- testo: garamond 12;
- interlinea "1,15 pt";
- titolo capitolo: garamond 12 grassetto;
- titoli paragrafi: garamond 12;
- evitare soprattutto i rientri (non inserire tabulazioni a inizio capoverso);
- non sillabare;
- evitare le virgolette a sergente «», ma usare solo virgolette alte (" ");
- non usare le virgolette semplici (' ') e preferire le virgolette inglesi (' ' " ");
- fare attenzione all'uniformità dello stile quando si fanno copia/incolla di testi soprattutto provenienti da Internet;
- evitare sempre il maiuscoletto e il maiuscolo e il sottolineato.

Un termine che ammette due grafie differenti deve sempre essere scritto nello stesso modo (per esempio, i termini "psicoanalisi" e "psicanalisi" sono entrambi corretti, ma è importante utilizzarne uno solo per tutto il testo).
Le parole in lingua straniera (ad es. in latino) ed espressioni quali *en passant* vanno scritte in corsivo.
Il riferimento alle illustrazioni va scritto nel seguente modo: (Fig. 1).
Corsivo e virgolette vanno evitati come effetti stilistici.
Si raccomanda il rispetto di alcune convenzioni come le seguenti: p. e pp. (e non pag. o pagg.); s. e ss. (e non seg. e segg.); cap. e capp.; cit.; cfr.; ecc.; vol. e voll.; n. e nn.; [*N.d.A.*] e [*N.d.T.*].
I numeri di nota dovranno sempre precedere i segni di interpunzione (punti, virgole, punti e virgole, due punti ecc.), ma seguire le eventuali virgolette di chiusura. Esempio: "Nel mezzo del cammin di nostra vita"[23].
La frase deve sempre finire con il punto. Esempio: Verdi, nel 1977 (87) si chiedeva: "Perché l'alleanza non resse?".

a. Note a piè di pagina

Per le note a pié pagina usare corpo 10 Times New Roman.

b. Elencazioni di punti

Rientrare di cm 0,5. Se sotto lo stesso punto sono riportati più periodi, rientrare la prima riga dei periodi successivi al primo di cm 1.
Quando l'elencazione è preceduta da una frase che finisce con due punti, fare minuscola la prima parola di ogni punto (se non è un nome proprio) e mettere il punto e virgola dopo l'ultima parola di ogni singolo punto. Quando invece la frase che precede l'elencazione finisce con il punto, fare maiuscola l'iniziale della prima parola e mettere il punto dopo l'ultima parola.
Preferire per contrassegnare i punti al trattino tradizionale un simbolo grafico, non variando ogni volta il simbolo usato.

c. Citazioni

- Citazioni nel testo

Le citazioni brevi (fino ad un massimo di due righe) vanno riportate tra virgolette. Citazioni più lunghe si riportano senza virgolette, ma vanno evidenziate lasciando una riga prima e dopo la citazione, in modo tale che quest'ultima rimanga distinta dal corpo del testo ma senza rientro.
Le omissioni si segnalano esclusivamente con tre puntini tra parentesi quadre: […].

- Citazioni da web

Delle fonti reperite in rete va dato conto con la stessa precisione (e anzi maggiore) delle fonti cartacee. Se ricostruibili, vanno indicati almeno autore, titolo, contenitore (ossia il sito, la rivista *online*, o il portale che contiene il documento citato), data del documento, URL (tra parentesi angolari), e data della visita (tra parentesi tonde), come nell'esempio sotto riportato. Gli indirizzi (URL) vanno scritti per esteso, senza omettere la parte iniziale, l'indicatore di protocollo (es.: http://), ed evitando di spezzarli (se necessario, andare a capo prima dell'indirizzo).
es.: Pellizzi F., *I generi marginali nel Novecento letterario*, in «Bollettino '900», 22 maggio 1997, <http://www3.unibo.it/boll900/convegni/gmpellizzi.html> (15 agosto 2004).

d. Figure

Tutte le figure devono essere numerate, in modo progressivo iniziando da uno per ogni capitolo. Nel testo è necessario indicare la posizione esatta in cui inserire le foto e le tabelle (nel caso creare un elenco a parte) e riportare la didascalia, comprendente eventuale indicazione dell'autore il soggetto, luogo, anno, la fonte.
In didascalia di solito si utilizza l'abbreviazione tab., fig..
Le immagini dovranno essere caricate in files a parte debitamente numerati con numerazione progressiva che rispetti l'ordine di inserimento nel saggio.
Nel testo non si può scrivere «come evidenzia la tabella seguente:...» dato che ciò creerebbe la rigidità di doverla necessariamente collocare dopo i due punti. È molto più vantaggioso numerare progressivamente per capitolo tutte le figure e le tabelle e scrivere ad es. «come evidenzia la tab. 2», in modo che questa può essere inserita in qualsiasi punto della pagina o addirittura in quella a fronte, dove risulta più comodo ed esteticamente più confacente: ad es. all'inizio pagina, sopra il riferimento nel testo.
Il formato dei file grafici deve essere tra i più diffusi, preferibilmente Jpeg o Gif o Tiff.
Per le tabelle e i grafici è da preferire il formato excel o trasformate in Jpeg.

e. Titoli e sottotitoli

Titolo capitolo: non centrarli sulla pagina ma allinearli a sinistra. La distanza tra il titolo, se è di una riga, e il testo o il titolo del paragrafo è di 10 spazi in corpo 12.
Titoli paragrafi, sottoparagrafi e sotto-sottoparagrafi e altri titoli o parole in evidenza su riga a sé: lasciare 2 righe bianche prima di digitarli e ancora una riga bianca dopo averli digitati. Se il titolo finisse a fine pagina spostarlo alla pagina successiva aumentando il numero di righe bianche (di norma una o due sono sufficienti). Anche i titoli dei paragrafi, sotto paragrafi, ecc. sono allineati a sinistra, senza rientro.

f. Bibliografia

Gli autori sono invitati a utilizzare la bibliografia secondo i criteri illustrati di seguito, perché consente di ridurre l'uso delle note bibliografiche che, per un testo visionabile sul video, distolgono l'attenzione dal contenuto.

◊ *titoli dei periodici e dei libri* in corsivo senza virgolette inglesi;

◊ *titoli degli articoli* tra "virgolette inglesi" (si trovano in "inserisci - simbolo");

◊ *nome autore*: nel testo il cognome dell'autore va preceduto, quando citato, dal nome; nella bibliografia alla fine del capitolo o del libro e nelle citazioni bibliografiche in nota mettere sempre prima il cognome. Non mettere la virgole tra il cognome e il nome dell'autore ma solo (nel caso di più autori) tra il primo autore e quelli successivi digitando preferibilmente una "e" prima del nome dell'ultimo autore;

◊ *data di pubblicazione:* metterla tra parentesi dopo il nome; per gli articoli dopo il nome della rivista o dopo il numero del fascicolo, sempre divisa da una virgola.

◊ *editore*: metterlo solo per i volumi, dopo il titolo, separato da questo da una virgola. Mettere, quindi, sempre dopo una virgola, il luogo di pubblicazione;

Esempi:

Mowen J.C., Mowen M.M. (1991), "Time and outcome evaluation", *Journal of marketing,* 55: 54-62.

Murray H.A. (1938), *Explorations in personality*, Oxford University Press, New York.

- Bibliografia nel testo

Le indicazioni bibliografiche devono essere espresse direttamente nel testo fra parentesi tonde, secondo il seguente schema.

• Nome dell'autore (se non espresso nel testo) e anno di pubblicazione senza virgola:

Uno studio recente (Neretti, 1999) ha confermato questa opinione.

Il recente studio di Neretti (1999) ha confermato questa opinione.

I recenti studi di Neretti (1999; 2000; 2001a; 2001b) hanno confermato questa opinione.

Recenti studi (Bianchi, 2000; Neretti, 1999; Vitali, 2001) hanno confermato questa opinione.

• L'eventuale numero della pagina in cui si trova la citazione, obbligatorio quando la citazione è diretta, è separato da virgola senza nessuna sigla (Neretti, 1999, 54).

- Riviste

Cognome dell'autore e iniziale del nome puntato, anno di pubblicazione fra parentesi, separato da uno spazio, *titolo in corsivo*, nome della rivista tra virgolette preceduto da "in", numero della rivista.

Esempio:

Alberti G. (1999), *Democratizzazione e riforme strutturali*, in "Politica Internazionale", nn. 1-2.

Per le riviste, non si ritiene necessario il luogo di pubblicazione, né l'indicazione della pagina esatta in cui si trova l'articolo.

- Articoli di periodico

titolo tra virgolette, nome del periodico - per esteso o in forma abbreviata in corsivo – numero del volume, pagine di riferimento:

Stevenson T. (2003), "Cavalry uniforms on the Parthenon frieze", *American Journal of Archeology* 104, 629-654.

Nel caso di un periodico composto da vari fascicoli con numerazione separata nell'ambito della stessa annata, si scrive: 104/4

- Articolo di giornale

Nelle citazioni da quotidiani, al nome dell'autore e al titolo dell'articolo si fanno seguire il titolo del giornale tra virgolette angolari, giorno, mese e anno della pubblicazione.

- Tesi di laurea

Dopo il nome e il cognome dell'autore e il titolo, che si riportano con le stesse norme usate per i libri, si aggiunge il nome del relatore, la Facoltà e l'Università di appartenenza, l'anno accademico in cui la tesi è stata discussa.

Il materiale deve essere inviato esclusivamente a: rivistapsicodinamica.criminale@gmail.com
Gli Autori riceveranno una mail di conferma del ricevimento del materiale.

I dati personali conferiti vengono trattati con il rispetto della normativa relativa alla tutela della privacy e in particolare ai sensi del D.Lgs. 196 del 2003.

Dichiarazione

La sottoscritta (o il sottoscritto)_____
Nata/o a_____il _____
Residente in via_____

Città_____ tel._____mail_____

con la presente

AUTORIZZA

la pubblicazione a titolo gratuito nella rivista on line open access "Rivista di psicodinamica criminale" dell'articolo dal titolo

_____

Firma _____

Data

Questa rivista è pubblicata sotto licenza Creative Commons Attribution 3.0.
ISSN 2037-1195
Editore proprietario: Associazione "Psicologo di strada"
e-mail: rivistapsicodinamica.criminale@gmail.com