AI per la Digital Forensics: affidabilità critica



L’AI per la Digital Forensics rappresenta una svolta cruciale nell’evoluzione delle indagini digitali, introducendo strumenti analitici potenti ma anche nuove vulnerabilità e complessità normative. Questo articolo analizza l’impatto dell’intelligenza artificiale sui processi forensi, evidenziando le opportunità di miglioramento in termini di accuratezza, velocità e automazione, ma anche le criticità legate a bias algoritmici, attacchi adversariali e mancanza di trasparenza.

L’integrazione dell’intelligenza artificiale nei processi di digital forensics rappresenta una trasformazione paradigmatica per le indagini digitali, introducendo capacità analitiche avanzate ma anche **criticità sistemiche immediate** che compromettono robustezza e affidabilità investigativa. Le evidenze scientifiche consolidate dimostrano che gli strumenti AI forensi, pur raggiungendo **performance del 98,7% in classificazione malware**, presentano vulnerabilità fondamentali ad attacchi adversariali e bias algoritmici che richiedono framework normativi rigorosi e metodologie di validazione specifiche prima del deployment operativo.

Architetture algoritmiche avanzate per evidenze digitali

L’ecosistema tecnico dell’AI forense si basa su architetture neurali specializzate con performance quantificate in contesti operativi. Le **Convolutionaal Neural Networks (CNN)** per classificazione malware attraverso visualizzazione di immagini binarie raggiungono accuratezza del 98,7% con validazione cross-fold k=10, mentre i modelli **Long Short-Term Memory**

(LSTM) per analisi malware Android ottengono F1-score di 0,967 utilizzando API calls dinamiche.

Gli approcci ensemble combinano CNN e Multi-Layer Perceptron per memory forensics con accuratezza del 99,2% su dataset Cuckoo sandbox, dimostrando resistenza a malware sofisticati memory-resident. Le **Graph Neural Networks (GNN)** abilitano l'analisi di correlazioni eventi complesse attraverso rappresentazioni grafiche di digital artifacts, mentre i modelli **1D-CNN** per classificazione traffico di rete superano significativamente le performance di sistemi basati su conversione immagini.

Le metodologie di feature extraction includono **Histogram of Oriented Gradients (HOG)** per facial recognition forensics, **Hash Indexed Sparse Distributed Representation (HI-SDR)** per anomaly detection IoT, e tecniche **Term Frequency-Inverse Document Frequency (TF-IDF)** per text analysis. I framework di validazione tecnica implementano **Adversarial Testing Framework (ATF)** per test robustezza modelli contro attacchi black-box, con metriche avanzate come Cohen's Kappa e Likelihood Ratio per DNA probabilistic genotyping.

Vulnerabilità sistemiche e attacchi adversariali

L'analisi scientifica rivela **criticità immediate** che compromettono l'affidabilità operativa dei sistemi AI forensi. Gli **attacchi adversariali transferable** basati su GAN ingannano efficacemente i sistemi CNN forensi con perturbazioni minime, dimostrando che attacchi progettati per un sistema specifico mantengono efficacia contro altri sistemi senza conoscenza esplicita degli algoritmi target.

I **deepfake detectors** stato-dell'arte falliscono contro modelli generativi personalizzati dall'utente e attacchi adversariali basati su foundation models. La ricerca evidenzia che in

contesti realistici (social media, post-processing), gli attacchi adversariali riducono l'accuratezza di detection fino al punto critico dove "i rischi derivanti dall'affidarsi ai detector superano i benefici".

Le **tecniche steganografiche avanzate** sfruttano catene di agenti AI multimodali per embedding nel dominio linguistico, rendendo detection tradizionale inefficace. I sistemi **Spread Spectrum Image Steganography (SSIS)** richiedono fine-tuning specifico con conseguenti trade-off di generalizzazione, mentre l'audio synchronous steganography in piattaforme video rappresenta una sfida emergente per la detection.

Quadro giuridico e procedurali per l'utilizzo dell'AI per la Digital Forensics

L'ammissibilità legale di prove generate o analizzate da AI presenta **complessità giuridiche significative** nel sistema italiano ed europeo. I precedenti giurisprudenziali chiave includono **Cass. Pen. n. 11905/2016** sulla necessità di best practices per attendibilità probatoria, e **Cons. Stato n. 881/2020** che stabilisce principi fondamentali per utilizzo AI: trasparenza rafforzata, controllo umano significativo, non discriminazione algoritmica e conoscibilità dell'algoritmo.

La **chain of custody** per evidenze digitali processate da AI richiede documentazione aggiuntiva di algoritmi utilizzati, dataset di training e parametri, introducendo complessità procedurali che possono compromettere l'ammissibilità. Il **Sistema SARI** della Polizia di Stato ha ricevuto parere negativo dal Garante Privacy per mancanza di base giuridica, evidenziando le sfide normative per sistemi di riconoscimento biometrico.

L'**AI Act europeo** (Regolamento UE 2024/1689) classifica gli

strumenti investigativi come sistemi ad alto rischio, richiedendo conformità entro agosto 2026. L'art. 5 vieta pratiche AI specifiche incluso il riconoscimento biometrico real-time, con eccezioni limitate per law enforcement. La **Direttiva 2016/680/UE** proibisce decisioni basate unicamente su trattamento automatizzato, richiedendo sempre controllo umano significativo.

Framework normativi e standard di certificazione

Il **NIST AI Risk Management Framework (AI RMF)** fornisce struttura volontaria per incorporare considerazioni di trustworthiness nell'uso dell'AI, con aggiornamenti specifici per Generative AI. Il **Computer Forensics Tool Testing (CFTT) Program** stabilisce metodologie di testing per strumenti forensi, con espansione prevista per strumenti AI-enabled.

ENISA ha sviluppato un framework multilayer per sicurezza AI con tre livelli (foundations, AI-specific, sector-specific), supportando l'implementazione dell'AI Act. Gli **standard ISO/IEC 27037-27043** definiscono linee guida per identificazione, acquisizione e analisi di evidenze digitali, mentre le certificazioni professionali **IACIS (CFCE)** e **ISFCE (CCE)** stabiliscono competenze richieste per investigatori.

Europol ha pubblicato l'AI and Policing Observatory Report evidenziando benefici e sfide AI per law enforcement, con framework di compliance per protezione dati e AI Act. Le **linee guida INTERPOL** per laboratori forensi digitali stabiliscono procedure per gestione evidenze elettroniche con metodologie tecnicamente robuste per ammissibilità legale.

Bias algoritmici e problematiche di fairness

I **bias sistemici** documentati in automation tools per digital forensics hanno influenzato casi civili e penali reali, con

particolare focus su bias nei dati di training, bias algoritmici nella progettazione e bias interpretativi nell'applicazione. L'**impostor bias** rappresenta una nuova categoria di bias cognitivo nella detection di deepfake, dove investigatori dubitano sistematicamente dell'autenticità di media reali.

La **discriminazione algoritmicamente mediata** perpetua bias esistenti nel riconoscimento facciale con performance differenziate per etnia, profilazione comportamentale discriminatoria e analisi linguistica con bias culturali. I framework di mitigazione richiedono audit sistematici per bias detection, diversificazione dataset training e integrazione di fairness metrics.

Explainable AI e interpretabilità giuridica

Il **dilemma black-box** degli algoritmi AI impedisce la spiegazione delle decisioni forensi, creando "present level of skepticism expressed by courts, legal practitioners, and the general public". I sistemi AI forensi faticano a soddisfare i requisiti **Daubert Standard** per ammissibilità legale per mancanza di trasparenza e interpretabilità.

I **framework XAI emergenti** per digital forensics integrano visualizzazioni interpretative, spiegazioni causali e metriche di confidence trasparenti. Il **computational reliabilism** propone valutazione affidabilità attraverso indicatori tecnici (validazione algoritmica), scientifici (consensus metodologico) e sociali (acceptance comunitario).

Implementazioni operative e casi di

studio

Le **implementazioni pratiche** documentano risultati operativi significativi: l'**FBI Sarasota County Sheriff's Office** ha utilizzato analisi AI per identificazione impronte digitali attraverso immagini smartphone, mentre l'**Oakland Police Department** ha ottenuto riduzione del 50% delle sparatorie utilizzando sistemi AI investigativi.

Il **Magnet.AI** integrato in AXIOM ha ridotto del 70% i tempi di investigazione in casi di sfruttamento minorile ed è utilizzato da FBI ed Europol per decrypt comunicazioni criminali. L'**University of Huddersfield** ha implementato sistemi DDM per classificazione età con riduzione del 36% dei tempi di processing e 26% per scoperta contenuti IIOC.

La **Polizia Scientifica italiana** ha sviluppato la IV Divisione con sistemi AI per applicazioni multimediali, mentre la **Guardia di Finanza** ha implementato metodologie CFDA per acquisizione evidenze digitali in indagini fiscali.

Competenze professionali e training requirements

Le **competenze richieste** includono Digital Forensics Examiner con competenze AI, Machine Learning Specialist per forensics, Computer Forensics Data Analyst (CFDA) e AI Ethics Specialist per law enforcement. Le **certificazioni identificate** comprendono Cellebrite Certified Physical Analyst (CCPA), Magnet Forensics certified examiner e specialized AI forensics training programs.

I **programmi formativi** includono Master in Forensic Psychology con AI components, university programs in AI-enabled digital forensics e continuous professional development in AI technologies. Le **best practices** enfatizzano Human-in-the-Loop Approach, standardized methodologies, continuous training ed

ethical frameworks.

Raccomandazioni immediate per deployment sicuro

L’implementazione responsabile richiede **adversarial training** con dataset rappresentativi, **explainable AI integration** obbligatorio con standardizzazione explanation frameworks, e **bias mitigation protocols** attraverso audit sistematici e diversificazione dataset training.

La **standardizzazione di validazione** deve includere sviluppo standard internazionali, cross-validation metodologie robuste e continuous performance monitoring. L’**approccio implementativo** deve essere graduale con validation processes costanti, legal compliance specifica e cross-agency collaboration per condivisione expertise.

Conclusioni

L’AI per digital forensics presenta potenzialità trasformative significative ma richiede approcci multidisciplinari immediati per addressare vulnerabilità sistemiche. Le **criticità immediate** identificate – attacchi adversariali, bias algoritmici, interpretabilità limitata e gaps normativi – rappresentano sfide che richiedono risoluzione prima del deployment operativo su larga scala.

La **convergenza** tra capacità tecniche avanzate e readiness operativa necessita investimenti urgenti in robustezza, fairness e interpretabilità per garantire applicazioni forensi affidabili ed eticamente sostenibili. Il **gap critico** tra innovation potential e practical deployment evidenzia la necessità di framework integrati che bilancino efficacia investigativa con tutela dei diritti fondamentali e requisiti di trasparenza giuridica.

L’evoluzione futura richiederà **standardizzazione internazionale**, continuous validation methodologies e sviluppo di competenze specialistiche per realizzare il potenziale trasformativo dell’AI mantenendo rigorous scientific standards e legal compliance.

Fonti:

Commissione Europea. (2024). *Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull’intelligenza artificiale (AI Act)*.

ENISA – European Union Agency for Cybersecurity. (2023). *Multilayer Framework for Good Cybersecurity Practices for AI.*

Europol. (2024). *AI and Policing Observatory Report: Benefits and Challenges of AI for Law Enforcement.*

FBI – Federal Bureau of Investigation. (2024). *Forensic Spotlight: Digital Forensic Examination – A Case Study*. FBI Law Enforcement Bulletin.

FBI – Federal Bureau of Investigation. (2023). *Conducting a Digital Forensics Capability Study*. FBI Law Enforcement Bulletin.

INTERPOL. (2024). *Digital Forensics Guidelines for Law Enforcement.*

NIST – National Institute of Standards and Technology. (2023). *AI Risk Management Framework (AI RMF 1.0)*.

NIST – National Institute of Standards and Technology. (2024). *Computer Forensics Tool Testing Program (CFTT).*

NIST – National Institute of Standards and Technology. (2023). *Digital Forensics Research and Standards.*

NIST – National Institute of Standards and Technology. (2024). *Answering the Call: How NIST Helps Crime Labs Recover Evidence from Mobile Devices.*

Polizia di Stato – Direzione Centrale per la Polizia Scientifica. (2024). *Innovazione Tecnologica e Sicurezza Cibernetica.*

UNODC – United Nations Office on Drugs and Crime. (2023). *Standards and Best Practices for Digital Forensics.*

U.S. Government Accountability Office. (2021). *Forensic Technology: Algorithms Strengthen Forensic Analysis, but Several Factors Can Affect Outcomes*. Report GAO-21-435SP.

Apruzzese, G., et al. (2023). "A comprehensive analysis of the role of artificial intelligence and machine learning in modern digital forensics and incident response." *Forensic Science International: Digital Investigation*, 44, 301544.

Benvenuto, C.J., et al. (2024). "Artificial Intelligence in Digital Forensics: Enhancing Forework with Machine Learning." *International Journal of Digital Forensics & Incident Response*, 15(2), 112-128.

Caviglione, L., et al. (2025). "Steganography Beyond Space-Time With Chain of Multimodal AI Agents." *arXiv preprint arXiv:2502.18547.*

Ferrara, M., et al. (2022). "Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models." *Forensic Science International: Digital Investigation*, 41, 301378.

Martini, B., et al. (2014). "Data Mining Methods Applied to a Digital Forensics Task for Supervised Machine Learning." In *Cyberspace Safety and Security* (pp. 270-284). Springer.

Condividi sui Social Network:
https://www.ictsecuritymagazine.com/articoli/ai-per-la-digital-forensics/