DIGITAL FORENSICS LAB
UNIVERSITY OF PAVIA

Intelligenze artificiali nella digital forensics

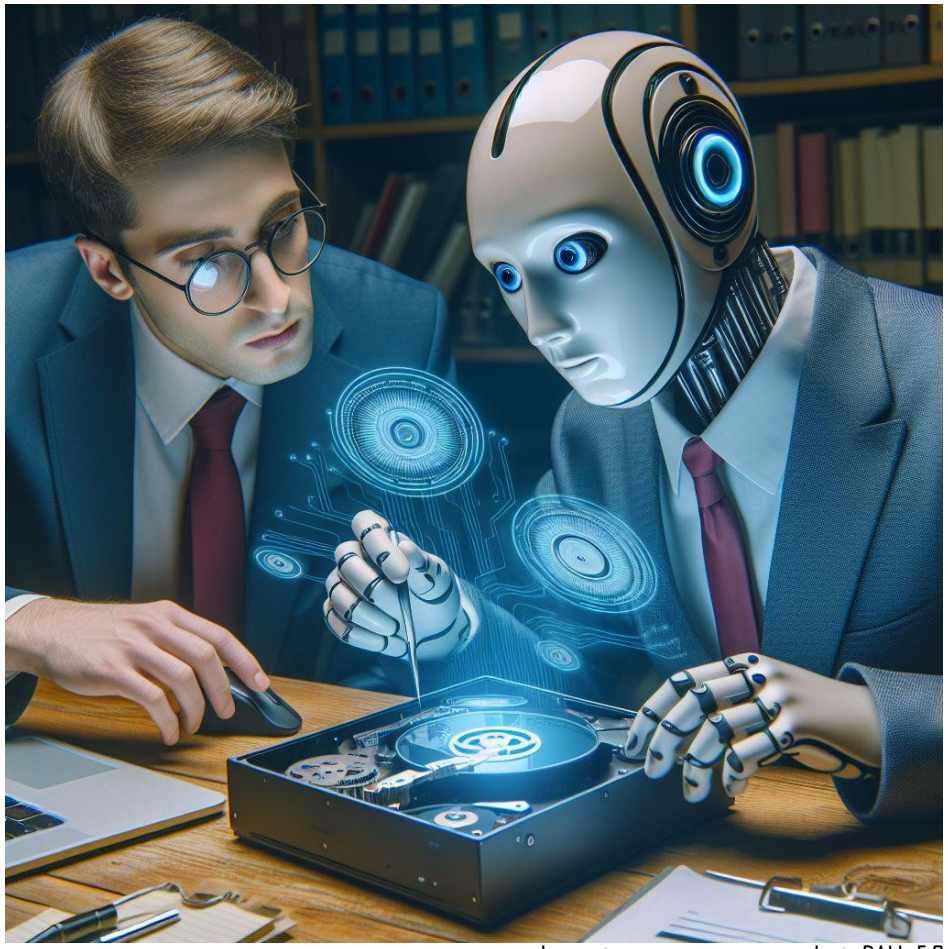


Immagine generata con tecnologia DALL-E 3

Webinar ISACA, 8 marzo 2024

Chi vi parla

Davide ‘Rebus’ Gabrini

- Professore a contratto in Informatica e Sicurezza Informatica presso il Dipartimento di Giurisprudenza dell’Università degli Studi di Pavia, A.A. 2023/2024
- Collaboratore del Laboratorio di Informatica Forense dell’Università degli Studi di Pavia, afferente al Laboratorio Nazionale di Cyber Security
- Membro del Comitato Scientifico dell’Area di Diritto e Informatica del Centro Ricerca e Didattica Universitaria del Collegio Ghislieri di Pavia
- Docente di sicurezza informatica e digital forensics per aziende, Pubbliche Amministrazioni e Università
- Certificazioni CIFI, ACE, AME
- Contributor di Tsurugi Linux, P.M. di Bento
- Socio fondatore di IHF e Nutria LUG
- Curatore della newsletter Rebus’ Digest

Artificial Intelligence for law enforcement

Artificial intelligence and robotics for law enforcement

- Dal 2018 l'Istituto interregionale delle Nazioni Unite per la ricerca sul crimine e la giustizia (UNICRI) e Interpol promuovono il "*Global Meeting on the Opportunities and Risks of Artificial Intelligence and Robotics for Law Enforcement*"
- Ai meeting sono stati presentati e discussi i contributi che IA e robotica possono dare alle attività di polizia e si sono esaminati casi d'uso a vari stadi di sviluppo da parte delle forze dell'ordine nazionali

United Nations Interregional Crime and Justice Research Institute

CENTRE FOR ARTIFICIAL INTELLIGENCE AND ROBOTICS

INTERPOL

Artificial intelligence and robotics for law enforcement

- Tra i più interessanti impieghi considerati ci sono:
  - Strumenti avanzati per l'autopsia virtuale, che aiutino a determinare le cause del decesso
  - Sistemi robotici autonomi di pattugliamento
  - Sistemi predittivi riguardo a luogo e tipologia dei reati che potrebbero compiersi
  - Software di visione artificiale per identificare auto sospette o rubate
  - Strumenti analitici per immagini, video e audio
  - Sistemi di riconoscimento avanzato dei volti
  - Strumenti per identificare i bambini sfruttati o a rischio
  - Strumenti di rilevamento comportamentale, per individuare taccheggiatori
  - Strumenti totalmente autonomi per identificare truffe online
  - Sistemi di realtà aumentata per le forze di polizia

Artificial intelligence and robotics for law enforcement

- Tra i possibili usi malevoli sono stati considerati:
  - cyber-attachi condotti da IA
    - spear phishing, exploiting automatizzati, ddos…
  - Attacchi di natura politica
    - proliferazione di fake news, propaganda, disinformazione, deepfake…
  - Attacchi cinetici
    - con l'uso di droni impiegati per colpire persone
  - Una IA potrebbe essere utilizzata anche per contrastare o sovvertire un altro sistema di IA (*adversarial models*)
    - ad esempio per avvelenare il dataset

Artificial intelligence: an overview of state initiatives

- Rapporto FutureGrasp, LLC - luglio 2019

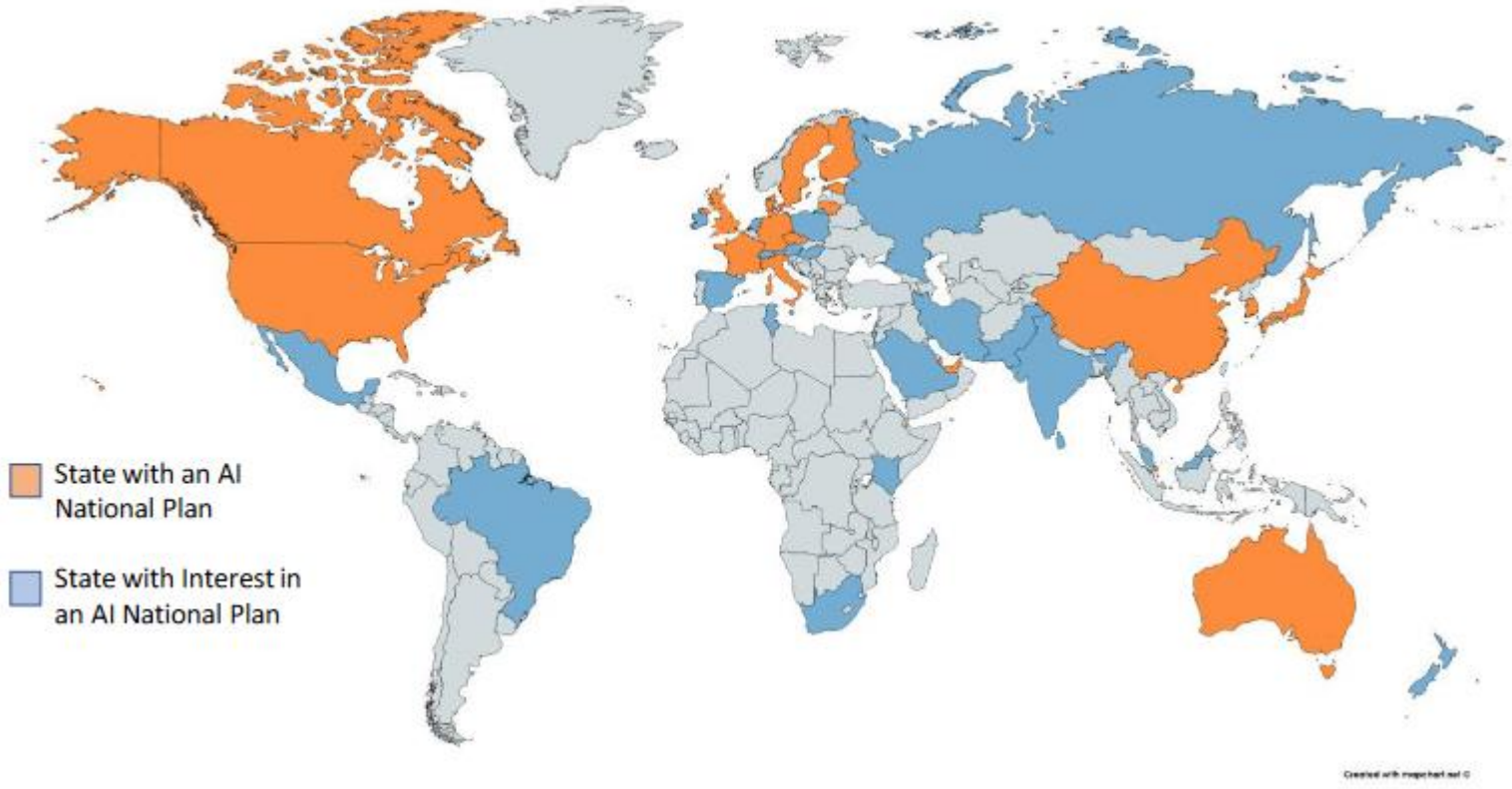


Figure 1. States with or having demonstrated interest in an AI national strategy or plan.

Toolkit for Responsible AI Innovation in Law Enforcement

- Col finanziamento dell'Unione Europea, UNICRI e Interpol hanno predisposto il *toolkit per l'innovazione responsabile dell'IA nelle forze dell'ordine*: un insieme di risorse pensate su misura per supportare le forze dell'ordine nell'istituzionalizzare e integrare l'uso di sistemi IA nel loro lavoro.
- Le risorse contenute nel AI Toolkit possono aiutare le forze dell'ordine nel miglioramento della comprensione, sviluppo, acquisizione e utilizzo dei sistemi di IA in un modo che sia allineato con le leggi sui diritti umani, l'etica e i principi di polizia.
- Sebbene destinato all'uso da parte della comunità delle forze dell'ordine, l'IA Toolkit è anche pertinente e accessibile a una vasta gamma di parti interessate, tra cui industria, accademia, società civile e pubblico generale.

RESPONSIBLE AI INNOVATION IN LAW ENFORCEMENT

AI Toolkit

PRIMARY GUIDANCE DOCUMENTS	PRACTICAL TOOLS	SUPPORTING DOCUMENTS
Introduction to Responsible AI Innovation	Organizational Readiness Assessment Questionnaire	Technical Reference Book
Principles for Responsible AI Innovation	Risk Assessment Questionnaire	
Organizational Roadmap	Responsible AI Innovation in Action Workbook	

Europol: The impact of LLM on Law Enforcement

- A fine marzo 2023 Europol Innovation Lab ha pubblicato un documento dal titolo "*ChatGPT: The impact of Large Language Models on Law Enforcement*" in cui esplora i possibili abusi criminali di LLMs come ChatGPT, così come considera i possibili usi per assistere gli investigatori
- Gli esperti che hanno partecipato ai workshop rappresentano l'intero spettro delle competenze di Europol, compresa l'analisi operativa, la criminalità organizzata e i *serious crimes*, la criminalità informatica e l'antiterrorismo

EUROPOL

INNOVATION LAB

Safeguards, prompt engineering, jailbreaks

- Le principali limitazioni di ChatGPT sono auto-imposte. Come parte delle policy di moderazione dei contenuti, ChatGPT non risponde a domande che sono state classificate come dannose o pregiudizievoli
- Un simile meccanismo di sicurezza deve essere continuamente aggiornato, e comunque può essere sovvertito con il giusto *prompt engineering*
- Alcune di queste scappatoie sono state chiuse, ma data la complessità del modello non mancano nuove soluzioni, scoperte dai ricercatori e dai *threat actors*.
- Nel caso di ChatGPT, i più comuni *workaround* includono:
  - *Prompt creation* (fornire la risposta e chiedere a ChatGPT di elaborare la corrispondente richiesta);
  - Chiedere a ChatGPT di fornire la risposta in forma di codice o simulando di essere un personaggio immaginario che discute un argomento;
  - Sostituire le parole chiave moderate e modificare successivamente il contesto;
  - Trasferimenti di stile/opinione (suggerire una risposta obiettiva e successivamente cambiare lo stile/prospettiva in cui è stata scritta);
  - Creare esempi fittizi facilmente trasferibili a eventi reali (ad esempio evitando nomi, nazionalità, ecc.)

Criminal use cases

- Gli esperti di Europol hanno individuato una vasta gamma di casi di uso criminale
- Se un potenziale criminale non sa nulla di una particolare area criminale, ChatGPT può velocizzare significativamente il processo di ricerca offrendo informazioni chiave che possono poi essere ulteriormente esplorate nelle fasi successive.
  - Si può accedere a informazioni su come entrare in una casa, sul terrorismo o sul cybercrime, o sugli abusi sessuali su minori.
- Le informazioni erano già pubbliche, ma così diventano più immediatamente fruibili

Fraud, impersonation, and social engineering

- Un LLM è uno strumento eccellente a scopo di phishing
- Molte truffe dozzinali di phishing sono facilmente rilevabili per grossolani errori grammaticali e sintattici, ma ora è più semplice impersonare un'organizzazione o una persona in modo molto realistico, anche con una conoscenza appena basilare dell'inglese
- Il contenuto delle mail di phishing può adattarsi facilmente alle necessità del criminale, che si tratti di proposte di investimento fraudolente, di BEC o CEO fraud
- ChatGPT offre ai criminali nuove opportunità per i crimini che utilizzano il social engineering, dando la possibilità di rispondere ai messaggi con pertinenza e adottando specifici stili di scrittura
  - IA generative sono di supporto alla creazione di *fake identity*: foto e clip video utili a registrare e popolare profili
- Con l'uso dei LLM, campagne di phishing e frodi online possono essere create più rapidamente, con più credibilità e su scala notevolmente aumentata

Harder, Better, Faster, Stronger

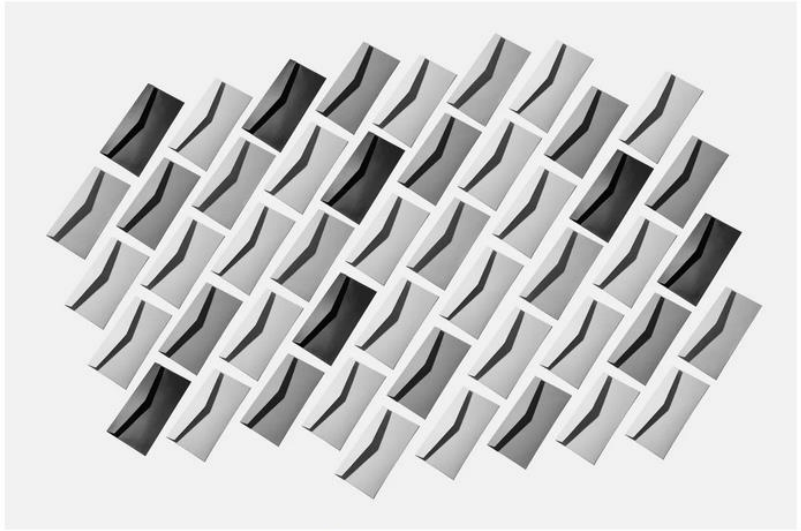
- A Black Hat 2021 un team della *Government Technology Agency* di Singapore ha presentato i risultati di un esperimento durante il quale sono state inviate email di spear phishing a 200 destinatari interni.
  - Alcuni messaggi erano creati dai componenti dell'agenzia, altri generati da GPT-3
  - Entrambi i messaggi contenevano link innocui ma traccianti
- I ricercatori sono rimasti sorpresi nello scoprire che un numero maggiore di persone ha cliccato sui link nei messaggi generati dall'IA rispetto a quelli scritti dall'uomo; lo scarto tra i due gruppi è stato piuttosto significativo.

≡ WIRED SIGN IN

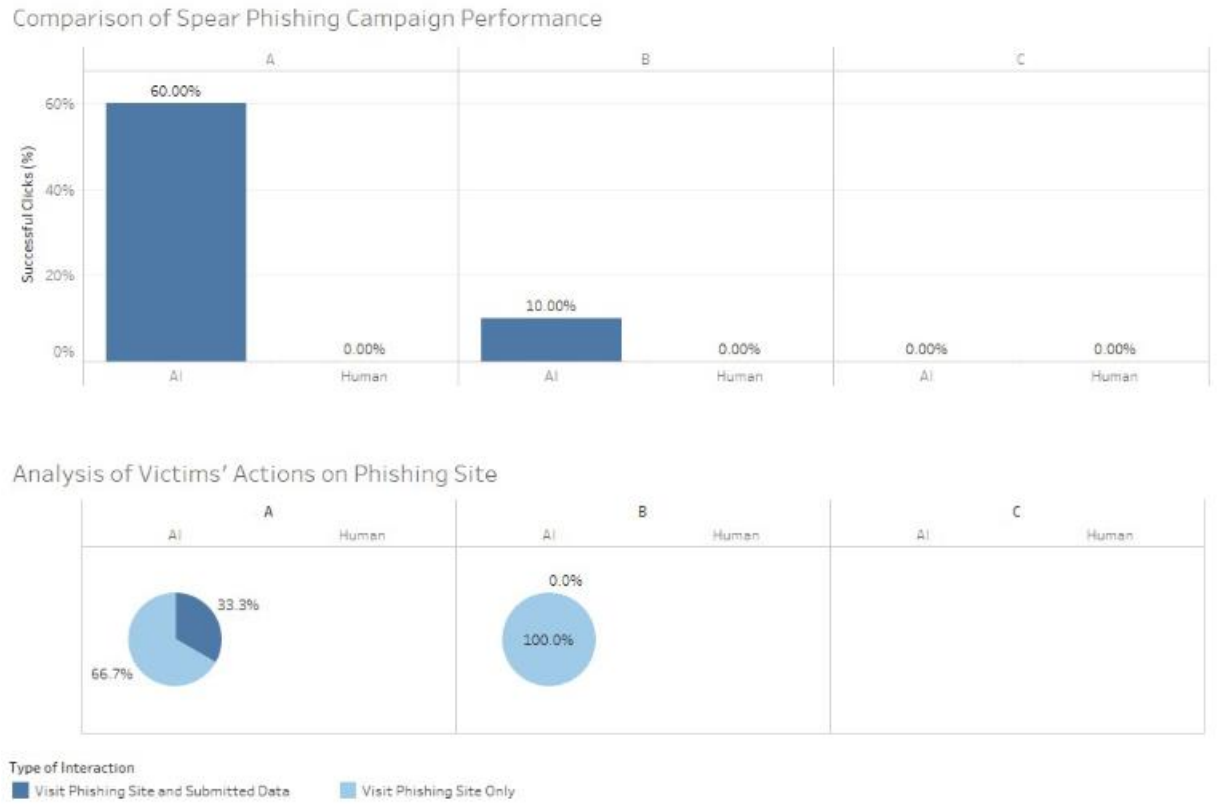
LILY HAY NEWMAN SECURITY AUG 7, 2021 7:00 AM

AI Wrote Better Phishing Emails Than Humans in a Recent Test

Researchers found that tools like OpenAI's GPT-3 helped craft devilishly effective spearphishing messages.



More people clicked the links in the AI-generated messages than the human-written ones—by a significant margin. PHOTOGRAPH: MIRAGEC/GETTY IMAGES



(you don't really) KYC

OPENAI JOURNEY

Guides Models

KYC Nightmare: Stable Diffusion User's Fake Image Raises Concerns For User Verification

By Ahfaz Ahmed • January 9, 2024

Deepfake for the masses

- Da Barack Obama a Michele Mirabella in 6 anni
  - dalle applicazioni belliche alle batterie di pentole…

IRCCS
HUMANITAS
RESEARCH HOSPITAL

Home » News » La trasmissione RAI Elisir denuncia in diretta un'intervista fake ad un medico di Humanitas, creata con l'Intelligenza Artificiale

La trasmissione RAI Elisir denuncia in diretta un'intervista fake ad un medico di Humanitas, creata con l'Intelligenza Artificiale

Pubblicato il Dicembre 14, 2023

WIRED

EVENTI NEWSLETTER MAGAZINE

DIEGO BARBERA LA TRUFFA 18.01.2024

Taylor Swift è stata trasformata in una venditrice di pentole dall'AI

Numerosi video comparsi su TikTok e Facebook ritraevano la cantante offrire costosi prodotti Le Creuset ai fan

I got some exciting news to share! I have teamed up with **Lé Creuset** for another epic gift-away - But this time we're giving away FREE 20 piece cookware sets!

Propaganda, fake news e manipolazione

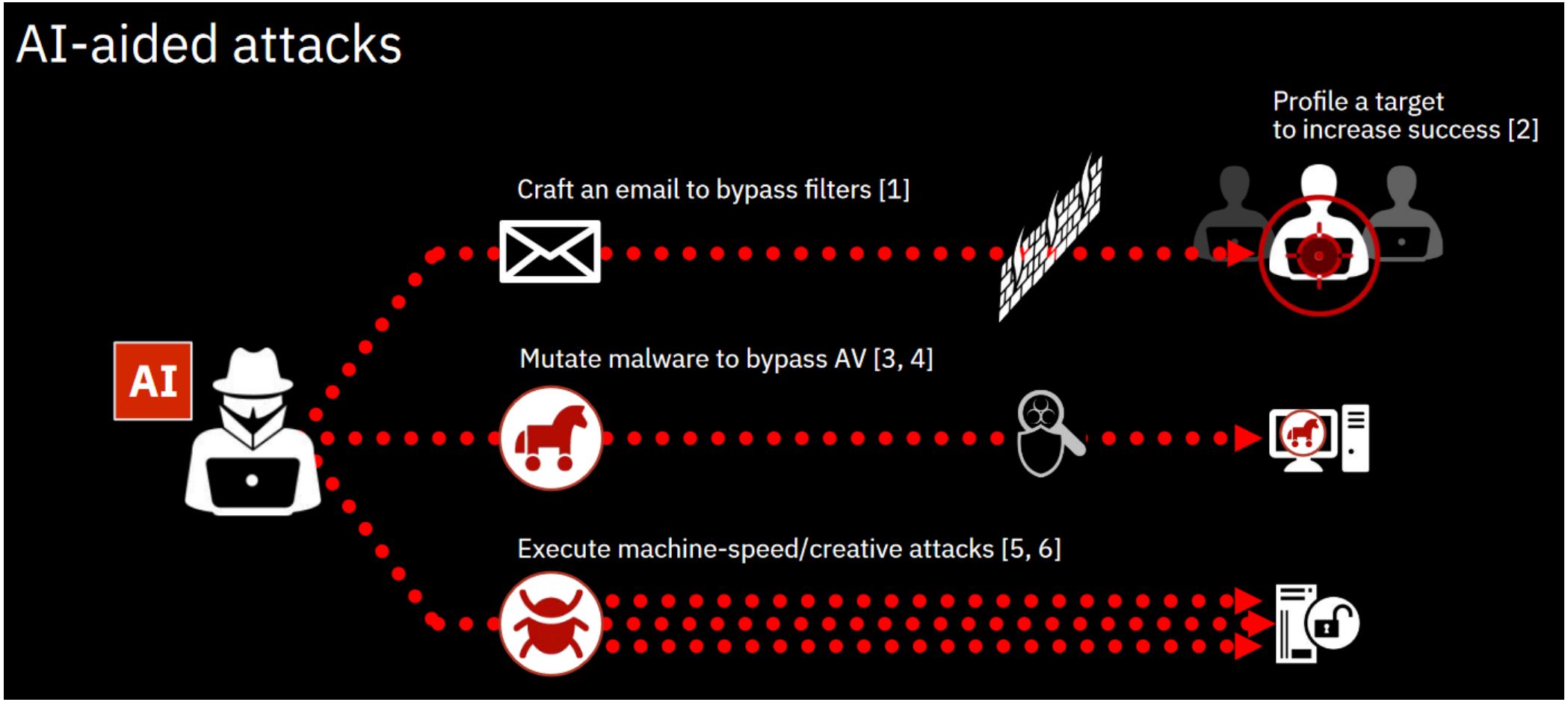
- I LLM sono potenzialmente utili in casi di abuso nel campo del terrorismo, della propaganda e della disinformazione.
- Ad esempio, il modello può essere utilizzato in generale per raccogliere più informazioni utili ad agevolare attività terroristiche, come ad esempio il finanziamento o la condivisione di file in anonimato
- ChatGPT eccelle nella produzione di testi apparentemente autentici, rapidamente e in quantità. Ciò rende il modello ideale per scopi di propaganda e disinformazione, poiché consente agli utenti di generare e diffondere messaggi che riflettono una narrativa specifica con uno sforzo relativamente minimo

Malware Development

- ChatGPT è in grado di produrre codice in diversi linguaggi di programmazione.
- La capacità di ChatGPT di trasformare le istruzioni del linguaggio naturale in codice funzionante è stata rapidamente sfruttata da soggetti malintenzionati per creare malware.
- Le misure di salvaguardia che impediscono a ChatGPT di fornire codice potenzialmente dannoso funzionano solo se il modello capisce cosa sta facendo. Se i prompt vengono suddivisi in singoli passaggi, è banale aggirare le policy di sicurezza.
- Poco dopo il rilascio pubblico di ChatGPT, in un post sul blog di Check Point Research del dicembre 2022 è stato mostrato come ChatGPT possa essere utilizzato per creare un flusso di infezione completo, dallo *spear-phishing* all'esecuzione di una shell inversa che accetta comandi in inglese
  - https://research.checkpoint.com/2022/opwnai-ai-that-can-save-the-day-or-hack-it-away

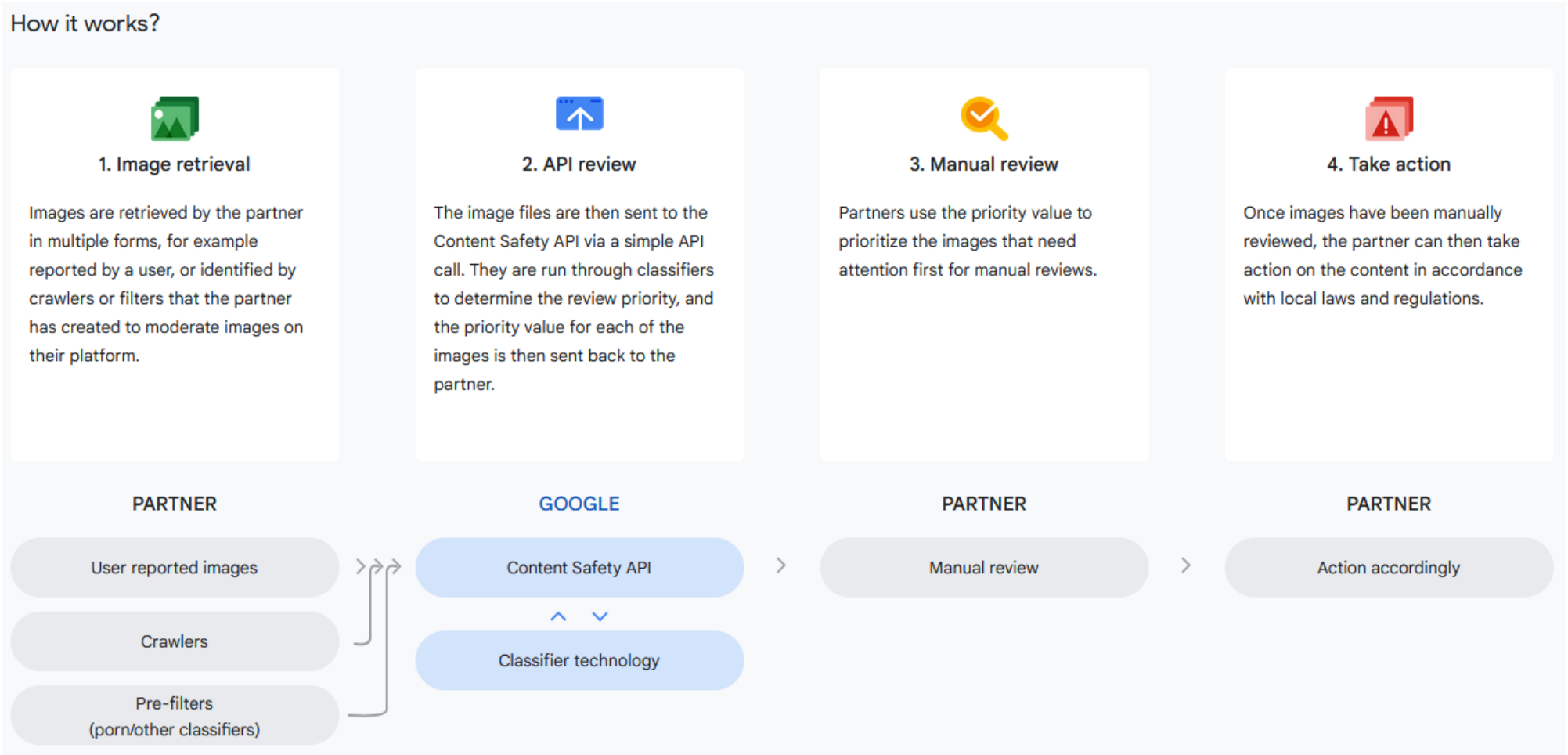
Un PoC di utilizzo malevolo: DeepLocker

- A BlackHat 2018, ricercatori IBM presentano DeepLocker, un malware sperimentale per condurre attacchi altamente mirati ed evasivi basato su tecnologia AI



Un utilizzo virtuoso: Content Safety API di Google

- Toolkit di Google messo gratuitamente a disposizione delle ONG partner per agevolare il riconoscimento di *child sexual abuse material* (CSAM)
- Anziché basarsi sul riconoscimento di contenuti già noti (comunque disponibile), impiega una IA per identificare contenuti ancora sconosciuti



Verso il regolamento UE

- Il 24 aprile è previsto il voto finale dell'Europarlamento sull'**AI Act**
- Tra i doveri degli sviluppatori, è previsto l'obbligo di pubblicare i materiali usati per l'addestramento degli algoritmi e l'obbligo di rendere riconoscibili i contenuti prodotti dall'AI.
- Sono inoltre contemplati gli usi ammessi e vietati dell'IA da parte delle forze dell'ordine
  - Sono generalmente vietati i sistemi di riconoscimento biometrico in tempo reale da remoto e negli spazi pubblici, con qualche eccezione…
    - Ricerca vittime di gravi reati o persone scomparse;
    - Prevenzione di minacce alla sicurezza pubblica o attacchi terroristici;
    - Localizzazione e identificazione dei sospettati per taluni reati (*serious crimes*);
  - …comunque limitata allo stretto necessario nello spazio e nel tempo e sottoposta ad autorizzazione preventiva o convalida da parte dell'Autorità competente;
  - Sono vietati i sistemi di polizia predittiva;
  - Sono ammessi sistemi di analisi del rischio che non profilano persone.
- Sono vietate inoltre le applicazioni ad alto rischio per i diritti fondamentali dei cittadini:
  - la categorizzazione biometrica che utilizza caratteristiche sensibili (razza, convinzioni politiche e religiose);
  - la raccolta indiscriminata di immagini dei volti per creare database di riconoscimento facciale;
  - il riconoscimento delle emozioni sul posto di lavoro e nelle scuole;
  - la classificazione sociale (*social scoring*);
  - gli usi volti a manipolare la volontà delle persone o a sfruttare i loro aspetti più vulnerabili (età, disabilità, condizioni di fragilità).

Artificial Intelligence for digital forensics

Il peso dei dati sui laboratori

- Tutti i reati sono in qualche modo informatici
  - Aumentano i casi che richiedono DFIR
  - Aumentano i reperti per caso
- La capienza degli storage aumenta
  - Quant'è un megabyte?
- L'accesso a risorse in cloud aumenta
- I dispositivi smart aumentano
- L'integrazione e i metadati aumentano

- Lo sviluppo tecnologico aumenta: nuovi servizi, piattaforme, sistemi operativi, app, artefatti... la frequenza dei rilasci
- Tutto questo rischia di causare un DDoS sui laboratori di digital forensics

Del rigore terminologico

- È corretto parlare di Artificial Intelligence?
- o di Big Data?
- di Machine Learning?
- di Deep Learning?
- Reti neurali?
- Expert systems?
- Machine vision?
- Natural Language Processing?

Si fa presto a dire "big data"

- Nel 2001, gli analisti si trovarono a fronteggiare circa mezzo milione di email relative al caso Enron

  Ricorrendo a tecniche di Social Network Analysis hanno potuto:

  - Scoprire gruppi nascosti ("*a group of individuals planning an activity over a communication medium without announcing their intentions*");
  - Scoprire la struttura organizzativa;
  - Dimostrare il modificarsi delle dinamiche comunicative durante situazioni di emergenza.
- Nel 2008, il caso TJX ha richiesto di elaborare 45 milioni di numeri di carte di credito.
  - Il CERT della Carnegie Mellon University sviluppò il programma CCFinder, che applicava tecniche di data mining e data reduction al fine di tracciare gli utilizzi abusivi, risalire al furto originale e agevolare la notifica alle vittime.

Si fa presto a dire "big data"

- Nel 2020 dal dump di un singolo smartphone è saltato fuori 1 milione di messaggi WhatsApp. True story.
  - Nel 2023 ho visto un iPhone con 5.5M di messaggi WeChat
- Abbiamo già provato in ogni modo a fronteggiare il problema con intelligenza:
  - Data mining
  - Data reduction
  - Link Analysis
  - Processing power
  - Distributed processing
  - Elastic cloud

- Tutto ancora utile, ma serve più intelligenza :-)

IA per la digital forensics

- L'uso di IA in analisi forense è un salto di paradigma
- Finora sono stati impiegati algoritmi sempre più sofisticati
  - Modelli deterministici
  - Elevato livello di specializzazione
  - Fondati su profonda comprensione del fenomeno da analizzare
- Gli algoritmi di IA invece sono basati sui dataset di addestramento
  - Imparano da esempi, non hanno necessità di comprendere il fenomeno e i suoi principi
- Coesisteranno entrambi gli approcci
- Le IA possono dare ausilio per:
  - risolvere problemi complessi, che con algoritmi ordinari sono intrattabili
  - risolvere meglio problemi che con algoritmi ordinari sono solubili, ma in modo insoddisfacente

IA per la digital forensics

L'ambito forense ha le sue esigenze, che sembrano difficili da conciliare con l'attuale natura delle IA, che per quanto deterministiche sono per lo più delle *black-box*

- Explainability

La capacità di comprendere e spiegare come l'IA sia giunta alla sua determinazione.

- Trustworthy

L'affidabilità intesa come fiducia che possiamo riporre nella bontà dei risultati.

Se l'IA diventa sufficientemente sofisticata, l'intelligenza umana non è più adeguata a comprendere le ragioni di un tale sistema, e pertanto dobbiamo ricorrere alla fiducia nei risultati senza poterli verificare.

Non è un problema da poco.

Ne consegue che l'impiego di IA nella Digital Forensics deve necessariamente prevedere un controllo umano

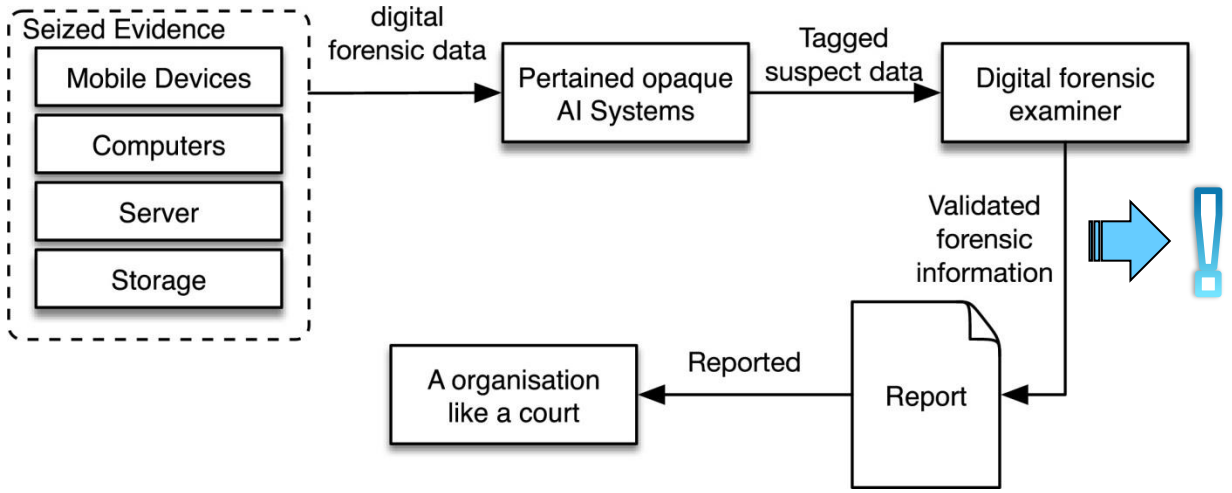
- AI-assisted investigation
- Human-in-the-loop

AI-assisted investigation

- Le IA possono essere utili come Decision Support Systems (DSS), poichè possono aiutare ad analizzare i dati raccolti come strumenti ausiliari, e non sostitutivi, del decisore umano.
- Ciò significa che abbiamo ancora molta strada da fare affinché i sistemi di IA dimostrino la loro mancanza di parzialità, la loro coerenza o la loro capacità di aiutare le forze dell'ordine nelle loro indagini.
- S. Costantini, G. De Gasperis, R. Olivieri: "Digital forensics and investigations meet artificial intelligence", Ann. Math. Artif. Intell., 86 (1) (2019), pp. 193-229, Springer

Human-in-the-loop

- I metodi dell'intelligenza artificiale vengono utilizzati per automatizzare attività minori come il rilevamento di droghe, armi, categorizzazione di chat e immagini. [...] argomenti come il rilevamento e il recupero dei dati, il triage dei dispositivi, l'analisi del traffico di rete, l'analisi forense dei dati crittografati, la ricostruzione di sequenze temporali e l'analisi forense di contenuti multimediali potrebbero trarre profitto dall'integrazione dei metodi dell'intelligenza artificiale nei loro campi di ricerca
- S.W. Hall, A. Sakzad, K.R. Choo: "Explainable artificial intelligence for digital forensics", WIREs Forensic Sci. (Jun. 2021)



- "Human investigators are responsible and accountable for the accuracy of the investigation. Machines are responsible for speed, humans for accuracy."
- J.Q. Chen: "Ai-Enabled Digital Forensic Evidence Examination", Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), vol. 1, Springer (2020), pp. 832-841

IA per la digital forensics

- Utilizzare un LLM di terze parti, magari pubblico, non è proponibile quando si trattano dati sensibili, privilegiati, confidenziali e coperti da segreto, quindi sarebbe necessaria un'installazione ospitata localmente.
- Tuttavia, i sistemi di terze parti possono essere impiegati per diversi task che non richiedono la condivisione di informazioni riservate
  - p.e. generazione automatica di scripts, query ed espressioni regolari
- *Question answering*: come con Siri o Alexa. La possibilità di formulare domande in linguaggio naturale per esplorare i dati di interesse giudiziario potrebbe essere estesa ad avvocati e magistrati, senza il supporto degli esperti forensi
- *Multilingual analysis*: la possibilità di formulare richieste nella propria lingua mentre il sistema può trovare informazioni pertinenti indipendentemente dal linguaggio in cui si presentano
- *Automated sentiment analysis*: la capacità di individuare velocemente in grandi quantità di comunicazioni contenuti minacciosi o violenti, *hate speech*, adescamento, molestie, phishing…

Perché tanta diffidenza?

Tra i rischi possiamo includere:

- **Bias ed errori**: come qualsiasi sistema di IA, il modello produce risultati basati sul dataset di addestramento
  - È valido? È completo? È aggiornato? È attendibile?
  - Il sistema non sa cosa è giusto o sbagliato moralmente o eticamente, è prevalentemente fatto per cercare di dare risposte in linguaggio umano.
- **Allucinazioni**: i sistemi attuali sono focalizzati sul dare comunque *una risposta*, piuttosto che *la risposta giusta*. Questo porta facilmente a risultati inaccurati, errati o del tutto inventati, presentati all'utente finale come *fatti* dalla forma plausibile. Senza la dovuta diligenza da parte degli utenti finali, ciò potrebbe erroneamente essere presentato come prova di un'assunzione.
- **Questioni legali**: l'uso di un LLM durante un'indagine potrebbe essere contestato in tribunale. A causa della sua architettura necessariamente complicata, si potrebbe perdere la capacità di spiegare il processo preciso seguito per identificare alcune prove incriminanti.
- **Eccessivo affidamento**: avere a disposizione un sistema facile da usare, potente e automatizzato può naturalmente comportare un eccesso di fiducia nel suo utilizzo.
- **Preoccupazioni etiche**: l'impiego di questa tecnologia in un contesto forense solleva alcune questioni etiche riguardanti la trasparenza, la privacy, l'equità, la non maleficenza e la fiducia
- **Mancanza di giudizio umano**: qualsiasi modello pre-addestrato potrebbe non essere in grado di fornire lo stesso livello di giudizio umano e intuizione necessari in molte indagini.
- **Limitazioni tecniche**: poiché questi modelli sono innanzitutto modelli linguistici, presentano gravi limitazioni sui dati che possono usare ed elaborare. Senza un'adeguata richiesta, qualsiasi risultato generato potrebbe non dichiarare i propri limiti, quali dati sono stati saltati o quali dati non erano utilizzabili.

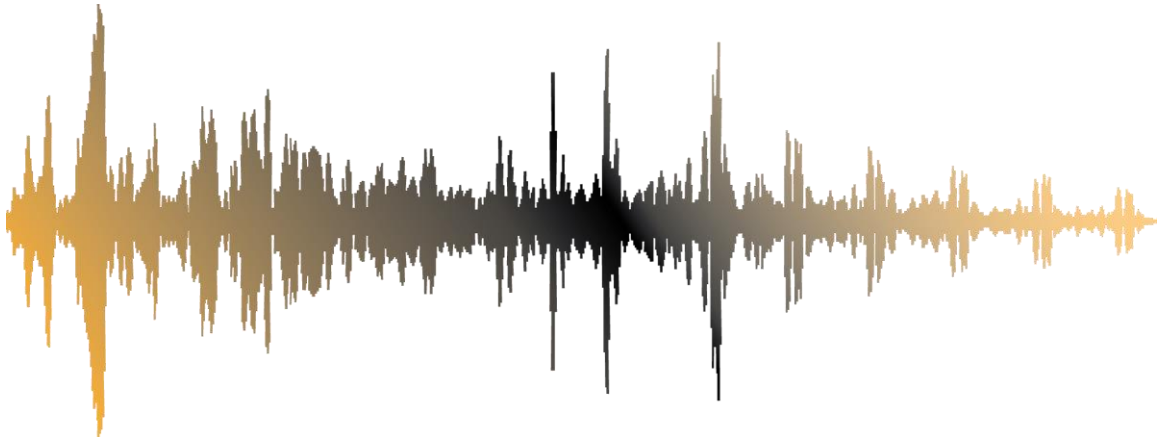
Adversarial models

- La possibile esistenza degli *adversarial models* (Nowroozi et al., 2021) è un problema intrinseco dei modelli di IA: se un modello è addestrato (e non è un'intelligenza artificiale generale) allora è possibile addestrare un modello antagonista, che crea degli input per il modello tali da produrre errori.
- A seconda del modello e del contesto, esistono diversi tipi di creazione di un modello antagonista, che crea un output precalcolato o casuale del modello AI (Zhang et al., 2020).
- L'esistenza di modelli antagonisti è un argomento contro l'uso dell'intelligenza artificiale in ambito forense.

Campi di applicazione per la digital forensics

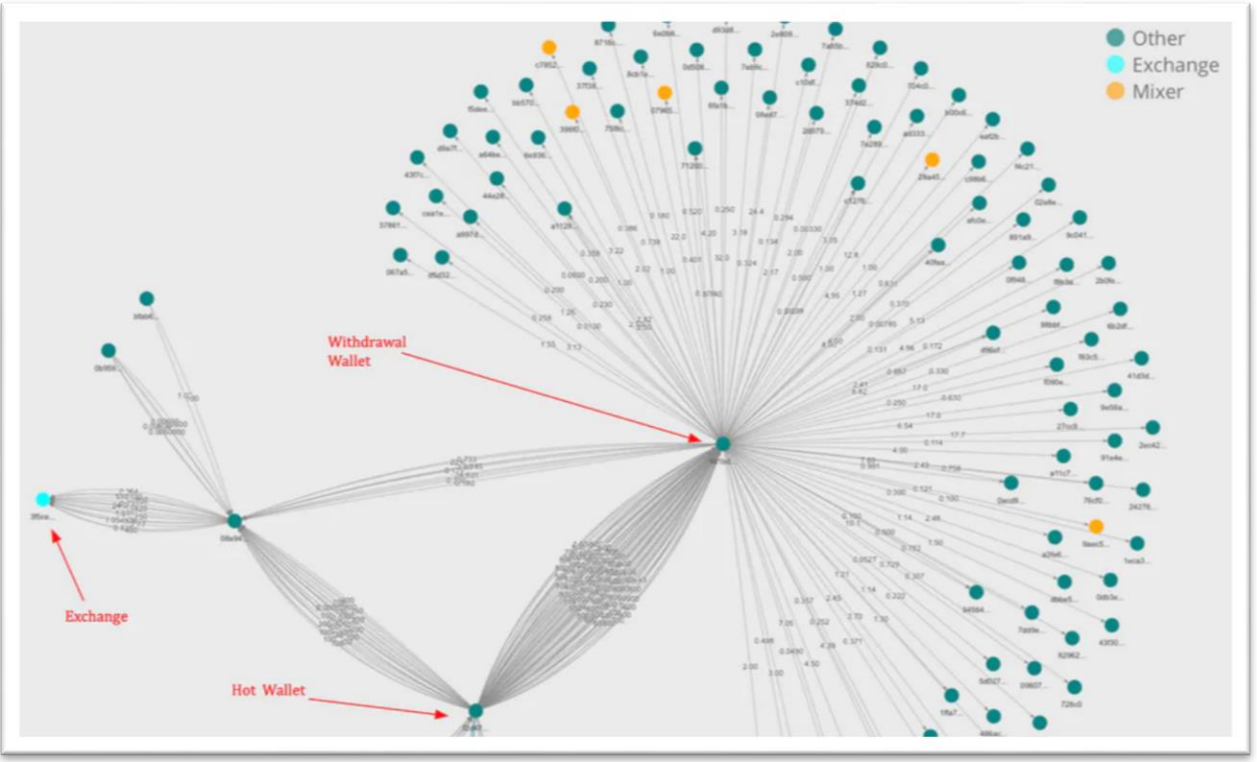
- Analisi semantica dei testi
  - Riassunto
  - Individuazione *key evidence*
  - Traduzione
  - Ricerche per significato
  - Sentiment analysis

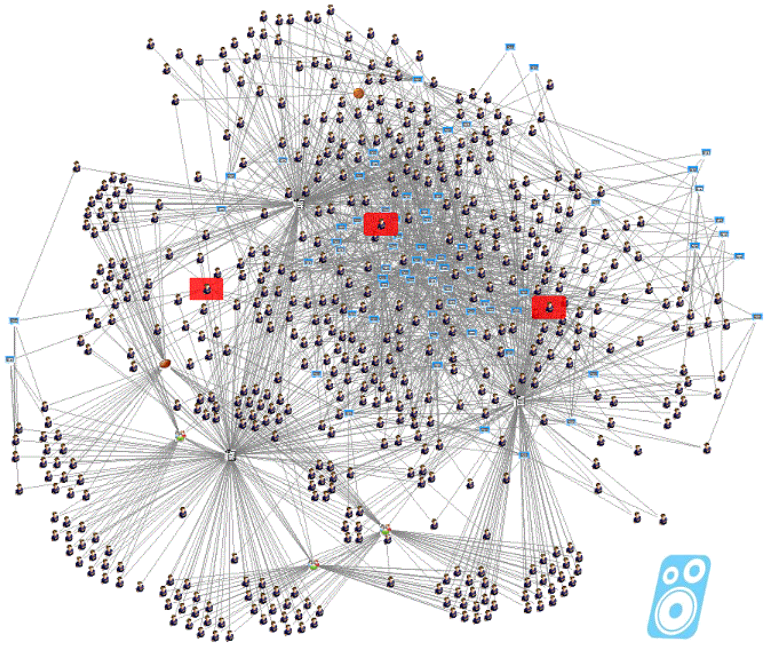
- Analisi file audio
  - Trascrizione messaggi vocali da diverse lingue
  - Riconoscimento del parlante
  - Rilevamento anomalie



Campi di applicazione per la digital forensics

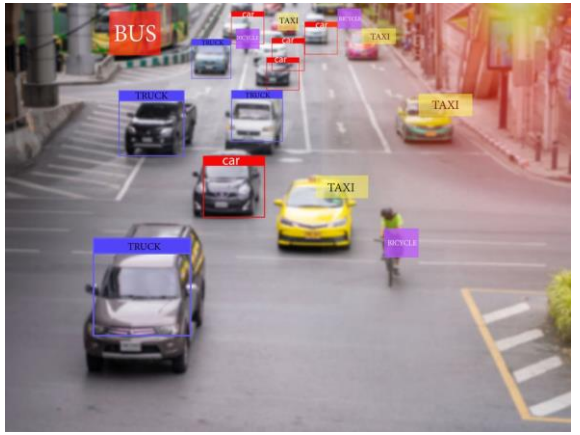
- Link analisys
- Triage
- Blockchain intelligence
- Malware analysis
- Network forensics





Campi di applicazione per la digital forensics

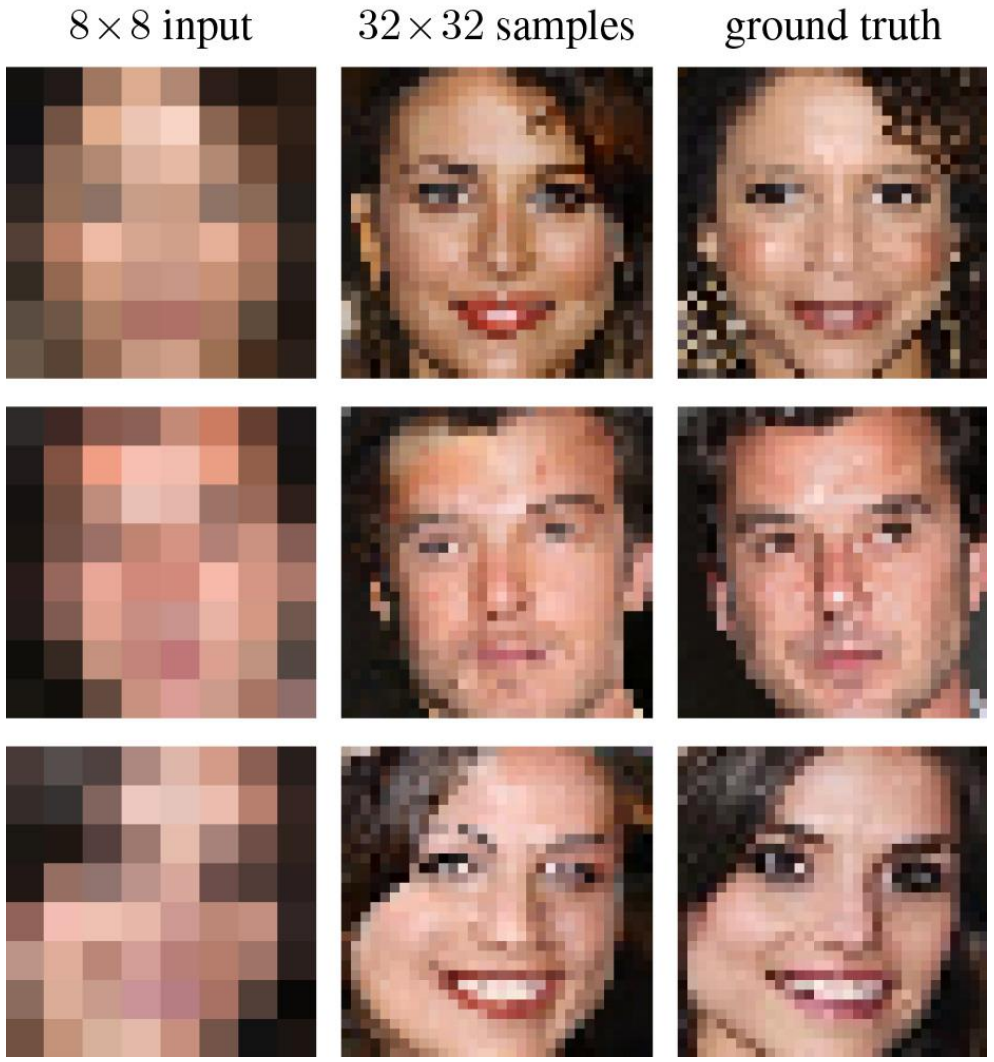
- Analisi immagini e video
  - Stima similarità (*approximate matching*)
    - Ricostruzione serie
  - Riconoscimento volti
    - Rilevamento, individuazione, aggregazione, categorizzazione…
    - Stima dell'età, del sesso o dell'etnia apparenti
  - Riconoscimento elementi
    - Persone, armi, droga, soldi, veicoli, targhe, documenti, mappe, pornografia, tatuaggi, simboli, screenshot…
  - Trascrizione testi
  - Rilevamento contraffazioni (inclusi *deepfake*)
- I video hanno una dimensione temporale:
  - Comprensione degli eventi dinamici
  - Analisi comportamentale e rilevamento anomalie



Miglioramento di immagini e video

- Si possono enfatizzare informazioni che non si vedono, ma ci sono

- Non si possono creare informazioni che non ci sono



Utilizzabilità

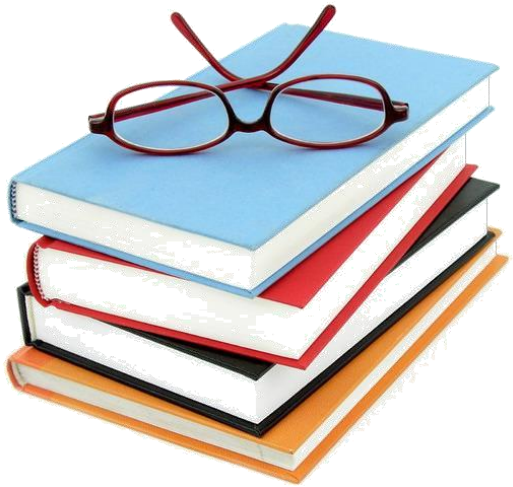
- Le elaborazioni di una IA difficilmente potranno avere una qualche rilevanza probatoria
  - Per quello sarà sempre richiesta la valutazione di un perito qualificato
- Possono invece essere uno spunto investigativo importante o addirittura necessario

Altrettanto come ausilio al giudice:

- Non possiamo lasciare che prendano decisioni sulle persone
- Ma possono essere un validissimo supporto decisionale
  - Possono fornire una stima di affidabilità
  - Possono mitigare i bias umani

Bibliografia

- Europol: "**ChatGPT: The impact of Large Language Models on Law Enforcement**", 2023
- Scanlon, Nikkel, Geradts: "**Digital forensic investigation in the age of ChatGPT**", Digital Investigation, Volume 44, 2023
- Sharon Ben-Moshe, Gil Gekker, Golan Cohen: "**OpenAI: AI That Can Save the Day or HACK it Away**", 2022
- Fähndrich et al.: "**Digital forensics and strong AI: A structured literature review**", Digital Investigation, Volume 46, 2023
- Hall, Sakzad, Choo: "**Explainable artificial intelligence for digital forensics**", WIREs Forensic Science, Vol.4, Issue 2, 2022
- Costantini et al. : "**Digital forensics and investigations meet artificial intelligence**", Annals of Mathematics and Artificial Intelligence, 2019
- Jim Q. Chen: "**AI-Enabled Digital Forensic Evidence Examination**", AISC book series, volume 1129, 2020
- Abiodun A. Solanke: "**Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models**", Digital Investigation, Volume 42 supplement, 2022
- Nowroozi, Dehghantanha, Parizi, Choo: "**A survey of machine learning techniques in adversarial image forensics**", Computers & Security, Vol.100, 2021
- Zhang, Sheng, Alhazmi, Li: "**Adversarial Attacks on Deep-learning Models in Natural Language Processing: A Survey**", ACM Transactions on Intelligent Systems and Technology, Volume 11Issue 3, 2020
- Lim, Tan, Hock, Lee: "**Turing in a Box: Applying Artificial Intelligence as a Service to Targeted Phishing and Defending against AI-generated Attacks**", BlackHat USA 2021
- Scanlon et al.: "**ChatGPT for digital forensic investigation: The good, the bad, and the unknown**", Digital Investigation, Volume 46 Supplement, 2023
- Kirat, Jang, Stoecklin: "**DeepLocker: Concealing Targeted Attacks with AI Locksmithing**", BlackHat USA 2018



Teniamoci in contatto...

Davide Rebus Gabrini

e-mail: davide.gabrini@unipv.it

GPG Public Key: www.tipiloschi.net/rebus.asc

KeyID: 0x176560F7

Queste e altre cazzate su

www.tipiloschi.net

facebook.com/gabrini

twitter.com/therebus

it.linkedin.com/in/rebus

- **Rebus' Digest**
  newsletter su cybercrime, hacking, digital forensics...
- **EventiLoschi**
  calendario delle conferenze pubbliche in materia