

DISPENSA DEL CORSO

Digital Forensics nell'Era dell'Intelligenza Artificiale

Traccia core e traccia specializzazioni

Vincenzo Calabrò

vincenzocalabro.it

Agosto 2026

Premessa

Questa dispensa accompagna il corso “Digital Forensics nell’Era dell’Intelligenza Artificiale” e ne raccoglie, in forma discorsiva, i contenuti teorici presentati a lezione. Le slide dei diciassette moduli restano lo strumento di riferimento per l’aula; qui lo studente trova il filo del ragionamento disteso in prosa, con i passaggi che una presentazione per sua natura comprime. I due materiali vanno letti insieme: le slide per la struttura e gli esempi discussi in classe, la dispensa per lo studio individuale e il ripasso.

L’impostazione del corso è duplice. Nella prima parte si costruisce il metodo della digital forensics classica: come si identifica, si acquisisce, si preserva e si analizza la prova digitale in modo che resista al vaglio di un procedimento giudiziario. Nella seconda parte lo stesso metodo viene messo alla prova dell’intelligenza artificiale, in una doppia prospettiva: l’AI come strumento al servizio dell’analista e l’AI come oggetto di indagine, quando il contenuto sintetico diventa esso stesso materia del contendere.

Un principio attraversa tutto il corso e conviene enunciarlo subito: l’AI seleziona, l’umano certifica. Gli strumenti automatici riducono il volume del materiale da esaminare, ordinano le priorità, suggeriscono piste; la responsabilità di ciò che entra nel fascicolo resta per intero in capo al consulente che firma la relazione. Ogni capitolo, in forme diverse, torna su questa divisione del lavoro e sulle sue conseguenze metodologiche e deontologiche.

Una nota sui materiali di esercitazione. Tutti i laboratori del corso impiegano esclusivamente dataset pubblici o sintetici, predisposti a scopo didattico (tra gli altri: NIST CFReDS, Digital Corpora, il corpus Enron). Nessuna attività didattica tratta dati personali reali: la regola vale per gli studenti anche fuori dal laboratorio, e anticipa un vincolo che il professionista ritroverà identico nella pratica.

Il corso è organizzato in due tracce. La traccia core, comune a tutti, si apre con il metodo forense e le basi dell’AI (moduli 1, 2, 3, 7 e 8) e si chiude con i profili giuridici ed etici e il laboratorio integrato (moduli 11 e 12). Fra le due fasi lo studente sceglie due dei cinque percorsi della traccia specializzazioni (S1-S5), ciascuno formato da un modulo base e da un modulo avanzato, trattato nei capitoli S1.2-S5.2. Il laboratorio integrato si svolge su un caso scelto fra quelli delle due specializzazioni seguite.

Capitolo 1. Fondamenti di digital forensics

La prova digitale

La digital forensics è la disciplina che applica il metodo scientifico all'identificazione, acquisizione, conservazione, analisi e presentazione del dato digitale a fini probatori. Il suo oggetto, la prova digitale, ha caratteristiche che la distinguono da ogni altra prova materiale. È latente: non è leggibile a occhio nudo, ma richiede strumenti e competenze per essere resa intelligibile. È fragile: un'accensione, un collegamento in rete, perfino la semplice apertura di un file possono alterarla in modo irreversibile. È duplicabile in modo perfetto: a differenza di un'impronta o di un reperto biologico, può essere copiata bit per bit, e la copia è indistinguibile dall'originale. È infine volatile in misura variabile: alcuni dati sopravvivono anni su un disco spento, altri svaniscono in millisecondi allo spegnimento della macchina.

Da queste caratteristiche discendono i due problemi fondativi della disciplina: garantire l'integrità del dato (che ciò che si analizza sia identico a ciò che si è acquisito) e documentarne la provenienza (che si possa sempre ricostruire chi ha fatto che cosa, quando e con quali strumenti).

Il processo forense

La prassi internazionale articola il lavoro in cinque fasi: identificazione, acquisizione, preservazione, analisi e presentazione. L'identificazione individua le potenziali fonti di prova sulla scena, fisica o logica che sia. L'acquisizione produce la copia su cui si lavorerà. La preservazione custodisce originale e copie garantendone l'integrità nel tempo. L'analisi estrae dai dati gli elementi pertinenti al quesito. La presentazione traduce i risultati in una relazione comprensibile a un destinatario non tecnico, tipicamente il giudice.

Le fasi non sono compartimenti stagni: ogni scelta compiuta a monte vincola ciò che sarà possibile a valle. Un'acquisizione condotta male non si corregge in analisi, e un'analisi corretta ma non documentata si dissolve in presentazione.

ISO/IEC 27037 e i ruoli

Lo standard ISO/IEC 27037 disciplina le prime fasi del processo e distingue due figure: il Digital Evidence First Responder (DEFR), chi interviene per primo sulla scena e compie le operazioni iniziali di identificazione e messa in sicurezza, e il Digital Evidence Specialist (DES), il tecnico qualificato per le operazioni specialistiche. Lo standard fissa quattro principi: auditability (ogni operazione deve essere documentata e verificabile), repeatability (a parità di condizioni e strumenti, il risultato deve potersi riprodurre), reproducibility (il risultato deve potersi ottenere anche con strumenti diversi) e justifiability (ogni scelta deve essere motivata).

Ripetibilità e codice di procedura penale

Nell'ordinamento italiano la distinzione tecnica tra operazioni ripetibili e irripetibili ha conseguenze processuali dirette. L'art. 359 c.p.p. disciplina gli accertamenti tecnici ripetibili del consulente del pubblico ministero. L'art. 360 c.p.p. si applica quando l'accertamento riguarda cose o luoghi il cui stato è soggetto a modificazione, e impone garanzie rafforzate: l'avviso all'indagato e al difensore, che possono partecipare con propri consulenti. Qualificare correttamente un'operazione come ripetibile o irripetibile è quindi una valutazione che il tecnico compie con effetti giuridici: un'acquisizione su fonte volatile, per esempio, è per natura irripetibile.

Legge 48/2008 e Convenzione di Budapest

La legge 18 marzo 2008, n. 48 ha ratificato la Convenzione di Budapest sul cybercrime e ha inserito nel codice di procedura penale i riferimenti espliciti alla prova digitale. Le norme su ispezioni (art. 244), perquisizioni (art. 247) e accertamenti urgenti (art. 354) impongono ora l'adozione di misure tecniche dirette ad assicurare la conservazione dei dati originali e ad impedirne l'alterazione, e la conformità delle copie all'originale. Il legislatore non prescrive una tecnica specifica, ma recepisce il principio: la prova digitale va trattata con metodo idoneo a garantirne integrità e genuinità.

La catena di custodia

La catena di custodia è il registro ininterrotto di ogni passaggio del reperto: chi lo ha preso in carico, quando, dove è stato custodito, quali operazioni ha subito, con quali strumenti e con quale esito di verifica. Una catena documentata consente di rispondere, anche a distanza di anni, alla domanda del controesame: come fate a sapere che questo è ancora il dato originale? Una catena interrotta non invalida automaticamente la prova, ma ne indebolisce il peso e apre lo spazio alle contestazioni.

Anti-forensics

Chi ha interesse a occultare agisce sulle stesse leve su cui lavora l'analista. Le tecniche principali: il wiping, la sovrascrittura deliberata dei dati; il timestomping, la manipolazione dei timestamp per falsare la cronologia; la cifratura, che rende il dato inaccessibile senza la chiave; la steganografia, che nasconde un contenuto dentro un altro; le tecniche fileless, che evitano di lasciare tracce su disco operando solo in memoria; la manomissione dei log. Conoscerle serve all'analista per riconoscerne i segni: quasi ogni tecnica di occultamento lascia a sua volta tracce caratteristiche, e l'assenza anomala di dati è essa stessa un dato.

Errori tipici

L'esperienza giudiziaria mostra errori ricorrenti: accendere la macchina sequestrata "per dare un'occhiata", lavorare sull'originale anziché su copia, omettere il calcolo dell'hash in acquisizione, documentare in modo generico strumenti e versioni, trarre in relazione conclusioni che i dati non sostengono. Ciascuno di questi errori torna, come caso di studio, nei capitoli che seguono.

Per il laboratorio. La prima esercitazione ricostruisce una catena di custodia completa su un reperto simulato, dalla presa in carico al verbale, e chiede di individuare gli errori inseriti ad arte in una

documentazione difettosa.

Lettere consigliate: ISO/IEC 27037:2012; E. Casey, *Digital Evidence and Computer Crime*, 3a ed., Academic Press; RFC 3227, *Guidelines for Evidence Collection and Archiving*.

Capitolo 2. L'acquisizione forense

La copia bit-a-bit

L'acquisizione è la fase in cui si produce la copia su cui si svolgerà tutto il lavoro successivo. La copia forense riproduce ogni settore del supporto, comprese le aree non allocate, lo slack space e i file cancellati, cioè tutto ciò che una copia logica dei soli file ignorerebbe. I formati in uso sono tre. Il formato RAW (dd) è il flusso grezzo dei byte: massima compatibilità, nessun metadato. Il formato E01 (Expert Witness Format) aggiunge compressione, hash integrati e metadati del caso, ed è lo standard di fatto nei laboratori. AFF4 è un formato aperto che supporta immagini sparse e acquisizioni parziali, utile sui volumi di grandi dimensioni.

Il write blocker

Durante l'acquisizione nessuna scrittura deve raggiungere la fonte. Il write blocker hardware è un dispositivo interposto tra reperto e workstation che intercetta i comandi di scrittura a livello di protocollo; la sua protezione non dipende dal sistema operativo né da configurazioni corrette, e per questo costituisce la prassi di laboratorio. Il blocco software (policy di sistema, mount in sola lettura nelle distribuzioni forensi) è più flessibile ed economico, ma la garanzia dipende dalla corretta configurazione, che va verificata e verbalizzata a ogni sessione. Senza protezione, un disco NTFS collegato a una macchina Windows viene toccato dal sistema in pochi secondi, e l'hash cambia prima ancora che un file venga aperto.

Hash e doppio hash

Le funzioni hash producono un'impronta a lunghezza fissa del contenuto: se la copia è identica all'originale, gli hash coincidono. MD5 presenta collisioni note e SHA-1 è deprecato; SHA-256 è lo standard attuale. La prassi del doppio hash (MD5 più SHA-256, calcolati in acquisizione e riportati a verbale) neutralizza in sede di controesame l'obiezione sulle collisioni: produrre due collisioni simultanee su funzioni diverse non è una possibilità concreta. Va chiarito che le collisioni MD5 richiedono una costruzione mirata e non rendono di per sé inattendibile una verifica di integrità; il doppio hash chiude comunque la discussione.

Il caso SSD

I dischi a stato solido hanno un comportamento che incide sulla ripetibilità. Il comando TRIM e la garbage collection interna cancellano fisicamente i blocchi marcati come liberi anche a disco inattivo, per cui due acquisizioni successive dello stesso SSD possono produrre hash diversi senza alcun intervento esterno. Il wear leveling distribuisce le scritture su celle diverse da quelle logicamente indirizzate, e l'area di over-provisioning resta inaccessibile all'acquisizione ordinaria. L'analista deve conoscere questi meccanismi per qualificare correttamente l'operazione e per spiegare in relazione un'eventuale discrepanza che nulla ha di anomalo.

Ordine di volatilità e live forensics

Quando la macchina è accesa si pone la scelta tra spegnere e acquisire a freddo oppure operare a caldo. La RFC 3227 ordina le fonti per volatilità decrescente: registri e cache, memoria centrale, stato della rete, processi, disco, supporti di backup. L'acquisizione live salva ciò che lo spegnimento distruggerebbe (memoria, connessioni attive, volumi cifrati montati), ma ogni operazione a caldo modifica il sistema: la scelta va motivata prima e documentata durante, comando per comando. È il caso tipico di accertamento irripetibile, con le garanzie processuali che ne conseguono.

RAID

Sui sistemi RAID il livello di acquisizione è una scelta tecnica con conseguenze. L'acquisizione fisica dei singoli dischi preserva tutto ma richiede di ricostruire in laboratorio la geometria dell'array (ordine dei dischi, dimensione dello stripe, parità); l'acquisizione logica del volume ricostruito è più semplice ma dipende dal controller e perde le aree che l'array non espone. Un errore classico riguarda il RAID 5 con un membro degradato: acquisire i dischi senza rilevare lo stato dell'array porta a ricostruzioni incoerenti.

Il verbale di acquisizione

Il verbale fissa ciò che la memoria non conserva: identificazione del reperto (marca, modello, numero di serie), contesto del prelievo, strumenti usati con versioni e firmware, formato della copia, orari di inizio e fine, hash calcolati e verificati, supporto di destinazione, sigilli, operatore. Un verbale completo consente a un altro tecnico di ripercorrere l'operazione; un verbale generico trasforma ogni domanda del controesame in un problema.

La conservazione

Dopo l'acquisizione il lavoro di custodia continua. Lo schema a tre livelli distingue l'originale (sigillato e toccato solo se indispensabile), la copia master (verificata e mai usata per lavorare) e le copie di lavoro, derivate dalla master per ogni analisi. Poiché un procedimento dura anni e i supporti invecchiano, la prassi prevede ridondanza su supporti distinti, ricalcolo periodico degli hash verbalizzato e migrazione programmata verso supporti nuovi, con hash di partenza e di arrivo. La conservazione è la coda lunga della catena di custodia.

Per il laboratorio. L'esercitazione prevede un'acquisizione completa con write blocker nei tre formati, la verifica degli hash e la stesura del verbale; una seconda prova confronta due acquisizioni successive dello stesso SSD e chiede di spiegare la discrepanza.

Lettere consigliate: NIST SP 800-86, *Guide to Integrating Forensic Techniques into Incident Response*; RFC 3227; B. Carrier, *File System Forensic Analysis*, Addison-Wesley (capitoli introduttivi sull'acquisizione).

Capitolo 3. File system e artefatti

NTFS e la Master File Table

Il file system è l'architettura che organizza i byte in file, e la sua conoscenza separa chi usa gli strumenti da chi li capisce. In NTFS ogni oggetto del volume è descritto da un record della Master File Table (\$MFT): nome, dimensioni, attributi, timestamp e, per i file piccoli, il contenuto stesso (file resident). Alla cancellazione il record viene marcato come libero ma non azzerato: finché non è riutilizzato, il file è ricostruibile per intero. I quattro timestamp di NTFS (creazione, modifica, accesso, modifica del record MFT) esistono inoltre in due copie, negli attributi \$STANDARD_INFORMATION e \$FILE_NAME: il secondo è assai più difficile da manipolare, e il confronto tra le due copie è il test classico contro il timestamping.

I journal: \$LogFile e USN Journal

NTFS tiene due registri delle operazioni. Il \$LogFile è il journal transazionale di basso livello: finestra temporale breve, dettaglio massimo. L'USN Journal registra a più alto livello le variazioni sui file (creazione, rinomina, cancellazione, modifica) con nome, momento e ragione, su una finestra di giorni o settimane: è la fonte di elezione per ricostruire la storia recente di un file, anche di file che non esistono più.

Volume Shadow Copies

Il servizio VSS di Windows fotografa periodicamente lo stato del volume con snapshot differenziali. Per l'analista significa poter leggere versioni precedenti di file modificati o cancellati, chiavi di registro comprese: una dimensione temporale verticale che si affianca alla timeline dei metadati. Gli snapshot si montano dalla copia forense con strumenti dedicati (libvshadow), senza toccare il reperto; lo spazio dedicato è limitato e gli snapshot più vecchi ruotano via, quindi la finestra coperta varia da giorni a mesi.

File carving

Quando le strutture del file system sono perse o deliberatamente distrutte, il carving ricerca i file direttamente nei byte grezzi, riconoscendone le firme di intestazione e coda (header e footer). Recupera contenuti dalle aree non allocate ma perde i nomi, i percorsi e i timestamp, e soffre la frammentazione. È la tecnica di ultima istanza, e insieme il banco di prova dei classificatori statistici che si vedranno nella seconda parte del corso.

Registry e tracce di interazione

Il registro di Windows è un database gerarchico che conserva, oltre alla configurazione, una quantità di tracce d'uso: dispositivi USB collegati (con numero di serie e date), reti note, programmi eseguiti,

documenti recenti. Tre famiglie di artefatti documentano in particolare l'interazione dell'utente. Gli shellbag registrano vista e posizione di ogni cartella aperta in Explorer, comprese cartelle su supporti rimovibili scollegati o cancellate. I file LNK, generati a ogni apertura di documento, conservano percorso originale, timestamp, volume e numero di serie del supporto, e sopravvivono spesso al file cui puntavano. Le Jump List collegano applicazione, file aperto e momento dell'apertura. Il valore probatorio nasce dalla convergenza: un singolo artefatto si contesta, tre coerenti tra loro molto meno.

Artefatti dei browser

Chromium e Firefox conservano cronologia, download, cookie e dati dei form in database SQLite. La cronologia registra anche i termini di ricerca, che in giudizio documentano l'intenzione oltre che l'azione. I timestamp usano epoche diverse tra loro (WebKit, Unix, PRTIME) e la conversione va verbalizzata. La navigazione in incognito non scrive questi database ma lascia tracce in memoria e nella cache DNS; la sincronizzazione cloud dei profili estende l'analisi oltre il singolo dispositivo, con le cautele giuridiche trattate nel capitolo 11.

Timeline analysis

La timeline è la disposizione cronologica di tutti gli eventi ricavabili dagli artefatti. La super-timeline, costruita con strumenti come Plaso e consultata in Timesketch, fonde in un'unica sequenza metadati del file system, journal, registry, log ed eventi applicativi: decine di sorgenti, milioni di righe. La sua forza è la correlazione (ciò che accade nello stesso minuto su sorgenti indipendenti); il suo rischio è la quantità, che richiede ipotesi di lavoro per essere interrogata. I fusi orari e la precisione delle diverse sorgenti vanno normalizzati esplicitamente, e ogni evento citato in relazione deve restare riconducibile all'artefatto di origine.

Per il laboratorio. Su un'immagine didattica gli studenti ricostruiscono la storia di un documento cancellato combinando MFT, USN Journal e una shadow copy, e verificano un sospetto di timestamping confrontando gli attributi \$SI e \$FN; la timeline completa viene poi generata con Plaso ed esplorata in Timesketch.

Lecture consigliate: B. Carrier, *File System Forensic Analysis*, Addison-Wesley; documentazione di Plaso e Timesketch; H. Carvey, *Windows Registry Forensics*, Syngress.

Capitolo 4. Memory forensics

Perché la memoria

La RAM contiene ciò che il disco non vede: processi in esecuzione, connessioni di rete attive, chiavi di cifratura, contenuti in chiaro di volumi cifrati, codice malevolo che non ha mai toccato il disco. L'analisi della memoria è quindi complementare a quella del file system, e in molti incidenti moderni è la sola via per ricostruire che cosa stava accadendo.

Acquisizione della RAM

L'acquisizione avviene su sistema acceso, con strumenti che leggono la memoria fisica e la riversano su un supporto esterno. È un'operazione per definizione irripetibile e invasiva: lo strumento stesso occupa memoria e modifica lo stato del sistema, per cui la scelta va motivata prima e documentata durante, in coerenza con quanto visto sulla live forensics. Il dump va accompagnato da hash e da annotazioni su strumento, versione e orari, come ogni acquisizione.

La memoria che resta sul disco

Una parte della memoria sopravvive allo spegnimento, riversata su disco dal sistema operativo. Il pagefile.sys contiene le pagine scaricate dal gestore della memoria virtuale: frammenti di processi anche terminati da tempo, privi però di struttura, da analizzare per ricerca di pattern. Lo hiberfil.sys è un'istantanea compressa dell'intera RAM al momento dell'ibernazione, convertibile in un formato analizzabile con gli strumenti ordinari. Nelle versioni recenti di Windows si aggiungono swapfile.sys e la compressione della memoria. La conseguenza pratica: anche da una macchina sequestrata spenta si recupera una parte dello stato volatile.

Strutture e analisi con Volatility

Il dump grezzo diventa leggibile interpretando le strutture dati del kernel: elenco dei processi, tabelle degli handle, moduli caricati, connessioni. Volatility 3 è lo strumento di riferimento: a partire dai simboli del sistema operativo analizzato, i suoi plugin ricostruiscono processi (pslist, pstree), righe di comando (cmdline), librerie (dlllist), file aperti (handles) e connessioni (netscan).

Cross-view analysis

Un principio metodologico attraversa l'analisi della memoria: confrontare viste diverse dello stesso oggetto. L'elenco dei processi ottenuto percorrendo le liste del kernel (pslist) può essere confrontato con la scansione diretta delle strutture in memoria (psscan): un processo presente nella seconda vista e assente nella prima è stato scollegato dalle liste, comportamento tipico dei rootkit. La discrepanza tra viste è il segnale; nessuna vista, da sola, è fededegna.

Dal dump al processo

L'analisi tipica procede in tre passi. Prima si individua il processo sospetto: nomi quasi legittimi, processi padre anomali, orari incoerenti. Poi lo si contestualizza: con quali argomenti è partito, quali librerie e file teneva aperti, con chi comunicava. Infine lo si estrae: l'eseguibile e la memoria del processo diventano file con hash propri, analizzabili staticamente o in sandbox. Ogni estrazione è una derivazione da documentare, con comandi, versioni e hash, in modo che un altro consulente possa ripercorrere il lavoro.

Credenziali in memoria

La memoria conserva spesso credenziali e chiavi: hash delle password, ticket di autenticazione, chiavi dei volumi cifrati caricate dai driver. Il recupero della chiave di un volume cifrato da un hiberfil può aprire, in modo ripetibile e documentato, un contenitore altrimenti inaccessibile. L'uso di queste possibilità incontra i limiti giuridici discussi nel capitolo 11.

Linux, cloud e container

Gli incidenti aziendali avvengono in larga parte su server Linux. L'acquisizione usa moduli kernel come LiME o strumenti come AVML; il modulo va compilato per il kernel esatto della macchina e il suo caricamento modifica lo stato del sistema, scelta da motivare come ogni operazione live. L'analisi con Volatility 3 richiede una symbol table coerente con il kernel. Nei container i processi appaiono nel kernel dell'host, per cui un solo dump copre tutti i container in esecuzione. Nel cloud lo snapshot della macchina virtuale offerto dal provider è spesso la via più pulita, con una catena di custodia da documentare insieme al provider.

Malware residente e fileless

Le minacce fileless operano interamente in memoria: iniezione in processi legittimi, codice caricato da script, persistenza affidata a registry o attività pianificate anziché a file su disco. Per queste minacce la memoria non è un complemento dell'analisi del disco: è la fonte primaria. I plugin dedicati cercano regioni di memoria eseguibili non riconducibili a moduli legittimi e incroci anomali tra processi e privilegi.

Per il laboratorio. Su un dump didattico gli studenti individuano un processo malevolo con la cross-view analysis, ne ricostruiscono il contesto, lo estraggono e ne verificano l'hash contro i repository di riferimento, documentando ogni passo come in una relazione reale.

Lecture consigliate: M. H. Ligh, A. Case, J. Levy, A. Walters, *The Art of Memory Forensics*, Wiley; documentazione ufficiale di Volatility 3; NIST SP 800-86 per l'inquadramento dell'acquisizione live.

Capitolo 5. Network e cloud forensics

Il traffico come fonte di prova

La rete registra le relazioni: chi ha parlato con chi, quando, quanto. La cattura completa del traffico (PCAP) conserva ogni pacchetto ed è la fonte più ricca, ma il suo costo in spazio la rende praticabile solo su finestre brevi o segmenti selezionati. L'analisi forense del traffico lavora quindi, nella pratica, su una gerarchia di fonti a dettaglio decrescente e copertura crescente: pacchetti, metadati arricchiti per protocollo (Zeek), flussi, log.

TLS e il valore dei metadati

La cifratura del trasporto ha cambiato il mestiere: il contenuto delle sessioni TLS non è leggibile, ma i metadati restano eloquenti. L'handshake espone il nome del server richiesto, i certificati, le suite crittografiche; dimensioni, tempi e ritmo dei pacchetti disegnano il profilo della comunicazione. Molte conclusioni investigative (quale servizio, quanto a lungo, con che volumi) non richiedono il contenuto.

DNS e beaconing

Le interrogazioni DNS tracciano le intenzioni di ogni processo che risolve un nome, e la loro analisi rivela domini generati algebricamente, tunneling e contatti con infrastrutture note. Il beaconing, il contatto periodico di un host compromesso con il suo centro di comando, si tradisce per la regolarità: intervalli quasi costanti, payload di dimensioni ripetitive, orari indifferenti all'attività umana. È un pattern che emerge dalle statistiche dei metadati, non dal contenuto.

NetFlow e IPFIX

Dove la cattura completa è impraticabile, i flussi NetFlow/IPFIX registrano i soli metadati di ogni conversazione: indirizzi, porte, protocollo, byte, pacchetti, orari. Il risultato è una memoria storica di mesi dell'intera rete a costo contenuto, che risponde alle domande strategiche: quando è iniziato il contatto con l'infrastruttura ostile, quali host interni l'hanno raggiunta, quanti dati sono usciti. Il PCAP, quando esiste, si usa poi in modo chirurgico sugli intervalli che i flussi hanno indicato. I metadati di flusso pongono inoltre meno problemi di proporzionalità rispetto alla cattura del contenuto.

Log e correlazione

I log di sistemi, apparati e applicazioni sono la fonte più diffusa e la più eterogenea. Il lavoro forense consiste nel normalizzarli (formati, fusi orari, identificatori) e correlarli: lo stesso evento osservato da sorgenti indipendenti acquista un peso che nessuna sorgente da sola possiede. Le cautele ricorrenti: i log si manomettono, la loro conservazione ha finestre brevi che impongono richieste

tempestive di congelamento, e un'assenza nei log prova poco se non si conosce la politica di registrazione.

Email forensics

L'email resta il vettore principale di frodi e compromissioni. Gli header Received, aggiunti da ogni server attraversato e letti dal basso verso l'alto, ricostruiscono il percorso del messaggio; il campo From è una semplice dichiarazione, falsificabile da chiunque. Tre controlli di autenticazione aiutano a valutare la provenienza: SPF verifica che l'indirizzo IP mittente sia autorizzato dal dominio, DKIM verifica la firma crittografica del dominio e l'integrità del contenuto firmato, DMARC stabilisce la politica da applicare quando i primi due falliscono e la coerenza con il From visibile. Gli esiti si leggono negli header Authentication-Results del messaggio originale, non di un inoltrato. Va ricordato il limite logico: un DKIM valido prova che il messaggio proviene dal dominio firmatario, non che il suo contenuto sia veritiero, e le frodi con domini somiglianti superano i controlli proprio perché autenticano il dominio sbagliato.

Log di proxy e web server

Il log del proxy aziendale è la controparte server-side della cronologia del browser: utente autenticato, URL, volumi, esito. Sopravvive alla pulizia del client e copre tutti i dispositivi della rete. Sul versante opposto, access log ed error log dei web server documentano chi ha visitato un sito e come, e nelle indagini sugli attacchi web mostrano i pattern tipici (scansioni, path traversal, injection nei parametri). Le cautele di lettura: fusi orari del server, NAT che nasconde più utenti dietro un indirizzo, user-agent falsificabili.

Cloud forensics

Nel cloud la prova vive presso un terzo. L'acquisizione passa per le API e gli snapshot del provider, per gli audit log dei servizi e, sul piano giuridico, per richieste formali i cui tempi vanno confrontati con le finestre di conservazione. La catena di custodia include il provider: chi ha estratto che cosa, con quali strumenti e garanzie va documentato insieme ai dati stessi.

Per il laboratorio. Su una cattura didattica gli studenti individuano un beacon con Zeek, quantificano un'esfiltrazione dai flussi e analizzano gli header di una email di frode, verificando SPF, DKIM e DMARC sul messaggio originale.

Lettere consigliate: documentazione di Zeek; RFC 7011 (IPFIX); RFC 7208, 6376 e 7489 (SPF, DKIM, DMARC); NIST SP 800-86.

Capitolo 6. Mobile forensics

Un oggetto diverso

Lo smartphone è il dispositivo più personale che esista: accompagna il suo utilizzatore ovunque, registra comunicazioni, spostamenti, abitudini e salute, ed è progettato per sincronizzarsi di continuo con il cloud. Per l'analista è insieme la fonte più ricca e la più difficile: cifratura di serie, sistemi chiusi, aggiornamenti che cambiano le regole a ogni versione.

Isolamento

La prima operazione sul dispositivo sequestrato è isolarlo dalla rete: una cancellazione remota può distruggere tutto in pochi secondi. Gli strumenti sono la modalità aereo documentata, le buste e le gabbie di Faraday, la disattivazione controllata delle connessioni. L'isolamento ha un costo che va messo a verbale: il dispositivo isolato non riceve più gli eventi che continuerebbero ad arrivare, e il suo stato si congela a quel momento.

Tecniche di estrazione

Le estrazioni si ordinano per profondità crescente e invasività crescente. L'estrazione logica recupera ciò che il sistema espone attraverso le interfacce di backup e sincronizzazione. L'estrazione del file system accede alla struttura completa delle partizioni utente. L'estrazione fisica riproduce la memoria a basso livello e oggi, sui dispositivi cifrati moderni, coincide spesso con tecniche avanzate che sfruttano stati particolari del dispositivo o vulnerabilità. La scelta dipende dal modello, dalla versione del sistema, dallo stato (acceso o spento, sbloccato almeno una volta dopo l'accensione oppure no) e va motivata rispetto al quesito.

La via laterale dei backup

Quando il dispositivo è irraggiungibile, bloccato o distrutto, i suoi dati spesso esistono altrove. Il backup locale su un computer sequestrato contiene messaggi, foto e database delle app; se protetto da password nota o recuperabile è anzi più ricco di quello in chiaro, perché include il keychain. I backup cloud (iCloud, Google One) conservano copie dei dati principali, raggiungibili per richiesta al provider o tramite token di autenticazione trovati sugli altri dispositivi, strade giuridicamente diverse da percorrere con i titoli adeguati. Il backup è una fotografia alla sua data: la differenza tra backup e dispositivo è essa stessa informazione su che cosa è stato cancellato, e quando.

SQLite, il formato della mobile forensics

La quasi totalità dei dati applicativi mobili vive in database SQLite. Oltre alle tabelle, il formato conserva spesso record cancellati non ancora sovrascritti nelle pagine libere e nel write-ahead log (WAL), che può contenere transazioni non ancora consolidate. Saper interrogare SQLite

direttamente, senza affidarsi solo al parser commerciale, è la competenza che distingue l'analista in grado di verificare i propri strumenti.

Le app di messaggistica

Tre architetture producono tre esiti forensi. WhatsApp cifra il trasporto, ma i messaggi riposano in chiaro nel database locale e spesso nei backup: con accesso al dispositivo il contenuto è di norma recuperabile. Telegram tiene le chat ordinarie sul cloud del provider: dal dispositivo si recupera la cache, il resto dipende dal gestore, e le chat segrete sono solo locali. Signal riduce al minimo le tracce: database locale cifrato con chiave nel keystore, niente backup cloud, metadati minimi anche lato server. La relazione deve spiegare perché gli esiti divergono: è l'architettura dell'app a determinare il perimetro del recuperabile.

Geolocalizzazione

Le posizioni si ricostruiscono da più sorgenti con precisioni diverse: database di localizzazione del sistema, metadati EXIF delle fotografie, dati delle app di navigazione e fitness, celle agganciate e reti Wi-Fi viste. Ogni sorgente ha errori e semantiche propri (posizione del dispositivo, non della persona; momento della registrazione, non necessariamente della presenza) che vanno dichiarati.

Attribuzione: il dispositivo non è la persona

Ogni artefatto mobile documenta un'azione del dispositivo, non l'identità di chi lo teneva in mano. Il passaggio dal dispositivo alla persona è un'inferenza da costruire con riscontri: biometria e codici di sblocco, coerenza con celle e badge, contenuti riconducibili all'interessato, testimonianze. La relazione che confonde i due piani si espone al controesame più ovvio.

Wearable e IoT personale

Orologi e braccialetti registrano battito, passi, sonno e allenamenti georeferenziati; gli assistenti vocali conservano la cronologia dei comandi nel cloud dell'account; serrature e termostati intelligenti registrano presenze. I dati vivono su tre livelli (dispositivo, app companion, cloud) e il livello intermedio è spesso il più accessibile. Il loro valore è tipicamente di riscontro: confermare o smentire un alibi, collocare un orario. Vale il limite di attribuzione appena visto: il sensore misura un corpo, non un'identità.

Cifratura e accesso

La cifratura di serie protegge i dati a dispositivo spento; gran parte delle possibilità di estrazione dipende dallo stato del dispositivo al momento del sequestro. Le richieste di codici e chiavi all'interessato, e l'uso di strumenti di sblocco, incontrano limiti giuridici diversi nei diversi ordinamenti: il tema è ripreso nel capitolo 11.

Per il laboratorio. Gli studenti analizzano con ALEAPP un'estrazione Android didattica, ricostruiscono una conversazione da un database SQLite recuperando record dal WAL e confrontano l'esito dell'estrazione con il backup disponibile, motivando le discrepanze.

Lecture consigliate: documentazione di ALEAPP e iLEAPP; E. Casey, *Digital Evidence and Computer Crime*, 3a ed. (parte mobile); documentazione SQLite sul formato di file e sul WAL.

Capitolo 7. Fondamenti di machine learning per il forenser

Questo capitolo fornisce le basi concettuali necessarie per valutare criticamente gli strumenti di intelligenza artificiale impiegati nelle indagini digitali. L'obiettivo non è formare data scientist, ma periti in grado di comprendere che cosa un modello sa fare, dove sbaglia e come se ne rende conto.

Paradigmi di apprendimento

Si distinguono tre famiglie principali. Nell'apprendimento supervisionato il modello apprende da esempi etichettati: coppie di input e risposta corretta. Nell'apprendimento non supervisionato il modello cerca struttura in dati privi di etichette: gruppi, regolarità, eccezioni. Nell'apprendimento per rinforzo un agente impara per tentativi, guidato da una ricompensa. In ambito forense dominano i primi due: la classificazione di file e la rilevazione di anomalie.

Dati e distribuzioni

Un modello conosce soltanto il territorio rappresentato dai suoi dati di addestramento. Se la distribuzione dei dati reali differisce da quella del training set, le prestazioni dichiarate non sono più garantite. Questa osservazione, apparentemente banale, è la radice della maggior parte delle contestazioni fondate: chiedere sempre su quali dati, di quale epoca e di quale provenienza, lo strumento è stato addestrato e validato. Nel corso si usano esclusivamente dataset pubblici o sintetici, e lo stesso criterio di tracciabilità va preteso dagli strumenti commerciali.

Feature engineering per artefatti forensi

Un modello non vede file o pacchetti: vede vettori numerici. La traduzione dal reperto al vettore, detta feature engineering, è il punto in cui la competenza forense entra nel sistema. Da un eseguibile si estraggono entropia delle sezioni, API importate, dimensioni; da una sessione di rete durate, byte scambiati, periodicità; da un messaggio lunghezze, lessico, orari. La scelta ha conseguenze probatorie: feature interpretabili rendono il modello spiegabile in aula, mentre rappresentazioni apprese automaticamente spostano la spiegazione su un piano più difficile da verbalizzare. Vale una regola pratica: se una feature non si riesce a spiegare a un giudice in una frase, il suo contributo diventerà un punto debole in controesame.

Overfitting e generalizzazione

Un modello troppo adattato ai dati di addestramento memorizza invece di apprendere: brilla sul training set e fallisce sul mondo reale. La difesa è metodologica: separare i dati in insiemi di addestramento, validazione e test, e misurare le prestazioni soltanto su dati mai visti dal modello.

Una relazione che cita metriche calcolate sugli stessi dati di addestramento riporta un numero privo di valore.

Metriche ed errori

Accuratezza, precision e recall misurano aspetti diversi del comportamento di un classificatore, e la matrice di confusione li riassume: veri positivi, falsi positivi, veri negativi, falsi negativi. In ambito investigativo i due errori hanno costi asimmetrici: il falso positivo consuma tempo di analista, il falso negativo lascia fuori una prova. Per questo la metrica va sempre letta rispetto al quesito: un triage orientato alla completezza privilegia il recall e accetta più falsi allarmi, documentandolo.

Clustering e riduzione dimensionale

Quando mancano le etichette si ricorre all'apprendimento non supervisionato. K-means partiziona i dati in k gruppi attorno a centroidi: veloce e leggibile, ma costringe ogni punto in un cluster, rumore compreso. DBSCAN e HDBSCAN raggruppano per densità, trovano forme arbitrarie e lasciano fuori gli outlier, che in chiave forense sono spesso proprio il risultato da esaminare. PCA e UMAP proiettano spazi a centinaia di dimensioni su due o tre assi visualizzabili: strumenti preziosi per esplorare e comunicare, con l'avvertenza che le distanze nella proiezione distorcono quelle reali e non autorizzano conclusioni probatorie. In tutti i casi il non supervisionato organizza e prioritizza il lavoro umano, non lo sostituisce.

Spiegabilità

Le tecniche di explainable AI, fra cui SHAP e LIME, attribuiscono a ciascuna feature il contributo dato alla singola decisione del modello. Permettono di rispondere alla domanda cruciale del contraddittorio: perché il sistema ha segnalato proprio questo elemento? La spiegazione prodotta è a sua volta un'approssimazione, con limiti noti in letteratura, e va presentata come tale.

Calibrazione delle probabilità

Molti modelli emettono uno score fra 0 e 1 che non è una probabilità: le reti neurali moderne tendono alla sovraconfidenza, e uno 0,92 può corrispondere a un tasso di correttezza reale molto inferiore. Un modello si dice calibrato quando, fra i casi cui assegna 0,9, circa il 90 per cento risulta davvero positivo. La calibrazione si misura con il reliability diagram e con l'Expected Calibration Error, e si corregge con tecniche come Platt scaling, regressione isotonica e temperature scaling, sempre su dati di validazione separati. In sede peritale il tema è decisivo: il numero riportato in relazione verrà letto come una probabilità, e deve quindi comportarsi come tale, oppure essere sostituito da fasce qualitative con frequenze d'errore dichiarate.

LLM e transformer in sintesi

I modelli linguistici di grandi dimensioni sono reti basate sull'architettura transformer, addestrate a predire testo su corpora molto ampi. Producono linguaggio fluente e sanno riassumere, tradurre,

estrarre informazioni; possono però generare contenuti plausibili ma falsi. Per l'uso investigativo valgono due regole: ogni output va verificato sulla fonte, e il loro ruolo legittimo è assistere la selezione e la redazione, mai certificare un fatto.

Per il laboratorio. Gli studenti addestrano un semplice classificatore su un dataset pubblico, ne misurano precision, recall e calibrazione su un test set separato e redigono una pagina di documentazione delle feature utilizzate, scritta in modo comprensibile a un lettore non tecnico.

Lettere consigliate: C. M. Bishop, “Pattern Recognition and Machine Learning”, Springer; A. Géron, “Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow”, O'Reilly; C. Molnar, “Interpretable Machine Learning” (disponibile online).

Capitolo 8. AI applicata I: triage e classificazione

I volumi dei sequestri moderni superano da anni la capacità di esame manuale. Il triage automatizzato risponde a questo problema: ordinare e filtrare i reperti perché l'attenzione dell'analista si concentri dove serve. Il capitolo presenta le tecniche principali e, con pari rilievo, gli errori di ragionamento che il loro uso introduce.

Ridurre il noto: gli hash set

Il primo filtro non richiede alcuna intelligenza artificiale. La National Software Reference Library del NIST pubblica insiemi di hash di file noti e innocui, tipicamente componenti di sistemi operativi e applicazioni commerciali: confrontando gli hash dei file del reperto con questi insiemi si esclude in blocco ciò che non merita esame. Tecniche analoghe, con insiemi di hash di materiale illecito noto, permettono di identificare contenuti già catalogati senza visione umana.

Classificare per contenuto

Rinominare un file è il più elementare degli occultamenti. La prima verifica è deterministica: i magic bytes, la firma d'intestazione del formato, confrontati con l'estensione dichiarata; la discrepanza è di per sé un segnale di triage. Il secondo livello è statistico: entropia e distribuzione dei byte distinguono testo, immagini compresse, archivi e contenuto cifrato anche su frammenti privi di intestazione, come i settori recuperati dal carving. Su questi frammenti i classificatori di machine learning superano le firme, perché apprendono il profilo statistico di ciascun formato. Resta un limite da dichiarare: contenuto cifrato e contenuto compresso convergono entrambi verso l'entropia massima, e la distinzione fine richiede test ulteriori e prudenza lessicale in relazione.

Hash percettivi

Gli hash crittografici cambiano completamente al variare di un bit, e dunque non riconoscono un'immagine ricompressa o ridimensionata. Gli hash percettivi, fra cui PhotoDNA e PDQ, sintetizzano invece l'aspetto visivo del contenuto: immagini simili producono hash vicini. Sono lo strumento standard per riconoscere varianti di materiale noto. La soglia di somiglianza è una scelta dell'analista e va motivata, perché determina il compromesso fra varianti mancate e falsi accostamenti.

La tutela dell'operatore

Nei procedimenti che coinvolgono materiale traumatico, l'esposizione quotidiana degli analisti ha costi documentati in letteratura, dallo stress traumatico secondario al burnout. L'automazione ha qui una funzione protettiva: il materiale già catalogato viene identificato senza visione umana, il nuovo viene prioritizzato e mostrato in condizioni controllate. Le misure organizzative completano il quadro: rotazione degli incarichi, limiti di esposizione, supporto psicologico strutturato. Resta fermo

che ogni elemento destinato al fascicolo richiede verifica umana: l'obiettivo è che ogni sguardo sia necessario, non che non ve ne siano.

Anomaly detection e fallacia del base rate

La rilevazione di anomalie segnala ciò che si discosta da un comportamento atteso: utile quando non si sa in anticipo che cosa cercare. Il suo uso richiede però di fare i conti con la fallacia del base rate. Un rilevatore accurato al 99 per cento, applicato a un fenomeno che riguarda un caso su diecimila, produce quasi soltanto falsi allarmi: su un milione di eventi segnala circa diecimila casi, dei quali appena un centinaio reali. Il valore predittivo di una segnalazione dipende dalla prevalenza del fenomeno, non solo dall'accuratezza dello strumento. Ogni relazione che riporti percentuali di accuratezza senza considerare la prevalenza presta il fianco a una contestazione fondata.

Soglie: una decisione da motivare

Ogni classificatore richiede una soglia operativa che traduce lo score in decisione. Abbassarla aumenta la completezza e i falsi allarmi; alzarla riduce il carico di verifica e aumenta il rischio di prove mancate. Non esiste la soglia giusta in astratto: esiste la soglia adeguata al quesito, e la sua scelta va documentata in relazione insieme alle metriche stimate e al trattamento riservato a ciò che resta sotto soglia.

Concept drift e manutenzione

Un classificatore apprende il mondo com'era nei suoi dati di addestramento, ma il mondo digitale cambia: nuovi formati, nuove applicazioni, nuove tecniche di occultamento. Le prestazioni decadono silenziosamente, mentre il modello continua a emettere score con la stessa apparente sicurezza. Si distingue il data drift, in cui cambia la distribuzione degli input, dal concept drift proprio, in cui cambia la relazione fra input ed etichetta. Le contromisure sono organizzative: monitoraggio delle distribuzioni, set di controllo periodici etichettati, riaddestramento programmato. Ne discende un'implicazione peritale precisa: le metriche hanno una data, e citarle senza indicarla rende l'affermazione incompleta.

Il protocollo di validazione umana

Il corso adotta un protocollo in tre passi per ogni impiego di triage automatico. Primo: campionare in modo casuale sia il materiale segnalato sia quello scartato. Secondo: esaminare il campione con il metodo manuale, in cieco rispetto all'esito del modello. Terzo: misurare la concordanza e riportarla in relazione insieme alle soglie e alle versioni degli strumenti. La macchina seleziona e ordina; la persona verifica e risponde di ciò che firma.

Per il laboratorio. Gli studenti applicano una pipeline di triage a un'immagine disco didattica, scelgono e motivano una soglia, eseguono il protocollo di validazione su campione e discutono i falsi positivi e negativi riscontrati.

Lecture consigliate: NIST, National Software Reference Library, documentazione ufficiale; J. Davis e M. Goadrich, “The Relationship Between Precision-Recall and ROC Curves”, ICML 2006; D. Quick e K.-K. R. Choo, “Impacts of increasing volume of digital forensic data”, Digital Investigation.

Capitolo 9. AI applicata II: NLP e analisi investigativa

Le indagini moderne producono testo in quantità: chat, email, documenti, verbali. Il natural language processing offre strumenti per interrogare questi corpora, e l'analisi dei grafi relazionali per ricostruire chi comunica con chi. Il capitolo presenta entrambi, con l'avvertenza che attraversa tutto il corso: queste tecniche selezionano e organizzano ciò che una persona leggerà e certificherà.

NLP su corpora investigativi

Il testo investigativo reale è rumoroso: abbreviazioni, errori, gergo, ironia. Le pipeline di NLP vanno quindi valutate sul materiale del caso, non sulle prestazioni dichiarate su corpora puliti. La prima operazione è di norma la normalizzazione e la deduplicazione; seguono le analisi di estrazione e organizzazione.

Named entity recognition

La NER individua nel testo nomi di persona, luoghi, organizzazioni, date, importi. Applicata a decine di migliaia di messaggi, produce l'inventario delle entità del caso e le loro co-occorrenze. Gli errori tipici riguardano soprannomi, nomi ambigui e lingue miste; per questo l'inventario è un indice di lavoro, da verificare prima di ogni uso probatorio.

Topic modeling

Le tecniche di topic modeling raggruppano i documenti per argomento senza etichette predefinite. Servono a farsi un'idea della struttura tematica di un corpus sconosciuto e a individuare i filoni da approfondire. I temi prodotti sono costrutti statistici privi di significato garantito: la loro interpretazione spetta all'analista.

Embedding semantici

Un embedding traduce un testo in un vettore in uno spazio a centinaia di dimensioni, costruito in modo che testi di significato simile risultino vicini. La conseguenza pratica è la ricerca semantica: trovare i messaggi pertinenti anche quando il corpus usa parole del tutto diverse da quelle della query. Per l'investigatore cambia la natura della ricerca, dalle parole chiave, che i soggetti evitano con codici e gerghi, alla somiglianza di significato. I limiti vanno dichiarati: la soglia di vicinanza è una scelta dell'analista, i modelli risentono di lingua e dominio di addestramento, e la vicinanza semantica è un criterio di selezione, non una prova.

Corpora multilingue

I corpora reali mescolano lingue, dialetti e traslitterazioni, spesso nella stessa frase. Le pipeline moderne identificano la lingua per segmento, traducono per consentire il triage e indicizzano con

embedding multilingue, che proiettano lingue diverse nello stesso spazio semantico: una ricerca in italiano recupera messaggi pertinenti in altre lingue. La qualità varia molto fra lingue ad alte e basse risorse e va dichiarata. Sul piano probatorio la regola è netta: la traduzione automatica è uno strumento di lavoro, e ciò che entra nel fascicolo richiede un traduttore o interprete qualificato, perché gergo, ironia e riferimenti culturali sono proprio i punti in cui la macchina sbaglia.

Grafi relazionali e centralità

Dai metadati delle comunicazioni si costruisce un grafo: nodi per i soggetti, archi per i contatti. Le misure di centralità ne leggono la struttura: il degree conta i contatti diretti, la betweenness individua gli intermediari che collegano gruppi altrimenti separati, l'eigenvector centrality pesa i legami con nodi a loro volta centrali. Queste misure generano ipotesi investigative sul ruolo dei soggetti; non le provano. Un'alta betweenness può indicare un intermediario criminale come un centralinista.

Stilometria e authorship attribution

La stilometria misura le regolarità inconsce della scrittura: frequenze delle parole funzionali, lunghezze di frase, punteggiatura, errori ricorrenti. Su questi profili si confronta un testo contestato con campioni noti dei candidati autori. I limiti sono parte integrante del metodo: servono campioni ampi e dello stesso registro; il risultato è una misura di compatibilità fra ipotesi alternative, mai un'identificazione; lo stile cambia con contesto e tempo; l'imitazione deliberata, oggi assistita da modelli linguistici, è un'ipotesi da escludere esplicitamente. La formulazione corretta in relazione è probabilistica, rispetto a ipotesi dichiarate, con il tasso d'errore del metodo su casi noti.

LLM nel workflow e RAG

I modelli linguistici assistono l'analista nel riassumere, confrontare versioni, preparare domande, redigere bozze. Il rischio specifico è la generazione di contenuti plausibili ma falsi, citazioni comprese. L'architettura retrieval-augmented generation riduce il rischio ancorando le risposte a documenti recuperati dal fascicolo, che il sistema cita: ogni affermazione resta così verificabile sulla fonte. La regola operativa non cambia: nessun output di un modello linguistico entra in una relazione senza verifica puntuale della persona che firma, e l'uso dello strumento va dichiarato.

Per il laboratorio. Su un corpus email pubblico gli studenti costruiscono l'inventario delle entità, il grafo relazionale con le misure di centralità e una ricerca semantica su un tema assegnato, quindi verificano a mano un campione dei risultati e documentano gli errori trovati.

Lettere consigliate: D. Jurafsky e J. H. Martin, "Speech and Language Processing", 3a ed. (bozza online); M. E. J. Newman, "Networks", Oxford University Press; P. Juola, "Authorship Attribution", Foundations and Trends in Information Retrieval.

Capitolo 10. Forensics dell'AI: contenuti sintetici e attacchi ai sistemi

Nei capitoli precedenti l'intelligenza artificiale era uno strumento dell'analista. In questo capitolo diventa l'oggetto dell'indagine: contenuti generati o manipolati da modelli, e sistemi di AI attaccati o ingannati. È il terreno in cui il metodo forense incontra la sua sfida più recente.

I modelli generativi

I generatori moderni, dalle architetture a diffusione per le immagini ai modelli autoregressivi per il testo e l'audio, producono contenuti di qualità tale da superare l'esame percettivo. Per l'analista ne discende un cambiamento di postura: l'autenticità di un contenuto digitale non si presume più dall'aspetto, ma si accerta con metodo, a partire dalla provenienza del file e dalla sua catena di acquisizione.

Deepfake di immagine e audio

Per le immagini, l'analisi cerca incoerenze di illuminazione e prospettiva, anomalie anatomiche residue, tracce statistiche del processo di generazione e incongruenze nei metadati. Per l'audio sintetico si esaminano spettro, prosodia, respiri e rumori di fondo, oltre alla coerenza del canale di registrazione dichiarato. In entrambi i casi i detector automatici forniscono un primo vaglio, con tassi d'errore che dipendono fortemente dal generatore, dalla compressione e dal dominio: le loro uscite sono indizi da integrare, non verdeti.

Video sintetico

Il video manipolato offre più superfici d'attacco, dal face swap al reenactment al lip-sync, ma anche più superfici di verifica, perché la coerenza va mantenuta fotogramma per fotogramma. L'analisi cerca incoerenze temporali, come lo sfarfallio dei bordi del volto e l'illuminazione che non segue il movimento; incoerenze fisiologiche, come il battito delle palpebre e i segnali di pulsazione cutanea rilevabili con tecniche rPPG; incoerenze di compressione, come tracce di doppia codifica e regioni con statistiche diverse dal resto del fotogramma. Nessun indizio è decisivo da solo: la relazione costruisce una convergenza. La compressione dei social degrada proprio i segnali più informativi, e per questo la disponibilità del file originale vale quanto l'analisi stessa.

Testo generato: i limiti dei detector

I rilevatori di testo sintetico misurano regolarità statistiche tipiche dell'output dei modelli, come perplexity e uniformità stilistica. Funzionano meglio su testi lunghi, in inglese, non editati; un intervento umano leggero o una parafrasi abbattano la rilevazione. Il problema più serio è la distribuzione dei falsi positivi, che colpiscono sistematicamente i non madrelingua e gli stili

formulaici. Allo stato dell'arte nessun detector di testo ha caratteristiche compatibili con l'uso probatorio isolato: il giudizio si costruisce su metadati, cronologia delle versioni, coerenza con i campioni noti dell'autore dichiarato e tracce d'uso di strumenti di generazione sul dispositivo.

Analisi di manipolazione ed ELA

Per le immagini fotografiche, l'Error Level Analysis evidenzia regioni con livelli di ricompressione incoerenti, possibile indizio di fotomontaggio. La tecnica è nota per la facilità con cui viene sovrainterpretata: molti fattori innocui producono pattern simili. Va usata come strumento esplorativo accanto ad analisi più solide, come lo studio dei metadati, delle tabelle di quantizzazione e del rumore del sensore.

Provenance e C2PA

Lo standard C2PA lega al contenuto credenziali firmate crittograficamente che ne documentano origine e trasformazioni. Quando presenti e valide, le credenziali offrono una verifica forte della provenienza; la loro rimozione, però, è banale, e l'assenza non dice nulla. Si tratta di metadati accanto al segnale, da distinguere dal watermarking, che agisce dentro il segnale.

Watermarking dei generatori

Il watermarking marca il contenuto alla nascita: pattern impercettibili nelle immagini, distorsioni statistiche controllate nella scelta delle parole per il testo, segnali fuori banda per l'audio. Il quadro normativo europeo spinge in questa direzione, richiedendo che i contenuti sintetici siano riconoscibili in formato leggibile dalla macchina. I limiti sono strutturali: il watermark prova la provenienza solo se c'è, perché i modelli a pesi aperti e gli strumenti ostili non marcano; la rimozione è spesso praticabile; la verifica può richiedere la collaborazione del generatore. È un segnale asimmetrico, utile quando presente e muto quando assente, e la relazione deve dirlo esplicitamente prima che lo faccia il controesame.

Attacchi ai sistemi di AI

I sistemi di intelligenza artificiale sono a loro volta bersagli. Gli esempi adversarial inducono errori di classificazione con perturbazioni impercettibili; il data poisoning inquina l'addestramento; il model stealing e la membership inference estraggono informazioni dal modello. Per i sistemi basati su LLM, la prompt injection inserisce istruzioni ostili nei dati che il modello elabora, piegandone il comportamento. La OWASP LLM Top 10 cataloga queste classi di rischio ed è il riferimento con cui valutare, anche in chiave peritale, la robustezza di un sistema coinvolto in un incidente.

Per il laboratorio. Gli studenti analizzano un insieme di immagini e brevi video, in parte autentici e in parte sintetici, documentando per ciascun elemento gli indizi raccolti e formulando conclusioni con il lessico della compatibilità; un esercizio separato mostra una prompt injection su un sistema RAG dimostrativo.

Lecture consigliate: H. Farid, “Photo Forensics”, MIT Press; M. Barni et al., survey sulla multimedia forensics in IEEE Signal Processing Magazine; OWASP, “Top 10 for Large Language Model Applications”.

Capitolo 11. Profili giuridici ed etici dell'AI forense

Le tecniche dei capitoli precedenti producono valore probatorio soltanto se superano il vaglio giuridico. Questo capitolo raccoglie i criteri di ammissibilità, i vincoli normativi e i doveri professionali che governano l'uso dell'intelligenza artificiale nelle indagini digitali.

Ammissibilità della prova algoritmica

L'output di un algoritmo non si autentica da solo: entra nel procedimento attraverso la persona che ne risponde, consulente o perito, chiamata a spiegare metodo, limiti e margini di errore. Un sistema opaco, di cui nessuno sa rendere conto, indebolisce il contraddittorio e viene guardato con sfavore. I principi generali della prova scientifica, dalla catena di custodia alla ripetibilità alla motivazione, si applicano agli strumenti di AI come a ogni altro strumento tecnico.

I criteri Daubert

La giurisprudenza statunitense sul caso Daubert ha fissato criteri che, pur non vincolanti negli ordinamenti europei, sono diventati il riferimento internazionale per valutare l'affidabilità di un metodo scientifico: verificabilità e falsificabilità, revisione paritaria e pubblicazione, tasso di errore noto, standard di controllo, accettazione nella comunità scientifica di riferimento. Per i sistemi di AI il criterio critico è il tasso di errore noto: deve essere misurato su dati pertinenti al caso, aggiornato nel tempo e riferito alla popolazione rilevante. Un tasso aggregato che nasconde forti differenze per sottogruppi non soddisfa il criterio nella sostanza.

L'art. 360 c.p.p. e gli accertamenti irripetibili

Nell'ordinamento italiano, l'accertamento tecnico che modifica lo stato delle cose è irripetibile e attiva le garanzie dell'art. 360 c.p.p.: avviso alle parti, facoltà di nominare consulenti, contraddittorio tecnico. La qualificazione di un'operazione come ripetibile o irripetibile va decisa prima di agire e motivata. L'uso di sistemi di AI aggiunge una dimensione: anche quando l'operazione sul reperto è ripetibile, la riproducibilità del risultato dipende da versione del modello, configurazione e soglie, che vanno quindi congelate e documentate.

Bias e fairness

L'equità di un sistema si misura. Le metriche principali confrontano i gruppi su cui il sistema opera: la demographic parity richiede tassi di esito simili, le equalized odds richiedono tassi di falso positivo e falso negativo simili, la calibrazione per gruppo richiede che lo stesso score significhi la stessa probabilità in ogni gruppo. Un risultato teorico importante stabilisce che queste proprietà sono in generale incompatibili fra loro quando i gruppi hanno prevalenze diverse: la scelta di quale garantire è una decisione di valore, da esplicitare. Gli studi del NIST sul riconoscimento facciale (programma FRVT) hanno documentato differenze di un ordine di grandezza nei tassi di falso positivo fra gruppi

demografici in molti sistemi commerciali: per il perito, chiedere le metriche disaggregate è parte della diligenza dovuta.

GDPR

Il trattamento di dati personali nelle indagini incontra i principi del regolamento europeo: liceità e base giuridica, minimizzazione, limitazione della conservazione, sicurezza. Due punti toccano direttamente la pratica forense. Il primo è la proporzionalità: i metadati di flusso, ad esempio, incidono sulla riservatezza meno della cattura integrale dei contenuti, e la scelta dello strumento meno invasivo adeguato allo scopo va documentata. Il secondo è l'art. 33, che impone al titolare la notifica delle violazioni di dati personali: le analisi forensi di un incidente alimentano direttamente il contenuto della notifica, dalla natura della violazione alla portata stimata.

AI Act

Il regolamento europeo sull'intelligenza artificiale introduce obblighi graduati per rischio. Per la materia di questo corso rilevano in particolare gli obblighi di trasparenza sui contenuti sintetici, che devono essere riconoscibili in formato leggibile dalla macchina, e la disciplina dei sistemi usati dalle autorità, fra cui l'identificazione biometrica, soggetta a condizioni stringenti. Il quadro attuativo è in evoluzione: le affermazioni al riguardo vanno datate.

Giurisprudenza in formazione

Sulla prova generata o selezionata dall'AI i precedenti consolidati sono pochi e gli orientamenti divergono fra ordinamenti. La prassi prudente per il consulente: datare ogni affermazione normativa e giurisprudenziale, distinguere norma vigente, orientamento e dottrina, e non far dipendere la tenuta della relazione da una tesi giuridica controversa. L'ancoraggio resta il metodo tecnico documentato, che sopravvive ai mutamenti di orientamento.

Responsabilità e deontologia

Ogni riga depositata appartiene a chi la firma: l'uso di strumenti di AI, dichiarato e verificato, è legittimo, ma non trasferisce la responsabilità. I nuclei comuni dei codici deontologici di settore valgono anche qui: competenza dichiarata, imparzialità, riservatezza, integrità del metodo, trasparenza sugli strumenti. Le certificazioni, dall'accreditamento ISO/IEC 17025 dei laboratori ai percorsi individuali delle organizzazioni internazionali di settore, non sono requisiti di legge in Italia ma pesano sulla credibilità in sede di esame. L'uso di AI aggiunge un dovere emergente: mantenere la competenza necessaria a capire e spiegare gli strumenti che si firmano.

Per il laboratorio. Gli studenti ricevono una relazione tecnica simulata che impiega un sistema di AI e redigono il parere di ammissibilità: criteri soddisfatti, carenze, domande da porre in sede di esame del consulente.

Lecture consigliate: Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993); P. Grother, M. Ngan, K. Hanaoka, “Face Recognition Vendor Test, Part 3: Demographic Effects”, NISTIR 8280; Regolamento (UE) 2024/1689 (AI Act), disposizioni sulla trasparenza dei contenuti sintetici.

Capitolo 12. Laboratorio integrato e casi di studio

Il laboratorio conclusivo mette alla prova l'intero percorso: i gruppi ricevono un caso realistico, lo lavorano con la doppia metodologia del corso e lo difendono in una discussione con controesame. Tutto il materiale didattico è sintetico o proviene da dataset pubblici.

Il formato

Ogni gruppo, di tre o quattro studenti, riceve un fascicolo: quesito, reperti, vincoli giuridici. La pipeline richiesta integra la metodologia classica e gli strumenti di AI, con replica manuale su campione e log completo delle operazioni. La consegna finale comprende la relazione peritale e i materiali di supporto che consentono a un terzo di ripercorrere ogni passo.

I sette casi di studio

I casi coprono lo spettro del corso e sono assegnati alle specializzazioni. Il primo è un ransomware aziendale, con la kill chain da ricostruire su immagine disco e dump di memoria e un triage dei log assistito dall'AI e validato in cieco. Il secondo è una frode con deepfake vocale, sul modello del caso Arup del 2024, in cui contano l'autenticità di audio e video e i protocolli anti-frode. Il terzo riguarda uno smartphone in sede penale: messaggistica, geolocalizzazione e attribuzione del dispositivo, con la discussione del precedente Apple-FBI. Il quarto rilegge il caso Silk Road come attribuzione di identità online tramite tracce testuali. Il quinto è un'analisi di email sul corpus Enron, con topic modeling e grafo relazionale. Il sesto chiede di stabilire se immagini e documenti prodotti in giudizio siano autentici o fabbricati con AI generativa. Il settimo ricostruisce la compromissione di un chatbot aziendale simulato, attaccato con prompt injection ed esfiltrazione. Ogni caso contiene elementi che richiedono sia strumenti di AI sia la loro validazione manuale.

Il caso 6 in dettaglio

Il sesto caso, la prova sospetta di essere fabbricata con AI generativa, merita una trattazione estesa perché integra più moduli di ogni altro. Nella variante svolta in aula, in un procedimento civile una parte produce lo screenshot di una conversazione che l'altra parte disconosce integralmente. Il quesito: la conversazione è autentica, fabbricata, o non determinabile? Il percorso atteso parte dalla gerarchia delle fonti, perché lo screenshot è il gradino più basso e la prima mossa è chiedere i dispositivi e i database originali delle app; prosegue con l'analisi dell'immagine, dai metadati alle tracce di fotomontaggio o generazione; cerca riscontri esterni in tabulati, backup e log dei provider; e si chiude con la formulazione, tre ipotesi esplicite con il peso dell'evidenza per ciascuna. Ogni gruppo riceve una variante diversa del materiale: in alcune la conversazione è autentica, in altre fabbricata, in altre il quadro resta indeterminabile. La conclusione "non determinabile", quando i dispositivi originali mancano e i riscontri tacciono, è un esito professionalmente corretto, e la valutazione premia chi lo riconosce: trovare una fabbricazione che il materiale non sostiene costa più punti che concludere onestamente per l'indeterminabilità.

Le trappole

I materiali contengono insidie deliberate, calibrate sugli errori professionali più frequenti: un hash che non torna in un verbale, un orario incoerente fra fusi, un falso positivo plausibile del triage, un dato che invita a una conclusione oltre l'evidenza. Le trappole non sono rivelate in anticipo; emergono, o non emergono, nella relazione e nella discussione. Il loro scopo è didattico: abituare a un livello di vigilanza che il materiale pulito non allena.

La peer review

Ogni gruppo riceve la relazione di un altro gruppo, su un caso diverso dal proprio, e produce una memoria critica di due pagine. La griglia di lettura riprende il mestiere del consulente di parte: catena di custodia verificabile, metodo dichiarato e ripetibile, versioni degli strumenti, salti logici fra dato e conclusione, alternative non considerate, limiti taciuti. La critica si rivolge al metodo e mai alla persona; ogni rilievo è specifico e verificabile; riconoscere ciò che è corretto rende credibile ciò che si contesta. La doppia esperienza, scrivere sapendo di essere letti da pari e leggere con il compito di trovare le crepe, è il correttivo più efficace contro le conclusioni eccessive.

La relazione peritale finale

La relazione segue la struttura professionale: quesito, reperti e catena di custodia, metodo, operazioni svolte, risultati, valutazione, limiti. Cinque voci sull'uso dell'AI sono obbligatorie e vengono verificate puntualmente: quali strumenti sono stati impiegati e in quale versione; su quali dati sono stati addestrati o validati, per quanto noto; quali metriche e quali tassi di errore sono applicabili; quale verifica umana è stata eseguita e con quale esito; quali limiti restano. Una relazione tecnicamente solida ma carente su queste voci non supera la soglia.

La discussione orale

La discussione simula l'esame del consulente in udienza e si articola in tre momenti: l'esposizione a un giudice non tecnico, che richiede chiarezza senza semplificazioni scorrette; il controesame, condotto da un docente nel ruolo del consulente avversario sui punti deboli della relazione depositata; un quesito a sorpresa, un elemento nuovo introdotto per valutare il ragionamento in tempo reale. I criteri premiano la condotta professionale: distinguere ciò che si sa da ciò che si inferisce, ammettere i limiti prima che vengano estorti, correggersi senza crollare, non difendere l'indifendibile. Il voto della discussione pesa quanto la relazione scritta.

Per il laboratorio. Il capitolo è esso stesso il laboratorio: la preparazione consiste nel rileggere i protocolli di validazione del capitolo 8 e le cinque voci obbligatorie, perché compariranno, in qualche forma, sia nella relazione sia nel controesame.

Lettere consigliate: E. Casey, "Digital Evidence and Computer Crime", Academic Press, capitoli sulla redazione della relazione; ENFSI, "Best Practice Manual for the Forensic Examination of Digital Technology"; NIST CFReDS e Digital Corpora, repository dei dataset didattici.

Capitolo S1.2. Incident response e intrusioni su host

Il capitolo segue l'analista che ricostruisce un'intrusione su un sistema Windows partendo da memoria, disco e registri degli eventi, e presuppone il modulo base di memory forensics.

Il ciclo di risposta agli incidenti

La pubblicazione NIST SP 800-61 Rev. 3 (aprile 2025) organizza la risposta agli incidenti attorno alle funzioni del Cybersecurity Framework 2.0 e la integra nella gestione complessiva del rischio, superando il ciclo in quattro fasi della revisione precedente. Per il forenser la conseguenza pratica riguarda la preparazione: la qualità dell'analisi dipende da decisioni prese prima dell'incidente, come quali registri conservare, per quanto tempo e con quali orologi sincronizzati. Senza registrazione degli eventi l'analista deve ripiegare su artefatti indiretti.

Gli event log di Windows come prova

Nel registro Security di Windows l'evento 4624 attesta un accesso riuscito e il campo logon type ne distingue la natura: 2 per l'accesso interattivo, 3 per l'accesso di rete, 10 per l'accesso tramite Desktop remoto. L'evento 4625 registra un accesso fallito; una serie di fallimenti seguita da un successo suggerisce tentativi di individuare la password. L'evento 4688 documenta la creazione di un processo e riporta la riga di comando soltanto se la relativa impostazione di audit è stata attivata. Sysmon, componente gratuita di Microsoft, aggiunge hash degli eseguibili, connessioni di rete e modifiche al registro.

La dimensione massima dei file determina quanto indietro si può risalire, perché gli eventi più vecchi vengono sovrascritti. La cancellazione dei registri lascia a sua volta una traccia, con l'evento 1102 nel registro Security e l'evento 104 nel registro System. I file .evtx vanno infine estratti dalla copia forense o raccolti con strumenti che ne calcolino l'hash, come ogni altro reperto.

Persistenza

Per sopravvivere ai riavvii l'attaccante installa un meccanismo di persistenza. I più comuni sono le chiavi di esecuzione automatica del registro (Run e RunOnce); i servizi, la cui installazione genera l'evento 7045 nel registro System; le attività pianificate, registrate con l'evento 4698 se l'audit è attivo; le sottoscrizioni WMI, che associano un filtro a un'azione e compaiono nel registro WMI-Activity. Ogni meccanismo lascia tracce in più fonti indipendenti, e il valore della ricostruzione cresce con la loro convergenza, come per gli artefatti del modulo 3.

Movimento laterale

Il movimento laterale porta l'attaccante da un sistema compromesso agli altri. Una sessione di Desktop remoto produce sul sistema di destinazione un evento 4624 con logon type 10 e un evento

1149 nel registro operativo di Remote Desktop Services, mentre sul sistema di origine restano la cronologia delle connessioni nel registro e la cache delle immagini della sessione. PsExec installa sul sistema remoto il servizio PSEXESVC, rilevabile con l'evento 7045. Il controllo remoto via WinRM e PowerShell lascia tracce nei registri dedicati e, se è attivo il logging dei blocchi di script, il testo dei comandi nell'evento 4104. Quando l'attaccante usa credenziali legittime sottratte, la sua attività somiglia all'amministrazione ordinaria, e la distinzione si ottiene correlando orari, sorgenti e sequenze sui sistemi coinvolti.

Ransomware e doppia estorsione

Un attacco ransomware segue di solito una cronologia riconoscibile: accesso iniziale, ricognizione, movimento laterale, sottrazione dei dati, cancellazione delle copie shadow e infine cifratura. Nel modello della doppia estorsione l'esfiltrazione precede la cifratura e alimenta la minaccia di pubblicazione. Datare l'esfiltrazione ha un rilievo giuridico immediato, perché la notifica al Garante prevista dall'art. 33 del GDPR va effettuata entro 72 ore dalla conoscenza della violazione e deve descriverne natura e portata. Le fonti utili sono i flussi di rete, i log dei proxy e le tracce di esecuzione di strumenti di archiviazione e trasferimento in artefatti come Prefetch e Amcache. La cifratura distrugge i contenuti ma raramente i metadati: Master File Table e USN Journal permettono spesso di datarne l'avvio.

Raccolta su larga scala ed EDR

Su centinaia di macchine la copia integrale di ogni disco è impraticabile, e strumenti come Velociraptor e KAPE raccolgono artefatti mirati in tempi brevi. La raccolta mirata è una scelta di proporzionalità e tempestività che va motivata, e il suo perimetro va dichiarato in relazione perché delimita ciò che l'analisi potrà affermare. L'integrità della raccolta remota si documenta con gli hash dei pacchetti, il log dello strumento e la configurazione usata. La telemetria delle piattaforme EDR è una fonte preziosa, conservata però dal fornitore per periodi limitati e condizionata da regole di raccolta che vanno conosciute e dichiarate.

Automazione e AI nella risposta

Le regole Sigma descrivono rilevazioni sui registri in un formato aperto, indipendente dal prodotto. Hayabusa le applica ai file .evtx e ne ricava una timeline degli eventi rilevanti. Le tecniche di anomaly detection sugli accessi e sui processi riprendono i temi della traccia core: prevalenza del fenomeno, soglie, falsi positivi. I modelli linguistici aiutano a riassumere una timeline o a interpretare un comando offuscato, purché ogni affermazione sia verificata sull'evento di origine. Il protocollo di validazione su campione del modulo 8 si applica anche alla risposta agli incidenti.

Contenimento senza distruggere la prova

Nelle prime ore l'esigenza di ripristinare i servizi entra in tensione con quella di preservare la prova. Isolare un sistema dalla rete conserva la memoria che lo spegnimento cancellerebbe; acquisire la

RAM prima di un riavvio e salvare lo stato di una macchina virtuale prima della reinstallazione sono operazioni rapide che evitano perdite irreparabili. Ogni azione di risposta è un'operazione sul reperto e va annotata con ora, operatore e motivazione. Nei procedimenti penali le operazioni irripetibili richiamano le garanzie dell'art. 360 c.p.p.; in azienda il registro delle azioni svolge la funzione del verbale e conserva il valore della prova per un eventuale giudizio.

Per il laboratorio. Su dataset pubblici di attacchi simulati (immagini disco, dump di memoria, event log) gli studenti ricostruiscono un'intrusione completa, ne mappano le tecniche sulla matrice MITRE ATT&CK e redigono un incident report con una sintesi in forma di relazione tecnica.

Lettere consigliate: NIST SP 800-61 Rev. 3, *Incident Response Recommendations and Considerations for Cybersecurity Risk Management*; J. Luttgens, M. Pepe, K. Mandia, *Incident Response & Computer Forensics*, 3a ed., McGraw-Hill; MITRE ATT&CK, matrice Enterprise.

Capitolo S2.2. Cloud, identità ed e-evidence

Nei servizi cloud la prova risiede presso un terzo e l'identità dell'utente diventa il perimetro da difendere e da analizzare. Il capitolo tratta le fonti dei provider, la compromissione degli account e gli strumenti giuridici per ottenere dati conservati presso terzi. Presuppone il modulo base di network e cloud forensics.

Gli audit log dei provider

AWS CloudTrail documenta le chiamate alle API con identità del chiamante, orario, indirizzo di origine e parametri; la cronologia degli eventi di gestione resta consultabile per 90 giorni, mentre una conservazione più lunga richiede un trail configurato in anticipo. Microsoft Entra ID conserva i log di accesso e di audit per periodi che dipendono dalla licenza, da 7 a 30 giorni. Lo Unified Audit Log di Microsoft 365 raccoglie le attività su posta, documenti e collaborazione, con una conservazione standard di 180 giorni estendibile con le licenze superiori. Google Workspace offre registri di audit per accessi, amministrazione, Drive e Gmail, con periodi di conservazione diversi a seconda del tipo.

Le finestre di conservazione sono quindi brevi e dipendono da configurazione e licenza. Il primo atto dell'analista è esportare i registri disponibili calcolandone l'hash e documentare le impostazioni di conservazione, che spiegano anche perché alcuni dati non esistono più.

La compromissione dell'identità

Le tecniche più diffuse mirano a sottrarre la sessione autenticata, oltre alle credenziali. Il phishing adversary-in-the-middle interpone un proxy fra la vittima e il servizio legittimo e cattura credenziali e cookie di sessione, aggirando l'autenticazione a più fattori tradizionale. Il consenso OAuth fraudolento induce l'utente a concedere permessi a un'applicazione controllata dall'attaccante, che conserva l'accesso anche dopo il cambio della password. L'abuso delle notifiche MFA consiste nell'inviare richieste ripetute finché l'utente ne approva una. L'analisi incrocia sequenze di accesso, indirizzi IP, user agent, identificativi di sessione e applicazioni autorizzate, per separare l'attività del titolare da quella dell'intruso quando entrambi usano lo stesso account.

Business Email Compromise in Microsoft 365

Nella frode BEC l'attaccante compromette una casella, osserva le conversazioni e interviene al momento opportuno, di solito chiedendo un pagamento su un conto diverso da quello abituale. Per nascondere le risposte crea regole di inoltramento o di spostamento dei messaggi, la cui creazione resta nello Unified Audit Log. I log di accesso datano la compromissione, il message trace ricostruisce i flussi di posta e l'evento MailItemsAccessed, la cui disponibilità dipende dal livello di audit attivo, aiuta a stabilire quali messaggi siano stati letti. La stima serve sia alla notifica prevista dal GDPR sia alla ricostruzione della frode per la parte offesa.

Container e Kubernetes

I carichi containerizzati possono vivere pochi minuti, e con essi scompaiono le tracce. Le fonti utili sono gli audit log dell'API server di Kubernetes, se abilitati, i log dei nodi e del runtime, le immagini conservate nei registry e gli snapshot dei volumi e dei nodi. Le funzionalità di checkpoint dei container, ancora in evoluzione, permettono di salvare lo stato di un carico in esecuzione. In relazione l'immagine va identificata con il suo digest crittografico, perché lo stesso tag può indicare contenuti diversi in momenti diversi.

Le prove presso i provider

Quando i dati utili sono conservati dal fornitore, la prima richiesta riguarda il congelamento, che ne impedisce la cancellazione mentre si prepara la richiesta di produzione. Occorre distinguere fra dati dell'abbonato, dati di traffico e contenuti, soggetti a regimi giuridici diversi. La catena di custodia include il provider: va documentato chi ha estratto i dati, con quale portale o strumento e con quali hash.

Il quadro e-evidence dell'Unione europea

Il Regolamento (UE) 2023/1543 è pienamente applicabile dal 18 agosto 2026 e consente all'autorità giudiziaria di uno Stato membro di rivolgersi direttamente al fornitore di servizi stabilito in un altro Stato. L'ordine europeo di produzione impone di consegnare i dati entro 10 giorni, ridotti a 8 ore nei casi di emergenza; l'ordine europeo di conservazione obbliga a preservarli per 60 giorni, prorogabili di 30 per consentire la richiesta di produzione. I dati dell'abbonato e quelli richiesti al solo fine di identificazione possono essere chiesti per qualsiasi reato, i dati di traffico e di contenuto soltanto per reati puniti con pena massima di almeno tre anni o per specifiche fattispecie, fra cui la criminalità informatica. Gli ordini sono indirizzati agli stabilimenti designati o ai rappresentanti legali nominati ai sensi della Direttiva (UE) 2023/1544.

Cooperazione internazionale

Fuori dall'Unione il riferimento è il Secondo Protocollo addizionale alla Convenzione di Budapest, aperto alla firma nel 2022, che prevede fra l'altro richieste dirette ai registrar per i dati di registrazione dei domini e ai provider per i dati degli abbonati, oltre a forme di cooperazione accelerata in emergenza. All'interno dell'Unione resta disponibile l'ordine europeo di indagine (Direttiva 2014/41/UE, attuata in Italia con il d.lgs. 108/2017), strumento fra autorità giudiziarie con tempi più lunghi, utile per atti diversi dalla produzione di dati. Gli atti spettano all'autorità giudiziaria; il consulente tecnico deve conoscerne tempi e contenuti per pianificare la raccolta e valutare la completezza dei dati ricevuti.

AI sugli accessi

Le piattaforme di sicurezza segnalano accessi anomali con tecniche come l'impossible travel, che individua accessi da luoghi incompatibili con i tempi di spostamento, e l'analisi comportamentale di

utenti ed entità (UEBA), che apprende profili abituali e segnala gli scostamenti. Su milioni di accessi anche un tasso di falsi positivi molto basso produce migliaia di allarmi, e VPN, proxy e reti mobili generano spostamenti apparentemente impossibili del tutto innocui. In relazione la segnalazione automatica compare come indizio, accompagnata dagli eventi grezzi che la sostengono e dalla verifica dell'analista.

Per il laboratorio. Gli studenti indagano su un tenant Microsoft 365 simulato vittima di BEC, con log sintetici, ricostruiscono un accesso illecito da eventi CloudTrail sintetici e redigono, a scopo di esercitazione, una richiesta di conservazione con i contenuti previsti per l'ordine europeo.

Lettere consigliate: NIST IR 8006, *NIST Cloud Computing Forensic Science Challenges* (2020); Regolamento (UE) 2023/1543 e Direttiva (UE) 2023/1544; documentazione di Microsoft Purview Audit e di AWS CloudTrail.

Capitolo S3.2. Mobile avanzato, IoT e veicoli

Il capitolo estende la mobile forensics del modulo base in due direzioni: verso l'interno dei dispositivi, con gli stati di sicurezza e l'analisi delle applicazioni, e verso l'esterno, con gli oggetti connessi e i mezzi di trasporto che accompagnano la vita quotidiana.

Stati di sicurezza e scelte di sequestro

Un dispositivo acceso può trovarsi in due condizioni molto diverse. Prima del primo sblocco successivo all'avvio (BFU, before first unlock) la maggior parte dei dati dell'utente resta cifrata con chiavi non disponibili; dopo il primo sblocco (AFU, after first unlock) molte chiavi sono caricate in memoria e le possibilità di estrazione aumentano. Su iOS i file appartengono a classi di protezione: la protezione completa li rende accessibili solo a dispositivo sbloccato, mentre la classe predefinita per i dati delle app li rende disponibili dopo la prima autenticazione. Android adotta la cifratura per file con due aree: quella cifrata con le credenziali dell'utente (CE), disponibile dopo lo sblocco, e quella cifrata con chiavi del dispositivo (DE), disponibile già dopo l'avvio.

Ne derivano scelte operative da verbalizzare. Un telefono trovato acceso e in stato AFU va mantenuto alimentato e isolato dalla rete, perché lo spegnimento lo riporterebbe allo stato BFU. Alcune versioni recenti dei sistemi operativi introducono il riavvio automatico dopo un periodo di inattività, con lo stesso effetto: il tempo che separa sequestro ed estrazione diventa così un fattore tecnico, oltre che organizzativo.

Le app oltre i parser automatici

I parser automatici coprono le versioni note delle applicazioni più diffuse, ma un aggiornamento può cambiare lo schema dei dati da un giorno all'altro. L'analista deve quindi saper leggere direttamente i formati: database SQLite con il relativo write-ahead log, property list binarie su iOS, messaggi protobuf (un formato binario che senza la definizione dello schema si decodifica solo in parte), archivi LevelDB usati da molte applicazioni basate su tecnologie web, cache e file multimediali. La verifica manuale di un campione di risultati permette di accertare che lo strumento abbia interpretato correttamente i dati.

Timeline delle attività dell'utente

I sistemi mobili registrano l'uso del dispositivo per funzioni di assistenza e di benessere digitale. Su iOS queste informazioni si trovano nel database KnowledgeC nelle versioni meno recenti e negli stream Biome nelle versioni attuali; su Android nei dati di Digital Wellbeing e nelle statistiche d'uso del sistema. Questi artefatti documentano applicazioni in primo piano, stato dello schermo e riproduzione multimediale, e richiedono cautela interpretativa. Un'app in primo piano descrive lo stato del dispositivo e non l'attenzione della persona; granularità e conservazione cambiano con le

versioni; il significato di un campo va verificato su un dispositivo di prova prima di affermarlo in relazione.

Smart home e assistenti vocali

Hub, telecamere, serrature e termostati conservano dati in tre luoghi: sul dispositivo, nell'app companion dello smartphone e nel cloud del produttore. Gli assistenti vocali mantengono la cronologia dei comandi, talvolta con frammenti audio, associata all'account dell'utente, che si acquisisce tramite l'app, con le esportazioni dei dati dell'account o con una richiesta al fornitore. Alcuni dispositivi conservano i propri log solo per poche ore, e la configurazione del router domestico, con l'elenco dei dispositivi connessi, completa il quadro.

Veicoli

I sistemi di infotainment conservano i dispositivi associati via Bluetooth, rubriche e registri delle chiamate sincronizzati, destinazioni di navigazione ed eventi come l'apertura delle porte o l'accensione. La telematica trasmette dati al produttore o a servizi connessi, acquisibili presso il fornitore. Gli strumenti di estrazione sono specializzati e supportano marche e modelli specifici; l'acquisizione è spesso fisica sul modulo e va documentata con il numero di telaio del veicolo. Il registratore di dati di evento, reso obbligatorio nell'Unione per le nuove autovetture e i veicoli commerciali leggeri dal Reg. (UE) 2019/2144, registra i parametri di guida nell'imminenza di un urto e va tenuto distinto dagli EDR informatici trattati nella specializzazione S1.

Droni

I droni registrano log di volo con posizione, quota, telemetria ed eventi, distribuiti fra velivolo, radiocomando e app di controllo, mentre foto e video portano metadati di georeferenziazione. Questi dati permettono di ricostruire traiettorie e punti di decollo, con le cautele dovute alla precisione del posizionamento satellitare e alla possibilità di manomissione dei file.

Validazione degli strumenti

Il programma Computer Forensics Tool Testing del NIST pubblica specifiche e rapporti di test su strumenti di acquisizione e analisi, inclusi quelli per dispositivi mobili. In laboratorio la validazione si esegue su dataset con verità nota, costruiti registrando le attività svolte sul dispositivo, e sul confronto fra strumenti diversi, spiegando ogni divergenza prima di trarre conclusioni. Le versioni di estrattori e parser vanno dichiarate, e la validazione interna dei metodi rientra fra i requisiti dei laboratori accreditati secondo ISO/IEC 17025.

AI nell'analisi mobile

Le estrazioni moderne producono centinaia di migliaia di immagini, lunghe raccolte di note vocali e chat in più lingue. Classificatori di immagini, trascrizione automatica, traduzione e OCR sugli screenshot accelerano il triage. Ogni elaborazione genera un derivato che va tracciato con modello e

versione e verificato su campione. Le trascrizioni e le traduzioni destinate al fascicolo vanno controllate da persona qualificata, secondo il principio, enunciato nella premessa, per cui l'AI seleziona e l'umano certifica.

Per il laboratorio. Su immagini di riferimento pubbliche gli studenti confrontano un'estrazione logica e una full file system, ricostruiscono la timeline di una giornata d'uso e redigono un breve rapporto di validazione degli strumenti impiegati, con le divergenze riscontrate e la loro spiegazione.

Lettere consigliate: NIST SP 800-101 Rev. 1, *Guidelines on Mobile Device Forensics* (2014); R. Tamma et al., *Practical Mobile Forensics*, 4a ed., Packt; NIST, rapporti del programma Computer Forensics Tool Testing sugli strumenti mobili.

Capitolo S4.2. LLM, RAG e visione artificiale nelle indagini

Il capitolo riguarda la progettazione e la valutazione di sistemi di intelligenza artificiale che lavorano sul materiale di un'indagine. Il criterio guida è quello della traccia core: ogni risultato destinato al fascicolo deve essere riconducibile a una fonte verificabile e difendibile da chi firma. Presuppone il modulo base di NLP e analisi investigativa.

La pipeline RAG sui fascicoli

Un sistema retrieval-augmented generation combina un motore di ricerca sul fascicolo con un modello linguistico che risponde usando soltanto i passaggi recuperati. La progettazione comincia dalla suddivisione dei documenti in segmenti, ciascuno con un riferimento stabile: identificativo del reperto, pagina o posizione, hash del documento di origine. I segmenti vengono indicizzati con embedding e recuperati in base alla domanda; il modello genera la risposta vincolata a quei passaggi e cita la fonte di ogni affermazione. Un'affermazione priva di citazione va trattata come non supportata. Il limite principale riguarda il recupero: se il sistema non trova un passaggio rilevante, la risposta negativa riguarda il comportamento dello strumento e non dice nulla sul contenuto del fascicolo.

Valutare un LLM su compiti forensi

La valutazione richiede un campione annotato a mano, preparato prima di esaminare gli output del sistema per evitare l'effetto di ancoraggio. Le misure utili sono la completezza delle entità estratte rispetto all'annotazione, la fedeltà dei riassunti, verificando che ogni affermazione trovi riscontro nel testo di origine, la quota di affermazioni non supportate e la correttezza delle citazioni. I valori ottenuti descrivono quel corpus e quella configurazione: cambiare modello, prompt o indice richiede una nuova misura, come accade per ogni strumento validato in laboratorio.

Riproducibilità

La documentazione del metodo comprende identificativo e versione del modello, prompt di sistema e utente, parametri di generazione come temperatura, lunghezza massima e, se disponibile, il seme, insieme alla versione dell'indice. Anche a temperatura zero l'output può variare fra esecuzioni o dopo un aggiornamento del servizio. Per questo si conservano gli output con data e hash, e i modelli eseguiti in locale consentono di congelare la versione usata. Il log così costruito svolge per l'analisi assistita la funzione che il verbale svolge per le operazioni tecniche.

Riservatezza

Il materiale di un'indagine penale è coperto dal segreto investigativo (art. 329 c.p.p.) e contiene dati personali, spesso appartenenti a categorie particolari. Per i trattamenti svolti dalle autorità competenti a fini di indagine e perseguimento dei reati il riferimento è la Direttiva (UE) 2016/680, attuata in Italia con il d.lgs. 51/2018; per consulenti di parte e aziende si applica il GDPR. In entrambi i casi sono da preferire modelli eseguiti in locale o su infrastrutture controllate. L'uso di servizi esterni richiede clausole contrattuali che escludano l'impiego dei dati per l'addestramento e ne definiscano conservazione e localizzazione, oltre all'accordo con il responsabile del trattamento previsto dall'art. 28 del GDPR.

Visione artificiale per il triage

Rilevamento di oggetti, riconoscimento del testo nelle immagini, classificazione delle scene e ricerca per somiglianza visiva permettono di orientarsi in archivi di centinaia di migliaia di file. Gli errori tipici riguardano categorie rare, immagini di bassa qualità e contesti diversi da quelli dei dati di addestramento. L'uso corretto è la prioritizzazione, con l'esame da parte di una persona di ogni elemento citato in relazione.

Riconoscimento facciale

Le prestazioni degli algoritmi di riconoscimento facciale variano fra gruppi demografici, come hanno documentato gli studi NIST FRVT su molti sistemi commerciali. L'AI Act vieta l'identificazione biometrica remota in tempo reale in spazi accessibili al pubblico a fini di contrasto, salvo eccezioni tassative e previa autorizzazione (art. 5), e classifica ad alto rischio l'identificazione a posteriori (Allegato III); per questi sistemi il regolamento omnibus sull'AI ha differito gli obblighi al 2 dicembre 2027. L'esito di una ricerca facciale resta in ogni caso un elenco di candidati, da trattare come pista investigativa e mai come identificazione.

Agenti AI e conferma umana

Un agente pianifica ed esegue azioni con strumenti: ricerche, interrogazioni, script. Nel lavoro forense i rischi sono azioni non tracciate, errori che si propagano lungo la catena e modifiche involontarie ai dati. Le regole operative discendono dalla catena di custodia. L'agente lavora solo su copie di lavoro e con privilegi minimi; ogni azione viene registrata con input e output; ogni passo con effetti richiede la conferma di una persona; nessuna conclusione dell'agente entra in relazione senza verifica.

Prompt injection contenuta nei reperti

I reperti possono contenere testo ostile. Un documento sequestrato può includere istruzioni rivolte al modello che lo elabora, visibili o nascoste in testo bianco, commenti o metadati, con effetti che vanno dall'omissione di passaggi nel riassunto alla classificazione alterata, fino ad azioni indesiderate di un agente. La OWASP Top 10 for LLM Applications colloca la prompt injection al primo posto fra i rischi. Le contromisure sono la separazione fra dati e istruzioni, l'assenza di azioni

automatiche basate sugli output, la segnalazione dei contenuti sospetti e la verifica umana dei risultati.

Per il laboratorio. Gli studenti costruiscono una pipeline RAG sul corpus pubblico Enron, ne valutano le citazioni su un campione annotato e ne misurano la resistenza a documenti che contengono istruzioni ostili, documentando configurazione e risultati come in una relazione.

Lettere consigliate: P. Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”, NeurIPS 2020; NIST AI 600-1, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* (2024); OWASP, *Top 10 for LLM Applications*.

Capitolo S5.2. Multimedia forensics e perizia sui contenuti sintetici

Il capitolo forma il perito chiamato a pronunciarsi sull'autenticità di immagini, audio e video e sul funzionamento dei sistemi di intelligenza artificiale coinvolti in un incidente. Presuppone il modulo base sulla forensics dell'AI e ne porta le tecniche al livello richiesto da una perizia.

PRNU e attribuzione alla fotocamera

Le minime differenze di sensibilità fra i pixel di un sensore producono un'impronta di rumore stabile, detta PRNU (photo response non-uniformity), che si stima da una serie di immagini di riferimento scattate con il dispositivo. L'attribuzione consiste nel misurare la correlazione fra questa impronta e il residuo di rumore dell'immagine contestata. Il metodo ha limiti noti: ricompressione forte, ridimensionamento, ritaglio, stabilizzazione digitale ed elaborazione computazionale degli smartphone moderni degradano o alterano il segnale. Servono immagini di riferimento del dispositivo in esame, e il risultato va espresso con i tassi di errore misurati nelle condizioni del caso.

Compressione, contenitori e piattaforme

Un'immagine JPEG salvata due volte con parametri diversi conserva tracce statistiche della doppia compressione nei coefficienti della trasformata, indizio di riesportazione o di modifica. I file video MP4 e MOV hanno una struttura a blocchi, detti atomi o box, come ftyp, moov e mdat: ordine, contenuto e metadati di questi blocchi rivelano l'encoder e talvolta la sequenza delle elaborazioni. Le piattaforme social ricomprimono i contenuti e rimuovono i metadati con parametri propri, lasciando un'impronta che aiuta a ricostruire la catena di condivisione. ExifTool, MediaInfo e FFprobe sono gli strumenti di base per queste analisi.

Audio forense ed ENF

L'analisi di un file audio verifica la continuità della registrazione: discontinuità nel rumore di fondo, nello spettro o nei livelli possono indicare tagli e inserimenti. Le registrazioni effettuate vicino a dispositivi alimentati dalla rete elettrica possono catturarne la frequenza, nominalmente 50 Hz in Europa, con le sue micro-variazioni. Confrontate con registrazioni di riferimento della rete, queste variazioni permettono di datare la registrazione e di individuare porzioni rimosse. Il segnale può mancare, ad esempio con dispositivi a batteria o in ambienti esterni, e l'analisi richiede dati di riferimento relativi alla rete pertinente.

C2PA e watermark

Lo strumento c2patool legge e verifica i manifest C2PA: firma, identità del firmatario, azioni dichiarate sul contenuto. Una firma valida attesta chi ha firmato e che il contenuto non è cambiato

dopo la firma; la corrispondenza fra ciò che il contenuto rappresenta e la realtà resta da accertare con gli altri metodi. I watermark dei generatori sono spesso verificabili solo con gli strumenti del fornitore. L'assenza di manifest o di watermark non prova l'autenticità, perché i metadati si rimuovono con facilità e molti generatori, fra cui i modelli a pesi aperti, non applicano alcuna marcatura.

Forensics dei sistemi di AI

Quando un sistema di intelligenza artificiale è coinvolto in un incidente, le fonti di prova sono i log di inferenza con input, output e orari, le versioni del modello e dei componenti, la configurazione (prompt di sistema, filtri, strumenti collegati) e, per quanto disponibili, i dati di addestramento o di adattamento. La provenienza del modello si documenta con gli hash dei file dei pesi, le schede descrittive e i registri di distribuzione. Il profilo NIST AI 600-1 (2024) offre un quadro di riferimento per i rischi dell'AI generativa e per la documentazione da richiedere.

Il quadro normativo aggiornato

Gli obblighi di trasparenza dell'art. 50 dell'AI Act si applicano dal 2 agosto 2026. I fornitori di sistemi che generano contenuti sintetici devono marcarli in un formato leggibile dalla macchina e rilevabile; i deployer che usano un sistema per produrre deepfake devono dichiarare che il contenuto è stato generato o manipolato artificialmente. Il regolamento omnibus sull'AI, in vigore dal 27 luglio 2026, ha lasciato invariata la sostanza di questi obblighi e ha concesso ai fornitori di sistemi generativi già presenti sul mercato tempo fino al 2 dicembre 2026 per la marcatura; nel luglio 2026 la Commissione ha adottato le linee guida sull'art. 50. Per il perito l'assenza di marcatura resta un dato neutro, che può dipendere da un inadempimento, da un sistema non soggetto agli obblighi o da una rimozione.

La perizia per ipotesi alternative

La linea guida ENFSI sul reporting valutativo (2015) chiede di valutare le osservazioni rispetto ad almeno due proposizioni alternative, ad esempio che il video sia una registrazione originale non alterata oppure che sia stato generato o manipolato. Il peso dell'evidenza si esprime come rapporto di verosimiglianza, cioè la probabilità delle osservazioni sotto la prima ipotesi divisa per quella sotto la seconda, eventualmente tradotto in una scala verbale. Il perito valuta la probabilità delle osservazioni date le ipotesi; la probabilità delle ipotesi alla luce di tutte le prove spetta al giudice, e confondere i due piani è l'errore noto come fallacia del condizionale trasposto. Il lessico della compatibilità e la dichiarazione dei limiti completano la relazione.

Dataset di ricerca e generalizzazione

I dataset pubblici più usati sono FaceForensics++, che raccoglie video manipolati con diversi metodi ed è accessibile su richiesta, il Deepfake Detection Challenge (DFDC) e, per la voce, ASVspoof. Un rilevatore addestrato su un dataset perde spesso accuratezza su generatori, compressioni e

condizioni di ripresa diverse, ed è questo il problema della generalizzazione. In perizia il rilevatore automatico entra come uno degli indizi, con l'indicazione del dominio su cui è stato validato e delle prestazioni misurate.

Per il laboratorio. Su un video di prova tratto da dataset di ricerca gli studenti svolgono una perizia completa: acquisizione con hash, analisi del contenitore, indizi visivi e audio, verifica delle credenziali C2PA e relazione per ipotesi alternative, con il rapporto di verosimiglianza espresso in scala verbale.

Lettere consigliate: H. Farid, *Photo Forensics*, MIT Press; C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, C. Busch (a cura di), *Handbook of Digital Face Manipulation and Detection*, Springer, 2022 (open access); ENFSI, *Guideline for Evaluative Reporting in Forensic Science* (2015).

Glossario essenziale

Acquisizione. Fase del processo forense in cui il contenuto di una fonte di prova digitale viene copiato in forma verificabile, di norma con copia bit-a-bit e calcolo degli hash.

AFU e BFU. Stati di un dispositivo mobile acceso: dopo il primo sblocco successivo all'avvio (after first unlock), con molte chiavi di cifratura disponibili in memoria, o prima del primo sblocco (before first unlock), con la maggior parte dei dati dell'utente inaccessibile. Lo stato condiziona le tecniche di estrazione praticabili.

AI Act. Regolamento (UE) 2024/1689 sull'intelligenza artificiale; introduce obblighi graduati per rischio, fra cui la riconoscibilità dei contenuti sintetici.

Anomaly detection. Famiglia di tecniche che segnalano scostamenti da un comportamento atteso, utile quando non si sa in anticipo che cosa cercare.

Anti-forensics. Insieme delle tecniche volte a ostacolare l'indagine: wiping, timestomping, cifratura, steganografia, manipolazione dei log.

ATT&CK (MITRE). Base di conoscenza pubblica di tattiche e tecniche degli attaccanti, usata per descrivere in modo uniforme le fasi di un'intrusione e confrontare casi diversi.

Base rate (fallacia del). Errore di valutazione che ignora la prevalenza del fenomeno: con fenomeni rari, anche un rilevatore accurato produce in maggioranza falsi allarmi.

Beacon. Traffico periodico e regolare generato da un malware verso la propria infrastruttura di comando e controllo.

C2PA. Standard per le credenziali di provenienza dei contenuti: metadati firmati crittograficamente che documentano origine e trasformazioni di un file.

Calibrazione. Proprietà di un modello i cui score si comportano come probabilità: fra i casi con score 0,9, circa il 90 per cento è davvero positivo.

Carving. Recupero di file o frammenti da spazio non allocato, basato su firme e strutture note anziché sul file system.

Catena di custodia. Documentazione ininterrotta di chi ha avuto accesso al reperto, quando, e con quali operazioni, dall'acquisizione al dibattimento.

Concept drift. Degrado delle prestazioni di un modello dovuto al cambiamento del fenomeno rispetto ai dati di addestramento.

Copia forense. Duplicato bit-a-bit di un supporto, comprensivo di spazio non allocato e slack space, verificato tramite hash.

Daubert (criteri). Criteri di affidabilità del metodo scientifico elaborati dalla giurisprudenza statunitense: verificabilità, revisione paritaria, tasso d'errore noto, standard, accettazione della comunità scientifica.

Deepfake. Contenuto audio, video o fotografico manipolato o generato da modelli di apprendimento automatico in modo da apparire autentico.

EDR (Endpoint Detection and Response). Software installato sui sistemi che registra e analizza processi, connessioni e modifiche; la sua telemetria è una fonte di prova, conservata dal fornitore per periodi limitati.

e-Evidence (ordini europei). Strumenti introdotti dal Reg. (UE) 2023/1543, applicabile dal 18 agosto 2026: ordine europeo di produzione e ordine europeo di conservazione, rivolti direttamente ai fornitori di servizi stabiliti in un altro Stato membro.

ELA (Error Level Analysis). Tecnica esplorativa che evidenzia regioni di un'immagine con livelli di ricompressione incoerenti; nota per la facilità di sovrainterpretazione.

Embedding. Rappresentazione di un contenuto come vettore numerico in uno spazio in cui la vicinanza riflette la somiglianza di significato.

ENF (Electric Network Frequency). Frequenza della rete elettrica, nominalmente 50 Hz in Europa, le cui micro-variazioni possono restare impresse nelle registrazioni audio e consentirne la datazione per confronto con dati di riferimento.

Entropia. Misura del disordine statistico di una sequenza di byte; valori massimi caratterizzano contenuti cifrati o compressi.

Equalized odds. Metrica di fairness che richiede tassi di falso positivo e falso negativo simili fra gruppi.

Event log. Registri degli eventi di Windows (ad esempio Security e System), conservati in file .evtx; documentano accessi, creazione di processi, installazione di servizi e altre attività rilevanti.

Fairness. Equità di un sistema automatico, misurata confrontando le prestazioni sui gruppi su cui il sistema opera.

Falso negativo / falso positivo. Rispettivamente, elemento rilevante non segnalato ed elemento irrilevante segnalato; i loro costi investigativi sono asimmetrici.

Feature. Grandezza numerica estratta da un reperto per alimentare un modello; la sua definizione e documentazione sono parte della riproducibilità.

Furto di token di sessione. Sottrazione dei cookie o dei token emessi dopo l'autenticazione, che consente all'attaccante di usare la sessione senza conoscere la password e senza ripetere l'autenticazione a più fattori.

Hash (funzione). Funzione che produce un'impronta a lunghezza fissa del contenuto; identico il contenuto, identica l'impronta. Standard attuale SHA-256.

Hash percettivo. Impronta che sintetizza l'aspetto visivo di un contenuto: immagini simili producono impronte vicine (PhotoDNA, PDQ).

Live forensics. Acquisizione e analisi condotte su un sistema acceso, necessarie per i dati volatili; modificano lo stato del sistema e vanno motivate e verbalizzate.

LLM. Large language model: rete neurale addestrata su grandi corpora testuali, capace di generare linguaggio; può produrre contenuti plausibili ma falsi.

Magic bytes. Firma d'intestazione che identifica il formato reale di un file, indipendentemente dall'estensione dichiarata.

Matrice di confusione. Tabella che incrocia esiti reali e predetti di un classificatore: veri e falsi positivi, veri e falsi negativi.

Movimento laterale. Insieme delle tecniche con cui un attaccante, ottenuto l'accesso a un sistema, si sposta verso altri sistemi della rete, spesso con credenziali legittime sottratte.

NER (Named Entity Recognition). Estrazione automatica di nomi di persona, luoghi, organizzazioni, date e importi da un testo.

NSRL. National Software Reference Library del NIST: insiemi di hash di file noti usati per escludere dal triage il software legittimo.

Overfitting. Adattamento eccessivo di un modello ai dati di addestramento, con prestazioni che non si trasferiscono a dati nuovi.

Persistenza. Meccanismi che permettono a un codice malevolo di sopravvivere ai riavvii, come chiavi di esecuzione automatica, servizi, attività pianificate o sottoscrizioni WMI.

Precision / Recall. Frazione di segnalazioni corrette sul totale delle segnalazioni; frazione di elementi rilevanti effettivamente trovati sul totale dei rilevanti.

PRNU (Photo Response Non-Uniformity). Impronta di rumore del sensore di una fotocamera, dovuta a minime differenze di sensibilità fra i pixel; consente di valutare se un'immagine provenga da un determinato dispositivo.

Prompt injection. Attacco che inserisce istruzioni ostili nei dati elaborati da un sistema basato su LLM per alterarne il comportamento.

RAG (Retrieval-Augmented Generation). Architettura che ancora le risposte di un modello linguistico a documenti recuperati e citati, rendendole verificabili.

Rapporto di verosimiglianza. Rapporto fra le probabilità delle osservazioni sotto due ipotesi alternative; esprime il peso dell'evidenza senza pronunciarsi sulla probabilità delle ipotesi, che spetta al giudice.

Regole Sigma. Formato aperto per descrivere rilevazioni sui log in modo indipendente dal prodotto, traducibile per diversi sistemi di analisi.

Ripetibilità. Possibilità di ripetere un accertamento senza alterare la fonte; la sua assenza attiva le garanzie dell'art. 360 c.p.p.

Shellbag. Artefatto del registro di Windows che documenta le cartelle aperte dall'utente, anche su supporti rimossi o cancellati.

Stilometria. Misura delle regolarità inconse della scrittura ai fini del confronto fra un testo contestato e campioni noti; produce compatibilità, non identificazioni.

Timeline. Ricostruzione cronologica unificata degli eventi di un sistema a partire da più fonti di timestamp.

Timestomping. Manipolazione deliberata dei timestamp di un file per ostacolare la ricostruzione cronologica.

Triage. Ordinamento e filtraggio dei reperti per priorità d'esame, oggi largamente assistito da strumenti automatici.

UEBA (User and Entity Behavior Analytics). Tecniche che apprendono il comportamento abituale di utenti e sistemi e segnalano gli scostamenti, ad esempio accessi da luoghi o in orari insoliti.

Volatilità (ordine di). Gerarchia delle fonti secondo la rapidità con cui i dati svaniscono (RFC 3227): registri e RAM prima, supporti e backup poi.

Watermarking. Marcatura impercettibile inserita in un contenuto alla generazione; prova la provenienza se presente, non dimostra nulla se assente.

Write blocker. Dispositivo o configurazione che impedisce ogni scrittura verso il supporto in acquisizione, a garanzia dell'integrità.

Bibliografia e riferimenti essenziali

Normativa

- Codice di procedura penale, in particolare artt. 244, 247, 352, 354 (ispezioni, perquisizioni e accertamenti urgenti sui sistemi informatici) e artt. 359 e 360 (consulenti tecnici e accertamenti non ripetibili).
- Legge 18 marzo 2008, n. 48, di ratifica della Convenzione del Consiglio d'Europa sulla criminalità informatica (Budapest, 2001).
- Regolamento (UE) 2016/679 (GDPR), in particolare i principi dell'art. 5 e la notifica delle violazioni ex art. 33.
- Regolamento (UE) 2024/1689 (AI Act), in particolare gli obblighi di trasparenza sui contenuti sintetici.

Standard e linee guida

- ISO/IEC 27037:2012, linee guida per identificazione, raccolta, acquisizione e conservazione della prova digitale.
- ISO/IEC 27042:2015, linee guida per l'analisi e l'interpretazione della prova digitale.
- ISO/IEC 17025:2017, requisiti generali per la competenza dei laboratori di prova.
- NIST SP 800-86, "Guide to Integrating Forensic Techniques into Incident Response".
- RFC 3227, "Guidelines for Evidence Collection and Archiving" (ordine di volatilità).
- ENFSI, "Best Practice Manual for the Forensic Examination of Digital Technology".

Manuali di riferimento

- E. Casey, "Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet", 3a ed., Academic Press.
- B. Carrier, "File System Forensic Analysis", Addison-Wesley.
- M. H. Ligh, A. Case, J. Levy, A. Walters, "The Art of Memory Forensics", Wiley.
- H. Farid, "Photo Forensics", MIT Press.
- D. Jurafsky, J. H. Martin, "Speech and Language Processing", 3a ed. (bozza disponibile online).
- C. Molnar, "Interpretable Machine Learning" (disponibile online).

Studi e documenti tecnici

- P. Grother, M. Ngan, K. Hanaoka, "Face Recognition Vendor Test, Part 3: Demographic Effects", NISTIR 8280.
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).

- OWASP, “Top 10 for Large Language Model Applications”.
- Coalition for Content Provenance and Authenticity (C2PA), specifiche tecniche.

Strumenti open source citati nel corso

- Volatility 3 (memory forensics).
- Plaso / Timesketch (super-timeline e analisi collaborativa).
- Zeek (analisi del traffico di rete).
- ALEAPP e ILEAPP (artefatti Android e iOS).
- FTK Imager (acquisizione; freeware).
- Autopsy / The Sleuth Kit (analisi di file system).
- scikit-learn (apprendimento automatico e metriche).

Riferimenti per i moduli avanzati

- NIST SP 800-61 Rev. 3, “Incident Response Recommendations and Considerations for Cybersecurity Risk Management: A CSF 2.0 Community Profile” (2025).
- J. Luttgens, M. Pepe, K. Mandia, “Incident Response & Computer Forensics”, 3a ed., McGraw-Hill.
- MITRE ATT&CK, base di conoscenza di tattiche e tecniche degli attaccanti (matrice Enterprise).
- NIST IR 8006, “NIST Cloud Computing Forensic Science Challenges” (2020).
- Regolamento (UE) 2023/1543 sugli ordini europei di produzione e di conservazione di prove elettroniche e Direttiva (UE) 2023/1544 sui rappresentanti legali.
- Secondo Protocollo addizionale alla Convenzione sulla criminalità informatica, sul rafforzamento della cooperazione e della divulgazione di prove elettroniche (2022).
- NIST SP 800-101 Rev. 1, “Guidelines on Mobile Device Forensics” (2014).
- R. Tamma et al., “Practical Mobile Forensics”, 4a ed., Packt.
- P. Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”, NeurIPS 2020.
- NIST AI 600-1, “Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile” (2024).
- OWASP, “Top 10 for LLM Applications”.
- C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, C. Busch (a cura di), “Handbook of Digital Face Manipulation and Detection”, Springer, 2022 (open access).
- ENFSI, “Guideline for Evaluative Reporting in Forensic Science” (2015).

Dataset didattici pubblici

- NIST CFReDS, Computer Forensic Reference Data Sets.
- Digital Corpora (immagini disco, memoria e traffico a scopo didattico).
- Corpus email Enron (analisi di comunicazioni e grafi relazionali).

Nel corso si impiegano esclusivamente dataset pubblici o generati sinteticamente; lo stesso criterio di tracciabilità delle fonti è richiesto in ogni esercitazione e nella relazione finale.