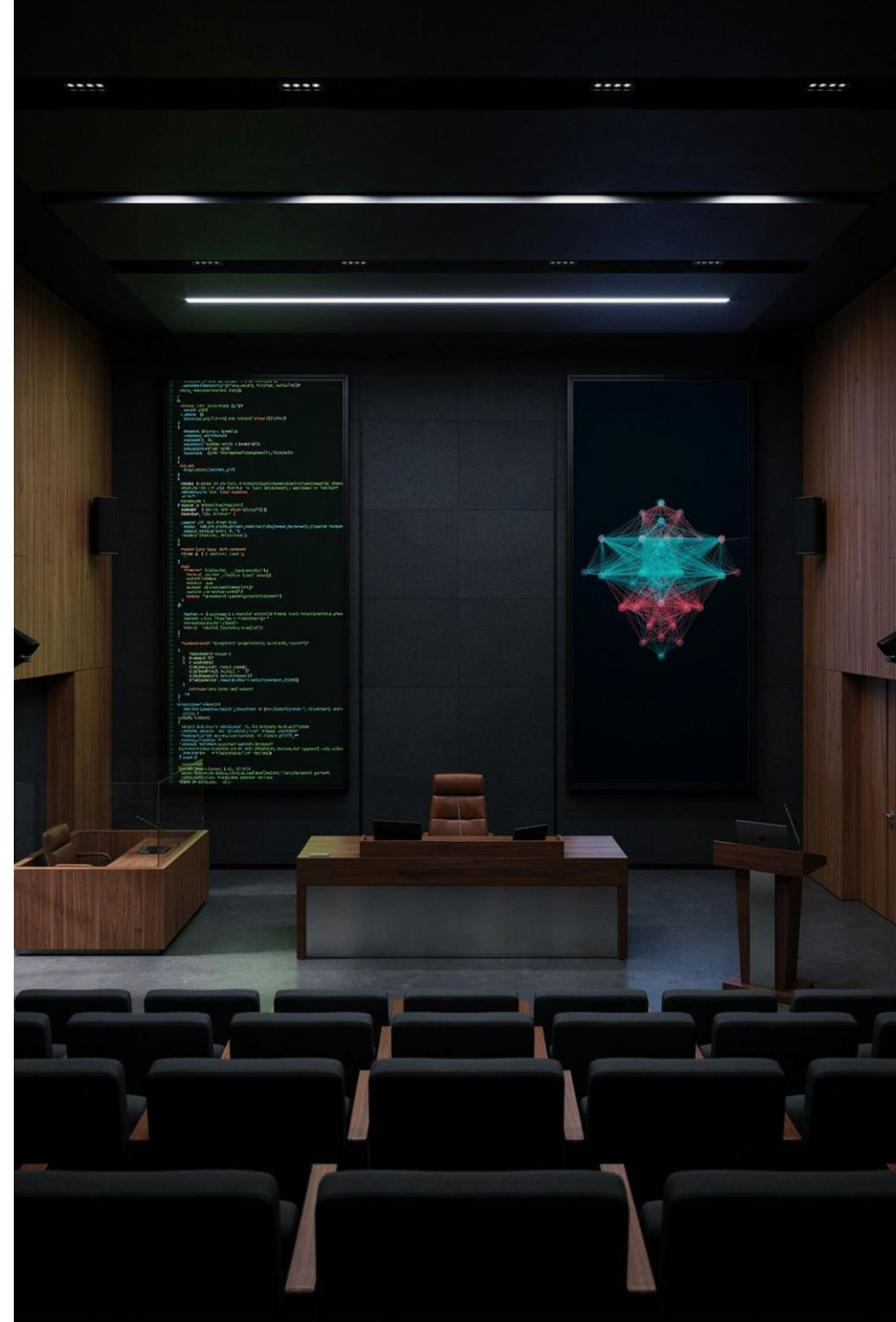


**ONIF - OSSERVATORIO NAZIONALE DI INFORMATICA FORENSE  
GENOVA, 16 OTTOBRE 2026**

# **INTELLIGENZA ARTIFICIALE E DIGITAL FORENSICS: VERSO UN NUOVO PARADIGMA PROBATORIO**

**TRA OPACITÀ ALGORITMICA E NUOVI REATI: IL CASO DELL'ART. 612-QUATER C.P.**





01000001 01001100 001000  
01000101 01001110 01010  
01010011 00100000 01000001 01001

# CHI SONO .. COSA HO FATTO... COSA FACCIO



## Professionista digital forensics expert

Attualmente consulente libero professionista, Local Digital Forensics Senior Auditor e Digital Forensics Manager presso BIT4LAW SRL di Bologna - Lead Auditor ISO/IEC 27001/2022, Digital Evidence Specialist (DES)



## Carriera nelle forze dell'ordine

Ufficiale di Polizia Giudiziaria ed Economico Finanziaria per oltre 38 anni nella Guardia di Finanza, con incarichi come Computer Forensics Data Analysis e Formatore First Responder.



## Attività accademica e divulgativa

Docente di Digital Forensics presso varie associazioni, enti e Università .Relatore in convegni di settore sulla Digital Forensics, Mobile Forensics, GDPR e Cyber Security.



## Pubblicazioni e affiliazioni

Autore di numerosi articoli tecnico-giuridici, coautore di saggi pubblicati sul Memberbook IISFA. Consigliere di Lab4Int, Socio IISFA, ONIF, OSINITALIA e AIP-ITCS

[www.linkedin.com/in/pier-luca-toselli-5a71a562](https://www.linkedin.com/in/pier-luca-toselli-5a71a562)

# L'EVOLUZIONE DELLA SCENA DEL CRIMINE DIGITALE

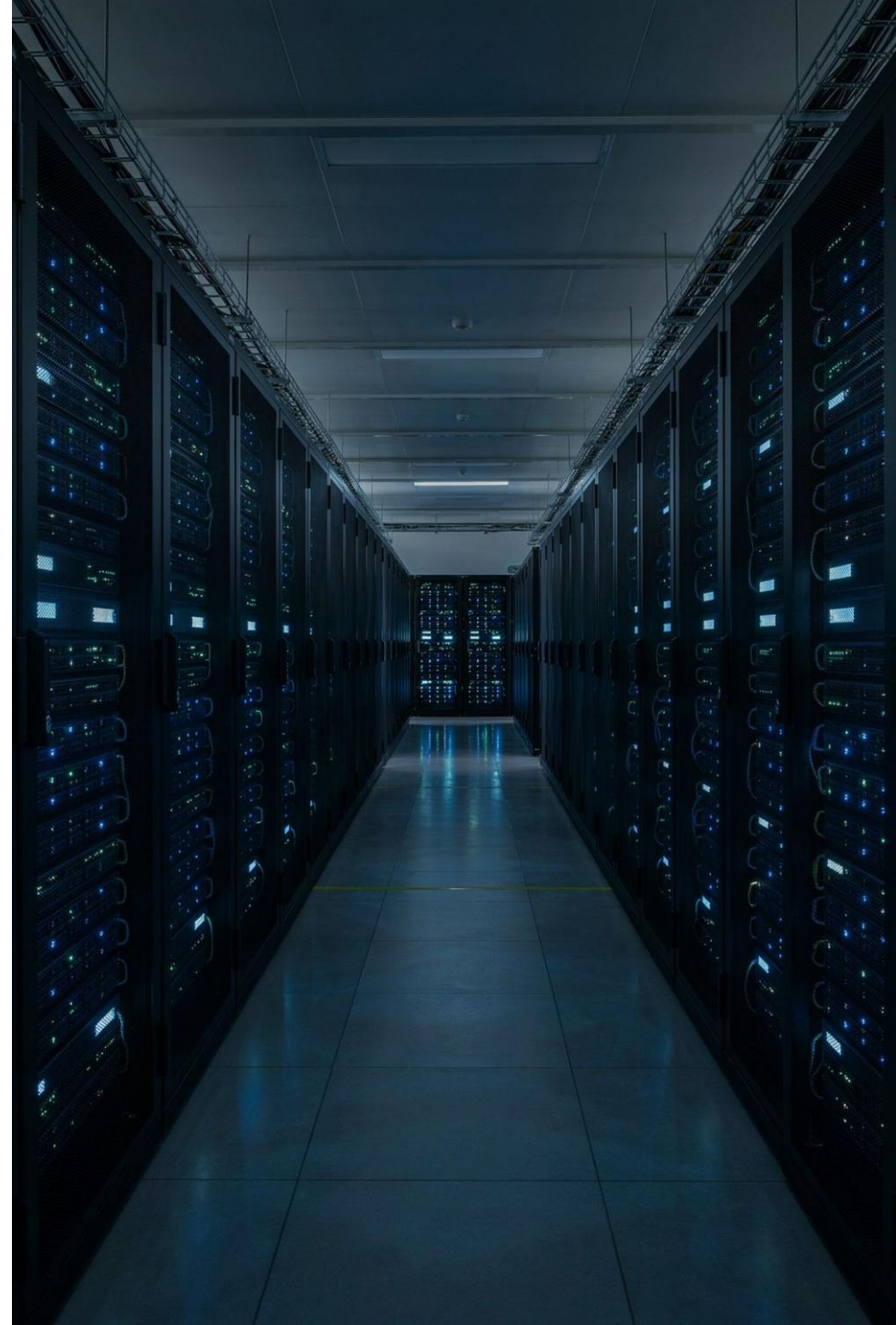
DALLA REPERTAZIONE FISICA STATICA AGLI ECOSISTEMI DINAMICI DISTRIBUITI

## PARADIGMA TRADIZIONALE (LEGACY FORENSICS)

- Reperto fisico ben circoscritto: disco fisso, memoria USB, smartphone sequestrato fisicamente.
- Immagine forense bit-stream (Raw/dd o E01) congelata nel tempo con hash invariabile.
- Processi software interamente deterministici basati su codice sorgente compilato staticamente.

## SCENA DEL CRIMINE CONTEMPORANEA (AI "ERA")

- Ecosistemi cloud nativi, cluster GPU distribuiti transnazionalmente e container effimeri.
- Algoritmi con auto-apprendimento continuo (Continuous Learning) e pesi sinaptici mutabili.
- L'AI non è un semplice file isolato: è una pipeline di dati, modelli matematici e risposte stocastiche.



# LA TRIPLICE DIMENSIONE PROCESSUALE DELL'INTELLIGENZA ARTIFICIALE

## INQUADRAMENTO DELLA CONDOTTA AI FINI DELL'INDAGINE TECNICA E GIURIDICA

### AI COME STRUMENTO (MEZZO)

L'agente umano utilizza l'AI generativa per perpetrare il reato: Deepfake diffamatori, spear phishing mirato, malware polimorfo.

Focus Peritale: Prompt engineering forense, correlazione oraria, estrazione log e chiavi API.

### AI COME VITTIMA (TARGET)

Il modello viene attaccato o corrotto dall'esterno: Data Poisoning, Adversarial Attacks, Model Inversion o Denial of Service algoritmico.

Focus Peritale: Integrità dei dataset di training, analisi dei gradienti, alterazione pesi.

### AI COME 'AUTORE' (AUTONOMIA)

Sistemi cyber-fisici autonomi (veicoli a guida autonoma, bot di High-Frequency Trading) che generano eventi lesivi.

Focus Peritale: Tracciabilità delle decisioni, alberi decisionali e rispetto dell'obbligo di controllo umano (Human-in-the-loop).

# IL LATO OSCURO DELL'AI NELLE INVESTIGAZIONI FORENSI

QUATTRO MINACCE DIRETTE ALLA GENUINITÀ E VERIFICABILITÀ DELLA PROVA

## GENERAZIONE DI PROVE SINTETICHE

Creazione di file audio, video o documentali sintetici indistinguibili a occhio nudo, utilizzati per calunniare o costruire falsi alibi processuali.

## ANTI-FORENSICS POTENZIATA

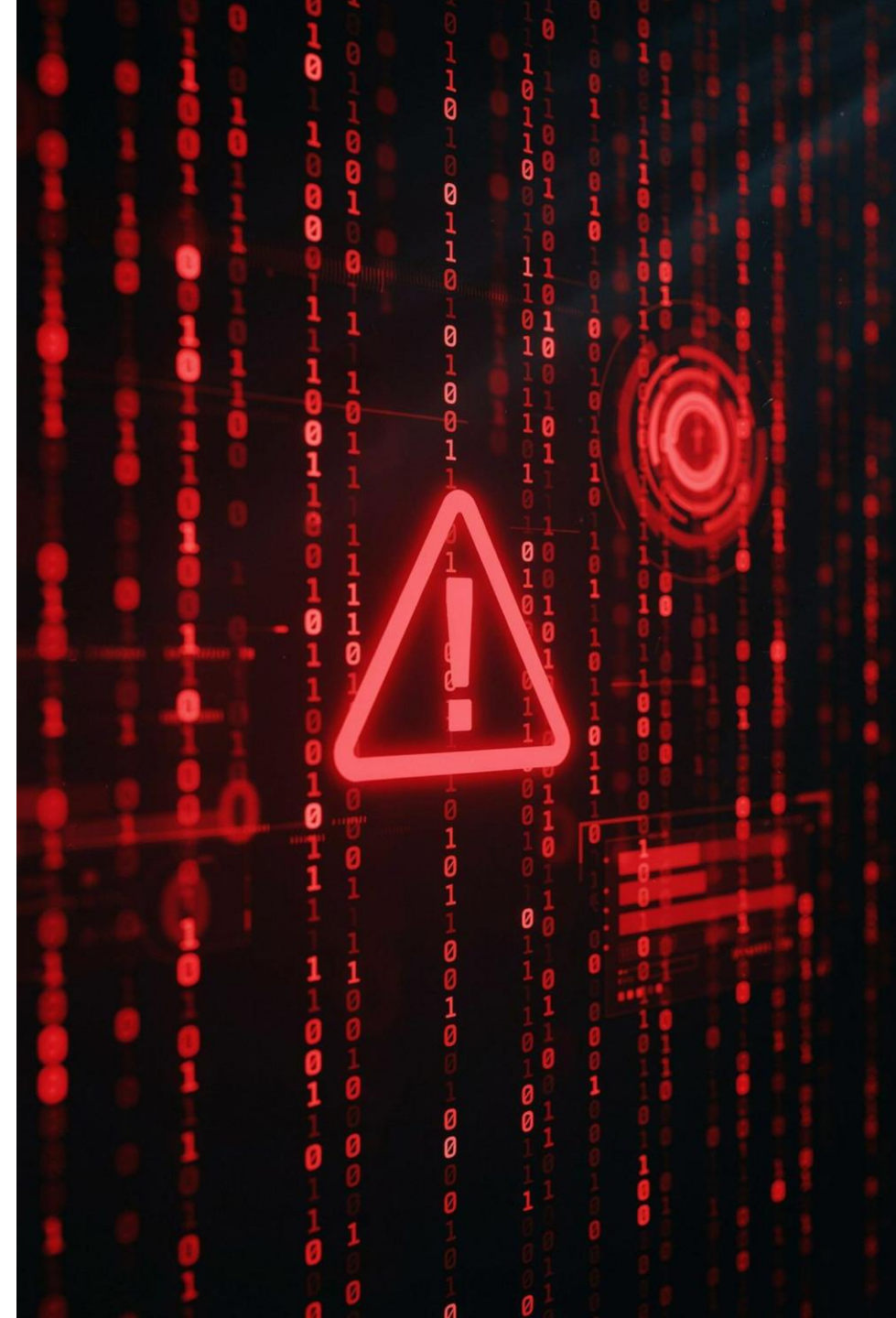
Malware dotati di micro-modelli neurali capaci di rilevare ambienti di analisi forense (sandbox/debugger) e cancellare selettivamente le tracce.

## OPACITÀ DECISIONALE (BLACK BOX)

Impossibilità strutturale di ricostruire punto per punto il percorso causale che ha portato una rete neurale complessa all'output incriminato.

## DATA POISONING STRUTTURALE

Corruzione mirata dei dataset di riferimento usati dai tool investigativi stessi (es. sistemi biometrici o parser forensi automatizzati).



# IL PROBLEMA DELLA 'BLACK BOX' NEL CONTRADDITTORIO PENALE

COME CONCILIARE L'OPACITÀ DELLE RETI NEURALI CON IL GIUSTO PROCESSO EX ART. 111 COST.



## **NATURA STOCASTICA DEL DEEP LEARNING:**

I modelli complessi elaborano dati attraverso milioni (o miliardi) di parametri sinaptici non lineari. Il risultato è un output statistico, non un sillogismo trasparente.



## **CRISI DELL'ESAME INCROCIATO (CROSS-EXAMINATION):**

Nel processo accusatorio, la difesa ha il diritto costituzionale di verificare come si sia formata la prova. L'affermazione peritale 'il software AI ha stabilito che è colpevole' non è legalmente ammissibile.



## **LA RESPONSABILITÀ GIUDIZIALE (PERITUS PERITORUM):**

Il Giudice non può motivare una sentenza recependo acriticamente l'esito di una 'scatola nera' algoritmica non ispezionabile, a pena di nullità per vizio di motivazione.

# IL MODELLO COME CORPO DEL REATO

## PROTOCOLLO OPERATIVO DI CRISTALLIZZAZIONE FORENSE DELL'ARTEFATTO ALGORITMICO

**1**

### **ACQUISIZIONE DEL CHECKPOINT**

Congelamento immediato dell'architettura e dei vettori di peso (.onnx, .safetensors, .pt, .h5). Calcolo contestuale dell'hash SHA-256 per garantire l'invariabilità dello stato neurale.

**2**

### **TRACCIAMENTO DEGLI IPERPARAMETRI**

Sequestro dei file di configurazione (config.json, tokenizers, optimizer state, seed casuali). Senza questi elementi è impossibile replicare esattamente il comportamento del modello.

**3**

### **AUDITING DEI DATI DI ORIGINE (DATA LINEAGE)**

Ispezione forense dei repository dei dataset di training e fine-tuning (sistemi DVC / Git LFS) per identificare inserimenti malevoli o alterazioni anomale della distribuzione statistica.

# DIGITAL FORENSICS TRADIZIONALE VS. MODEL FORENSICS

DUE DISCIPLINE SINERGICHE PER LA COSTRUZIONE DELLA PROVA INFORMATICA

| PARAMETRO PERITALE  | DIGITAL FORENSICS “CANONICA”                                                                      | MODEL FORENSICS APPLICATA                                                                       |
|---------------------|---------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| OGGETTO PRINCIPALE  | Supporti fisici (Hard Disk, SSD, USB), partizioni, file system, registri, memoria volatile (RAM). | Modello matematico, tensori, pesi sinaptici, file di architettura, prompt, dataset di training. |
| DOMANDA D'INDAGINE  | 'Chi ha compiuto l'azione e tramite quale dispositivo o utenza di sistema?'                       | 'Perché l'algoritmo ha preso tale decisione? Quali iperparametri hanno causato l'output?'       |
| STRUMENTI APPLICATI | FTK, EnCase, X-Ways Forensics, Autopsy, Volatility Framework, Cellebrite UFED.                    | SHAP (Shapley Additive exPlanations), LIME, Integrated Gradients, Captum, PyTorch Profiler.     |
| STATO DEL REPERTO   | Statico: cristallizzazione di un'immagine bit-a-bit (E01, RAW) congelata nel tempo.               | Dinamico/Evolutivo: Continuous Learning, fine-tuning incrementale e drift stocastico.           |

# DATA POISONING E ATTACCHI AVVERSARI: L'INDAGINE PERITALE

COME INDIVIDUARE LA CORRUZIONE DOLOSA DEL PROCESSO DECISIONALE ALGORITMICO

## MODALITÀ DI ATTACCO ALGORITMICO

- Backdoor Poisoning: inserimento silente di pattern 'trigger' nel dataset per condizionare il modello a sbagliare unicamente in determinate circostanze.
- Adversarial Perturbations (es. FGSM): iniezione di rumore calibrato invisibile all'occhio umano ma capace di ingannare totalmente classificatori biometrici.
- Model Inversion & Extraction: esfiltrazione non autorizzata dei dati sensibili contenuti nella memoria del modello tramite query appositamente costruite.

## APPROCCIO INVESTIGATIVO INTEGRATO

- Digital Forensics: individua l'intrusione telematica sui server di sviluppo, le sessioni abusive e i log di commit malevoli (Art. 615-ter c.p.).
- Model Forensics: esamina la superficie di decisione del modello, misura lo shift prestazionale e dimostra il nesso causale tra intrusione ed errore del sistema.
- Integrazione peritale: la convergenza probatoria individua ed inchioda l'autore dell'attacco e dimostra scientificamente l'alterazione del software.

# LA RIFORMA PENALE IN MATERIA DI INTELLIGENZA ARTIFICIALE

LEGGE N. 132 DEL 23 SETTEMBRE 2025: IL RECEPIMENTO DELL'AI ACT E LA RISPOSTA SANZIONATORIA

## L'ADEGUAMENTO DELL'ORDINAMENTO NAZIONALE:

La Legge n. 132/2025 recepisce e coordina l'impianto penalistico italiano con il Regolamento Europeo sull'AI (Reg. UE 2024/1689), colmando pericolosi vuoti di tutela.

## LA TIPIZZAZIONE DI UNA NUOVA FATTISPECIE AUTONOMA:

Introduzione dell'Art. 612-quater c.p., norma ad hoc per punire la creazione e diffusione non consensuale di contenuti multimediali alterati con sistemi AI.

## LA NUOVA CIRCOSTANZA AGGRAVANTE COMUNE:

Inasprimento sanzionatorio per qualsiasi reato comune previsto dal codice penale quando sia commesso mediante l'impiego insidioso di strumenti di intelligenza artificiale.

# L'ART. 612-QUATER C.P.

## STRUTTURA DELLA FATTISPECIE CRIMINOSA E CONDOTTE PERFEZIONATIVE DEL REATO

**1**

### **OGGETTO MATERIALE DELLA CONDOTTA**

Immagini, filmati o registrazioni vocali ritraenti la persona offesa, falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale, senza il consenso dell'avente diritto.

**2**

### **EVENTO DI DANNO TIPICO**

Produzione di un danno ingiusto alla persona: lesione dell'onore, della reputazione sociale/professionale, violazione della riservatezza o pregiudizio di natura patrimoniale.

**3**

### **CONDOTTE CONDANNABILI (ALTERNATIVE)**

La norma punisce disgiuntamente tre condotte:

- Cessione a singoli destinatari determinati
- Pubblicazione su piattaforme o bacheche aperte
- Diffusione attraverso qualsiasi circuito telematico o chat

# IL DUPLICE REQUISITO

I DUE REQUISITI PER LA SUSSISTENZA DEL REATO EX ART. 612-QUATER C.P.

## PILASTRO 1: PROVA DELLA MANIPOLAZIONE CON AI

- Non è sufficiente provare un generico ritocco o fotoritocco tradizionale (Photoshop / Cut-and-Paste).
- Il consulente tecnico ha l'onere di dimostrare che la falsificazione è scaturita da un sistema di intelligenza artificiale (GAN, VAE, Reti di Diffusione).
- Occorre individuare gli artefatti matematici specifici lasciati dall'architettura neurale impiegata.

## PILASTRO 2: IDONEITÀ DEL FILE AD INGANNARE

- Il contenuto deve possedere una verosimiglianza oggettiva idonea a trarre in errore i destinatari circa la sua genuinità.
- Il falso grossolano o manifestamente surreale esclude la tipicità della condotta criminosa prevista dalla norma.
- Valutazione contestuale: il perito deve rapportare la perizia tecnica al canale di diffusione e al target del pubblico ricevente.

# CASO PRATICO: LA SERIALITÀ DELLE PIATTAFORME DEEPPFAKE

## LA DIMENSIONE CRIMINOLOGICA DEL CASO:

Piattaforme illegali (es. caso SocialMediaGirls e circuiti Telegram collegati) con milioni di iscritti dedite alla generazione e scambio seriale di contenuti 'deepnude' non consensuali a danno di giornaliste, volti noti e comuni cittadini.

## LA PIPELINE TECNOLOGICA CRIMINALE:

1. Scraping massivo di immagini da profili pubblici e fotogrammi da videocamere connesse.
2. Elaborazione automatizzata tramite bot e script di swapping/nude-generator open source.
3. Condivisione seriale su canali anonimi e server offshore per monetizzazione illecita.

## LA RISPOSTA DELLA DIGITAL FORENSICS:

Acquisizione forense conforme a ISO/IEC 27037 delle evidenze online, tracciamento dei flussi crittografici di pagamento e disamina dei log di download/upload.

# LE SFIDE OPERATIVE NELLA RICERCA DELLA PROVA

## QUATTRO OSTACOLI STRUTTURALI ALLA CRISTALLIZZAZIONE DEI DEEPPFAKE

### **RICOMPRESSIONE LOSSY SUI SOCIAL**

WhatsApp, Instagram e TikTok comprimono aggressivamente file audio/video, cancellando micro-artefatti e frequenze ad alta precisione.

### **METADATA STRIPPING SISTEMATICO**

Le piattaforme eliminano di default i metadati EXIF e le intestazioni di origine, impedendo l'attribuzione diretta del dispositivo.

### **EFFETTO SCREENSHOT DEPOSITATO**

Spesso la parte offesa fornisce come prova lo screenshot di uno smartphone, moltiplicando la perdita di integrità del dato primario.

### **RAPIDITÀ EVOLUTIVA DEI MODELLI**

L'uscita incessante di nuovi modelli generativi rende rapidamente obsoleti i classificatori addestrati sulle generazioni precedenti.

# L'ACCERTAMENTO FORENSE

MISURARE L'EVIDENZA TECNICA TRA PERCEZIONE UMANA E RIGORE SCIENTIFICO

## APPROCCIO METROLOGICO / QUANTITATIVO

- Misurazione della fedeltà biometrica tra la persona offesa e il soggetto falsificato (Face Recognition Match Score > 95%).
- Analisi della coerenza fotometrica, delle ombre, delle riflessioni corneali e della fusione delle giunzioni anatomiche.
- Quantificazione scientifica del livello di verosimiglianza audiovisiva espresso in scala percentuale oggettiva.

## APPROCCIO CONTESTUALE / CIRCOSTANZIALE

- Canale di trasmissione: un video inviato in una chat privata ha una credibilità intrinseca superiore rispetto a un forum palesemente satirico.
- Presenza di didascalie o messaggi di accompagnamento mirati a confermare la veridicità dell'accaduto.
- Il consulente tecnico deve fornire parametri oggettivi affinché il magistrato valuti la concreta idoneità decettiva della condotta.

# METODOLOGIE FORENSI DI DETECTION SU IMMAGINI E VIDEO

## VETTORI FISICI E SPETTRALI PER ACCERTARE LA NATURA SINTETICA DEL FILE

**1**

### **ANALISI DEL RUMORE DI SENSORE (PRNU)**

Ogni sensore CCD/CMOS reale imprime sui fotogrammi un'impronta siliconica fissa (Photo Response Non-Uniformity). Nei volti o nelle immagini interamente generate da AI il segnale PRNU risulta assente o incoerente.

**2**

### **ANALISI SPETTRALE DI FOURIER (FFT)**

Le reti neurali generative applicano filtri di upsampling che lasciano specifiche frequenze armoniche cicliche. L'esame nello spazio di Fourier evidenzia picchi geometrici invisibili all'occhio umano.

**3**

### **FOTOPLETISMOGRAFIA REMOTA (RPPG)**

Rilevazione microscopica dei cambi di colore cutaneo provocati dal flusso sanguigno cardiaco. Nei volti deepfake sintetizzati frame-per-frame il pattern cardiovascolare biologico è caotico o nullo.

# MATRICE TECNICA COMPARATIVA DEI PROTOCOLLI INVESTIGATIVI

AFFIDABILITÀ SCIENTIFICA, VETTORE ANALIZZATO E CRITICITÀ OPERATIVE IN SEDE DIBATTIMENTALE

| AMBITO D'INDAGINE   | TECNICA FORENSE              | VETTORE ANALIZZATO                            | CRITICITÀ OPERATIVA                 | ATTENDIBILITÀ      |
|---------------------|------------------------------|-----------------------------------------------|-------------------------------------|--------------------|
| VOICE CLONING       | Spettrogrammi MFCC / LFCC    | Discontinuità formanti e transitori sillabici | Compressione codec GSM/AMR/Opus     | ALTO (AUDIO RAW)   |
| VIDEO DEEPPFAKE     | Segnale Biometrico rPPG      | Micro-pulsazione del flusso sanguigno         | Illuminazione variabile e basso FPS | MOLTO ALTO (HD/4K) |
| IMMAGINI SINTETICHE | Analisi Spettrale FFT & PRNU | Residui di filtri neurali e sensore ottico    | Downscaling e filtri di rete social | MEDIO-ALTO         |
| TRACCIABILITÀ       | Standard Crittografico C2PA  | Metadati firmati PKI e hash SHA-256           | Stripping dei metadati di transito  | DETERMINISTICO     |

# LA CORSA AGLI ARMAMENTI: GENERAZIONE VS. DETECTION

## IL PARADOSSO DELL'EVOLUZIONE CONTINUA NELLE SCIENZE FORENSI DIGITALI

### IL DILEMMA STRUTTURALE DEL DISCRIMINATORE

- Meccanismo intrinseco delle GAN: ogni qualvolta viene pubblicato un nuovo metodo per rilevare deepfake, gli sviluppatori lo integrano nella loss function per azzerare l'artefatto.
- Asimmetria temporale: il perito forense deve analizzare ex post contenuti generati con modelli sempre più recenti e resistenti all'analisi.
- Inefficacia di regole fisse: nessun artefatto statico può essere considerato un indicatore definitivo ed eterno di manipolazione.

### DALLA CERTEZZA ASSOLUTA AL 'LIKELIHOOD RATIO'

- Superamento della logica binaria (100% Vero vs 100% Falso): la consulenza forense moderna deve quantificare il Rapporto di Verosimiglianza.
- Trasparenza del margine d'errore: la relazione peritale deve esplicitare i limiti strumentali in aderenza alla giurisprudenza "Cozzini".
- Il primato della metodologia umana: l'AI velocizza l'analisi preliminare ma l'interpretazione scientifica resta esclusivo compito del perito.

# CONCLUSIONI: IL PERITO FORENSE NELL'ERA DELL'AI

TRE IMPERATIVI INDISPENSABILI PER MAGISTRATI, AVVOCATI E CONSULENTI TECNICI

## 1. RIGORE METODOLOGICO IRRINUNCIABILE

Il rispetto scrupoloso delle norme ISO/IEC 27037 e ISO/IEC 27042 e la documentazione della Catena di Custodia rimangono l'unico scudo valido contro qualsiasi eccezione formale di inutilizzabilità.

## 2. RIFIUTO DELLA 'BLACK-BOX JUSTICE'

Nessuna perizia o consulenza tecnica può fondarsi su strumenti chiusi, opachi e non verificabili nel contraddittorio. Senza trasparenza e spiegabilità scientifica, la prova informatica deve cadere.

## 3. INTEGRAZIONE MULTIDISCIPLINARE

Il moderno professionista della digital forensics deve padroneggiare contestualmente DIVERSI pilastri inscindibili: per es. ingegneria dei dati/modelli, digital forensic, rito del processo civile e penale, etc.



**GRAZIE PER L'ATTENZIONE**