Numero 6 - luglio 2025

INTELLIGENZA ARTIFICIALE: IMPATTO, RISCHI E OPPORTUNITA' NELL'ECOSISTEMA OSINT

MIRKO LAPI

INDICE

Mirko Lapi è Consulente e Formatore specializzato in Open Source Intelligence (OSINT), sicurezza delle informazioni e sviluppo del pensiero critico applicato ai processi decisionali. Ha prestato servizio nelle Forze Armate italiane per 27 anni, di cui 16 anni all'interno del II Reparto Informazioni e Sicurezza dello Stato Maggiore della Difesa, dove ha lavorato come analista presso il Centro Intelligence Interforze dello Stato Maggiore della Difesa e come docente di Intelligence presso il Centro Interforze di Formazione Intelligence (CIFIGE). È Professore a contratto di Open Source Intelligence (OSINT) presso il Dipartimento di Giurisprudenza dell'Università degli Studi di Foggia. Inoltre, ricopre il ruolo di Professore a contratto di Cyber Security per il Corso di Laurea Magistrale in Ingegneria dei Sistemi Intelligenti presso la Facoltà di Ingegneria, Università Campus Bio-Medico di Roma.

Torna all'indice

EXECUTIVE SUMMARY

Questo report analizza la trasformazione radicale dell'Open-Source Intelligence (OSINT) guidata dall'Intelligenza Artificiale (IA). La tesi di fondo è che l'IA non si configura come un semplice potenziamento incrementale, ma come un vero e proprio agente di ristrutturazione sistemica che ridefinisce ogni fase del ciclo dell'Intelligence. Tale assunto deriva dalla consapevolezza che questa tecnologia amplifica le capacità analitiche a una scala, velocità e profondità senza precedenti, consentendo di estrarre valore da un universo informativo altrimenti ingestibile.

Analisi SWOT

Punti di Forza: Il vantaggio primario dell'IA risiede nella sua capacità di automazione su larga scala, nell'analisi predittiva e nel riconoscimento di pattern complessi e correlazioni non visibili ad "occhio nudo". Queste capacità consentono di trasformare grandi volumi di dati grezzi e non strutturati in Intelligence azionabile e tempestiva,

superando i limiti cognitivi e operativi umani.

- Punti di Debolezza: Parallelamente, emergono vulnerabilità intrinseche e critiche. Il problema della "*black box*" (opacità decisionale), i bias algoritmici che perpetuano e amplificano pregiudizi, le "allucinazioni" dei modelli generativi che producono falsità credibili e una pericolosa erosione del pensiero critico dell'analista rappresentano sfide fondamentali all'integrità e all'affidabilità dell'Intelligence prodotta.
- Opportunità: L'applicazione dell'IA-OSINT apre nuove frontiere operative. Tra queste spiccano il contrasto avanzato alla disinformazione e ai deepfake, una gestione più efficace delle crisi umanitarie e dei disastri naturali e una lotta più incisiva alla criminalità finanziaria globale e al terrorismo, oltre a una parziale democratizzazione degli

strumenti di analisi.

- Minacce: Le minacce sono altrettanto significative. La strumentalizzazione dell'IA da parte di attori statali ostili e gruppi criminali, la possibilità di una sorveglianza di massa pervasiva e l'erosione della privacy individuale, unite a un profondo vuoto normativo ed etico, costituiscono minacce sistemiche alla sicurezza nazionale, alla stabilità democratica e ai diritti fondamentali.

L'integrazione dell'IA nell'OSINT non è più un'opzione ma un imperativo strategico. Tuttavia, il suo successo e la sua sostenibilità dipendono dall'adozione di un modello di "intelligenza aumentata", in cui l'analista non viene sostituito, ma potenziato, mantenendo un ruolo centrale e insostituibile di supervisione critica, giudizio contestuale e decisione strategica.

La governance di questa tecnologia, attraverso lo sviluppo di quadri normativi efficaci e tempestivi, l'adozione di standard etici e l'implementazione di tecnologie di Explainable AI (XAI) è, già oggi, un imperativo non più procrastinabile.

LA NUOVA FRONTIERA DELL'INTELLIGENCE - DEFINIRE L'OSINT NELL'ERA DELL'IA

Definizione ed evoluzione dell'OSINT
L'Open-Source Intelligence (OSINT) è definita come l'Intelligence derivata dalla raccolta, analisi e utilizzo di informazioni che sono pubblicamente e legalmente disponibili. Il suo scopo è supportare i processi di Intelligence e decisionali in una vasta gamma di contesti. Le sue fonti si sono evolute drasticamente nel tempo.
Storicamente, l'OSINT si basava su media tradizionali come libri, giornali, riviste e trasmissioni radiofoniche.

Con l'avvento di Internet e la successiva esplosione del web 2.0, l'universo delle fonti si è espanso in modo esponenziale, includendo oggi siti web, blog, forum, social network (Facebook, Twitter, LinkedIn), database pubblici (rapporti governativi, dati demografici, atti legislativi), dati geospaziali (mappe, immagini satellitari), metadati incorporati in file digitali e ancora la cosiddetta "letteratura grigia" (pubblicazioni scientifiche, atti di conferenze).

Sebbene la pratica di raccogliere Intelligence da fonti aperte abbia radici che risalgono a centinaia di anni, la rivoluzione digitale ha introdotto una sfida senza precedenti: il "sovraccarico informativo". La vastità e la velocità con cui vengono generati nuovi dati hanno reso i metodi di analisi manuale tradizionali inefficienti e, in molti casi, insostenibili.

In questo contesto, è fondamentale distinguere l'OSINT dalla semplice raccolta di informazioni (OSINF - Open Source Information).

L'OSINF rappresenta i dati grezzi, mentre l'OSINT è il prodotto finale di un processo strutturato. Questo processo, noto come ciclo dell'Intelligence, trasforma i dati grezzi in "*actionable Intelligence*", ovvero un prodotto analizzato, validato, contestualizzato e disseminato a un pubblico specifico per rispondere a una precisa esigenza informativa. Questa trasformazione è ciò che conferisce all'OSINT il suo valore strategico e la distingue da altre discipline di raccolta come la Human Intelligence (HUMINT) o la Technical Intelligence (TECHINT), che può comunque integrare e potenziare.
Il valore dell'OSINT è oggi riconosciuto in molteplici settori. Per le agenzie governative e i servizi di sicurezza, rappresenta una componente primaria per la protezione degli interessi nazionali politici, militari ed economici. Per le forze dell'ordine, è uno strumento investigativo fondamentale. Nel settore privato, le aziende la utilizzano per ottenere un vantaggio competitivo, monitorare i concorrenti, proteggere la propria

reputazione e identificare minacce emergenti. Infine, per il giornalismo investigativo e la società civile, l'OSINT è diventata una risorsa indispensabile per la verifica dei fatti e la documentazione di eventi di interesse pubblico. Questi sono alcuni esempi dei possibili campi di applicazioni a cui bisogna poi aggiungere l'altra faccia della medaglia ovvero che l'OSINT può essere utilizzato, come vedremo, anche da un'ampia gamma di attori ostili.

L'Intelligenza Artificiale come agente di trasformazione

L'avvento dell'Intelligenza Artificiale (IA) sta segnando un punto di svolta epocale per la disciplina OSINT. Infatti l'IA non rappresenta solo un semplice miglioramento incrementale degli strumenti esistenti, ma un vero e proprio agente di "completa ristrutturazione" e "trasformazione" che sta riscrivendo le regole del gioco. Quindi la sua integrazione nel ciclo dell'Intelligence non è più un'opzione, ma un "imperativo" per chiunque voglia mantenere efficacia e rilevanza nell'era digitale.

L'impatto dell'IA segna un "prima e un dopo", una vera e propria rivoluzione metodologica.

Il cuore di questa trasformazione risiede nella capacità dell'IA di processare volumi di dati che sono semplicemente insostenibili per qualsiasi team umano, e di farlo in tempo reale, superando le limitazioni cognitive e operative dell'analista. Tali considerazioni ci conducono quindi al concetto di **intelligenza aumentata**.

In questo paradigma, l'IA non è concepita per sostituire l'analista, ma per "iper-abilitarlo". Ciò significa che l'IA si fa carico delle attività ripetitive, laboriose e su larga scala, come la raccolta e il filtraggio dei dati, permettendo all'analista di liberare le proprie risorse cognitive per concentrarsi su compiti a più alto valore aggiunto: l'analisi strategica, l'interpretazione contestuale, il giudizio critico e le decisioni finali. Questa simbiosi uomo-macchina sta accelerando la fusione di discipline un tempo distinte. Ad esempio l'OSINT è ormai parte determinante della Cyber Threat Intelligence (CTI). Il processo logico è chiaro: l'OSINT si basa su fonti pubbliche; la CTI si concentra sulle minacce informatiche; l'IA, applicata all'OSINT, automatizza il monitoraggio continuo di fonti pubbliche per la CTI (dark web, i forum di hacker e i social media) al fine di rilevare in modo proattivo campagne di phishing, la vendita di credenziali rubate, nuove vulnerabilità, ecc...

Di conseguenza, l'IA non solo migliora l'OSINT, ma la trasforma in una componente proattiva e predittiva della CTI, spostando la postura di sicurezza da reattiva ad anticipatoria e preventiva.
Tuttavia, questa trasformazione porta con sé un paradosso. Se da un lato l'accesso a potenti strumenti AI-OSINT sembra "democratizzare" le capacità di Intelligence, rendendole disponibili a giornalisti, ricercatori e ONG, dall'altro rischia di creare un nuovo e profondo divario digitale. I modelli di IA più avanzati, i dataset di addestramento più vasti e la potenza di calcolo necessaria per operare su larga scala rimangono appannaggio esclusivo di grandi aziende tecnologiche e agenzie governative ben finanziate. Ciò crea una disparità significativa tra chi può permettersi un' "IA di frontiera" e chi deve affidarsi a strumenti open-source o a basso costo, che possono essere meno accurati, più lenti o più vulnerabili.
L'implicazione strategica è che la vera democratizzazione è limitata e il potere informativo rischia di

concentrarsi ulteriormente nelle mani di pochi attori d'élite, con conseguenze significative per l'equilibrio di potere tra stati e per il ruolo della società civile.

Il Ciclo dell'Intelligence riscritto dall'IA

Il tradizionale ciclo dell'Intelligence, un processo metodico e sequenziale, viene potenziato e accelerato in ogni sua fase dall'integrazione dell'Intelligenza Artificiale. Vediamo come fase per fase:

- **Definizione del problema e pianificazione:** Tradizionalmente, la fase di pianificazione è un'attività puramente umana, incentrata sulla definizione dei requisiti e degli obiettivi. Tuttavia, l'integrazione dell'IA sta trasformando questo stadio iniziale in un processo più dinamico di generazione di ipotesi e di esplorazione strategica, dove l'IA agisce come un "partner di ideazione" per potenziare l'intelletto umano. Sfruttando la sua capacità di analizzare insiemi di dati vasti

e complessi, l'IA può identificare pattern, correlazioni e anomalie che sfuggirebbero all'osservazione umana, fornendo così le basi per formulare ipotesi più solide e basate sull'evidenza. Ciò permette quindi di allargare la visione oltre la semplice definizione di obiettivi, per esplorare attivamente nuove piste investigative e pensare in maniera controintuitiva (capacità umana oggi sempre più rara). Sistemi di IA avanzati possono essere addestrati per generare autonomamente ipotesi di ricerca innovative, scoprire proprietà inaspettate nei dati e proporre scenari ipotetici, supportando così gli analisti a sfidare lo status quo e i propri preconcetti. Invece di limitarsi a rispondere a domande predefinite, l'IA aiuta a porre domande migliori e più pertinenti, trasformando la pianificazione in una vera e propria fase di "scoperta strategica".

- **Raccolta:** Questa è la fase dove l'automazione guidata dall'IA ha l'impatto più immediato. L'IA automatizza il *web scraping* e il

crawling di una moltitudine di fonti, inclusi social media, forum, siti di notizie e persino il deep e dark web. Bot e crawler intelligenti sono in grado di adattarsi dinamicamente alle strutture dei siti web, superare restrizioni come i CAPTCHA e raccogliere dati in modo continuo e instancabile, un compito impossibile da svolgere manualmente su tale scala.*

- **Elaborazione:** I dati raccolti sono tipicamente voluminosi, grezzi e non strutturati. L'IA interviene per elaborarli, convertendoli in formati adatti all'analisi. Questo include il filtraggio del "rumore" (informazioni irrilevanti), la standardizzazione dei dati, la rimozione di duplicati e la traduzione automatica di contenuti multilingue. In questa fase, l'IA migliora notevolmente la capacità di "vagliare" le fonti (discrimination), distinguendo tra informazioni potenzialmente rilevanti e affidabili e quelle fuorvianti o false.

*Fondamentale comprendere sempre i limiti normativi entro i quali è possibile operare. Rimando in questo senso alla lettura: OSINT: strumenti legali per aziende e investigatori privati

- **Analisi:** È in questa fase che l'IA dispiega il suo potenziale più profondo e trasformativo. Algoritmi di Machine Learning (ML), Natural Language Processing (NLP) e Computer Vision vengono applicati ai dati elaborati per estrarre insight significativi. L'IA può identificare pattern e correlazioni complesse, condurre analisi del sentiment su larga scala, mappare reti sociali e analizzare contenuti multimediali (immagini e video) per estrarre informazioni come la geolocalizzazione o il riconoscimento di volti e oggetti.
- **Diffusione:** Infine, l'IA ottimizza anche la fase di diffusione dell'Intelligence prodotta. Può generare automaticamente report personalizzati, creare dashboard interattive e sistemi di allerta in tempo reale. Questo permette di adattare l'output alle esigenze specifiche dei decisori finali, garantendo che le informazioni giuste arrivino alla persona giusta, nel formato più utile e al momento opportuno.

Tabella 1 - Come le specifiche tecnologie di IA si applicano alle diverse fasi del ciclo OSINT.

Fase del Ciclo OSINT	Tecnologia AI applicata	Funzione specifica
Pianificazione (Planning)	Intelligenza Artificiale Generativa (GenAI), Machine Learning (ML)	Generare ipotesi, identificare gap informativi, sfidare i preconcetti, suggerire nuove piste di indagine.
Raccolta (Collection)	Web Scraping e Crawling potenziati da IA	Automatizzare la raccolta continua di dati da social media, forum, siti di notizie, deep e dark web.
Elaborazione (Processing)	Natural Language Processing (NLP), Machine Learning (ML)	Filtrare il "rumore", classificare i dati per rilevanza, tradurre automaticamente contenuti multilingue, estrarre entità nominate.
Analisi (Analysis)	Machine Learning (ML), NLP, Computer Vision, Analisi di Reti Sociali (SNA)	Riconoscere pattern e anomalie, analisi del sentiment, mappare e analizzare reti di influenza, riconoscimento facciale, geolocalizzazione da immagini, rilevamento deepfake.
Diffusione (Dissemination)	Intelligenza Artificiale Generativa (GenAI), Piattaforme di Business Intelligence (BI)	Creare automaticamente report di Intelligence, generare riassunti esecutivi, costruire dashboard interattive e sistemi di allerta personalizzati.

Torna all'indice

PUNTI DI FORZA: L'AMPLIFICAZIONE DELLE CAPACITÀ ANALITICHE

L'integrazione dell'Intelligenza Artificiale nell'OSINT non è solo un cambiamento di scala, ma un salto qualitativo che potenzia le capacità analitiche in modi prima inimmaginabili. I punti di forza principali risiedono nella capacità di superare i limiti umani in termini di velocità e volume, di scoprire pattern nascosti all'interno di dati complessi e di spostare il paradigma dell'Intelligence da reattivo a predittivo.

Velocità e scala: Superare i limiti umani

Il vantaggio più evidente e immediato dell'IA nell'OSINT è la sua capacità di operare a una velocità e su una scala che trascendono completamente le capacità umane. L'universo delle fonti aperte genera dati a un ritmo vertiginoso; un team di analisti, per quanto numeroso e qualificato, non potrebbe mai sperare di monitorare

raccogliere e processare manualmente questa mole di informazioni in modo tempestivo. L'IA affronta direttamente il problema cronico del "sovraccarico informativo" che affligge l'Intelligence moderna, trasformandolo da ostacolo a opportunità.
L'automazione delle attività ripetitive e a basso valore cognitivo — come la raccolta dati, il filtraggio iniziale, la classificazione e la traduzione — libera un tempo prezioso per gli analisti. Invece di passare ore a setacciare manualmente fonti disparate, gli analisti possono delegare questi compiti all'IA e concentrarsi su attività a più alto valore aggiunto: la validazione critica delle fonti, l'interpretazione contestuale, il ragionamento strategico e la formulazione di giudizi complessi. Questa simbiosi permette di ottenere *insight* quasi immediati da flussi di dati in tempo reale, riducendo drasticamente il "*time-to-insight*". Un esempio emblematico di questa capacità è la

la gestione dei disastri naturali. Piattaforme come l'ODET (Open-Source Intelligence Disaster Event Tracker) hanno dimostrato un'efficacia straordinaria durante eventi come l'uragano Harvey nel 2017 e il terremoto in Turchia del 2023. Utilizzando modelli di IA non modificati, la piattaforma è stata in grado di analizzare enormi volumi di dati provenienti da social media (tweet) su base oraria, filtrando i contenuti irrilevanti, classificando le informazioni per pertinenza, costruendo grafi di conoscenza e generando report di consapevolezza situazionale. I risultati sono stati notevoli: i report generati automaticamente hanno raggiunto un punteggio di accuratezza fino all'89% (utilizzando la metrica AlignScore per confrontarli con articoli di Wikipedia scritti a posteriori), dimostrando un salto quantico in termini di velocità e scala rispetto a qualsiasi approccio manuale.

Riconoscimento di pattern nascosti

Al di là della velocità, la vera potenza trasformativa dell'IA risiede nella sua capacità di identificare pattern, anomalie, correlazioni e connessioni nascoste all'interno di dataset vasti, eterogenei e complessi. Gli algoritmi di Machine Learning (ML) sono specificamente progettati per questo scopo, andando a scoprire strutture e relazioni che sarebbero praticamente invisibili all'analisi umana.

L'analisi delle reti sociali è uno dei campi in cui questo potenziale si manifesta con maggiore chiarezza. L'IA può mappare automaticamente le relazioni tra persone, organizzazioni, eventi e risorse digitali, visualizzando reti complesse che rivelano centri di influenza, cellule criminali o terroristiche, strutture societarie occulte e legami finanziari nascosti. Parallelamente, l'analisi comportamentale permette all'IA di

stabilire una "*baseline*" del comportamento normale all'interno di un sistema o di una rete e di identificare deviazioni anomale. Ciò è ad esempio fondamentale per rilevare minacce sofisticate come account falsi, botnet che coordinano campagne di disinformazione, truffe online, ecc... Un caso di studio illuminante proviene dalla sicurezza nazionale statunitense. L'FBI è riuscita a smascherare un complesso schema in cui centinaia di lavoratori IT nordcoreani, utilizzando identità false, erano riusciti a farsi assumere da remoto da aziende statunitensi, incanalando i loro stipendi verso il regime di Pyongyang. Collegando "puntini" apparentemente non correlati sparsi in un'enorme quantità di dati pubblici non classificati, l'analisi di pattern su questa scala, ha permesso di scoprire affiliazioni estere non dichiarate e altre anomalie.

Analisi predittiva: Dalla reazione all'anticipazione

Forse il punto di forza più strategico dell'IA nell'OSINT è la sua capacità di spostare il paradigma dell'Intelligence da un'attività puramente reattiva (analisi di eventi passati) a una proattiva e anticipatoria (previsione di eventi futuri). Sfruttando l'analisi di dati storici e il monitoraggio dei trend in tempo reale, i modelli di IA possono essere addestrati per prevedere con un certo grado di probabilità l'insorgere di crisi economiche, instabilità politica, attacchi informatici o persino l'aumento di determinate tipologie di crimine. Nel campo della Cyber Threat Intelligence (CTI), l'IA potrebbe consentire di anticipare potenziali attacchi informatici prima che vengano lanciati. Ad esempio, analizzando il "*chatter*" su forum underground e sul dark web, potrebbero essere identificate discussioni di interesse, la vendita o lo sviluppo di nuovi exploit,

la pianificazione di campagne di phishing di massa o la compravendita di credenziali rubate. Tutti elementi che analizzati consentono di esprimere probabilità e quindi potenziali preavvisi ai team di sicurezza.
Nella previsione dell'instabilità geopolitica, l'analisi del *sentiment* su larga scala, applicata ai social media e ad altre fonti aperte, offre un potente strumento per "tastare il polso" di una società. È possibile monitorare l'evoluzione del pensiero pubblico, identificare la diffusione di narrative estremiste e rilevare "segnali deboli" (weak signals) di disordini sociali o politici emergenti. Questa capacità è di valore inestimabile per supportare missioni di peacekeeping, prevenire conflitti e supportare le attività diplomatiche.
Infine, nel Crime Forecasting, l'analisi combinata di dati OSINT in tempo reale con dati esterni, come le statistiche sulla criminalità locale o i report meteorologici, può aiutare le forze dell'ordine e le agenzie di sicurezza privata a prevedere e allocare le risorse in

modo più efficace per mitigare i rischi per la sicurezza fisica di persone e infrastrutture.
Questa transizione verso l'analisi predittiva non rappresenta solo un vantaggio operativo, ma un'arma strategica. La capacità di agire sulla base di informazioni quasi istantanee, o addirittura prima che un evento si verifichi, crea un vantaggio decisionale asimmetrico. In scenari competitivi — siano essi militari, finanziari o politici — l'attore che può anticipare le mosse dell'avversario o le dinamiche del contesto operativo detiene un vantaggio decisivo. Di conseguenza, il possesso di capacità AI-OSINT superiori non è più solo una questione di efficienza, ma un fattore determinante del potere nazionale o aziendale, in grado di influenzare l'esito di conflitti, crisi di mercato e campagne elettorali.
L'IA, in questo senso, non sta semplicemente aiutando a "raccogliere" informazioni; sta trasformando l'OSINT in una disciplina capace di rivelare fenomeni completamente nuovi e imprevisti.

DEBOLEZZE: LE VULNERABILITÀ INTRINSECHE E LE SFIDE OPERATIVE

Nonostante l'enorme potenziale, l'integrazione dell'Intelligenza Artificiale nell'OSINT introduce una serie di debolezze significative e sfide operative. Queste vulnerabilità non sono semplici difetti tecnici, ma problemi fondamentali che possono minare l'affidabilità, l'equità e la stessa integrità del processo di Intelligence. Le principali debolezze risiedono nell'opacità dei modelli, nei bias algoritmici, nella tendenza a "inventare" fatti e, forse la più insidiosa, nell'impatto negativo sul pensiero critico dell'analista.

Il Problema della "Black Box": Opacità e mancanza di spiegabilità

Molti dei modelli di IA più potenti e avanzati, in particolare le reti neurali profonde (deep learning), operano come "scatole nere" (black boxes). Questo significa che, sebbene possano produrre risultati estremamente accurati, il loro

processo decisionale interno rimane oscuro e incomprensibile persino per gli ingegneri che li hanno progettati. A differenza del software tradizionale, dove la logica è esplicitamente codificata, questi modelli apprendono da soli pattern complessi dai dati, rendendo quasi impossibile tracciare un percorso logico chiaro dalla richiesta all'output.

Questa opacità rappresenta, di fatto, un problema estremamente critico in un contesto di Intelligence. *Se un analista non è in grado di spiegare perché un sistema di IA ha segnalato un individuo come una minaccia o ha identificato una particolare narrazione come disinformazione, l'Intelligence prodotta perde gran parte della sua credibilità e utilizzabilità. In contesti decisionali ad alto rischio, come operazioni militari o procedimenti legali,* ***un'informazione non giustificabile è un'informazione inutilizzabile***.

La mancanza di trasparenza mina la fiducia nel sistema, rende quasi impossibile l'*accountability* (responsabilità) in caso di errore e impedisce di identificare e correggere efficacemente i bias nascosti nel modello.
Per affrontare questa sfida, è emerso il campo dell'Explainable AI (XAI), o IA Spiegabile. L'XAI mira a sviluppare tecniche e modelli che siano intrinsecamente trasparenti o che possano fornire giustificazioni comprensibili per le loro decisioni.
Metodi come LIME (Local Interpretable Model-agnostic Explanations) e SHAP (SHapley Additive exPlanations) cercano di far luce sulla "scatola nera", mostrando quali caratteristiche dei dati di input hanno maggiormente influenzato una specifica previsione.
Tuttavia, queste tecniche sono ancora in fase di sviluppo, spesso forniscono solo spiegazioni approssimative o *post-hoc*, e la sfida di rendere i modelli più complessi pienamente interpretabili rimane in gran parte irrisolta.

Esiste una tensione fondamentale e forse irrisolvibile: i modelli più potenti tendono ad essere i più opachi, mentre i modelli più semplici e trasparenti sono spesso meno performanti su compiti complessi. Questo pone i decisori di fronte a un difficile compromesso tra accuratezza e spiegabilità, una scelta di policy che ha profonde implicazioni etiche e di gestione del rischio.

Bias algoritmico

Una delle debolezze più pervasive e dannose dell'IA è il bias algoritmico. I modelli di IA apprendono dai dati con cui vengono addestrati; se questi dati riflettono i pregiudizi, ad esempio le disuguaglianze e le distorsioni presenti nella società, l'IA non solo li apprenderà, ma li codificherà, li sistematizzerà e li amplificherà.

Nel contesto dell'OSINT, il bias può insinuarsi in ogni fase del ciclo di Intelligence, portando a conclusioni errate e decisioni potenzialmente discriminatorie:

- Bias di selezione: Si verifica quando un analista o un sistema

automatizzato privilegia determinate fonti a scapito di altre. Ad esempio, un'analisi sul *sentiment* politico globale basata prevalentemente su dati di Twitter in lingua inglese ignorerà le conversazioni che avvengono su piattaforme come Telegram, VK (in Russia) o Weibo (in Cina), fornendo una visione del mondo parziale e pericolosamente distorta. La tendenza a fare affidamento su fonti ad alto volume e di facile accesso può portare a trascurare fonti di nicchia che potrebbero contenere informazioni utili.

- Bias di piattaforma: Gli strumenti che utilizziamo non sono neutrali. Senza intervento dell'utente, gli algoritmi dei motori di ricerca come Google e delle piattaforme social personalizzano i risultati in base alla cronologia di ricerca, alla posizione e al profilo dell'utente. Ciò può creare "bolle di filtraggio" che possono ingannare un analista, facendogli credere che una certa narrazione sia molto più diffusa o importante di quanto non sia in realtà.

Un analista potrebbe vedere un trend nel suo feed localizzato e concludere erroneamente che si un fenomeno globale.

- Bias culturale e linguistico: L'affidamento a traduzioni automatiche senza una profonda comprensione del contesto culturale è una fonte comune di errori. Le macchine possono faticare a interpretare correttamente il tono, il sarcasmo, le espressioni idiomatiche o i riferimenti culturali, portando a gravi fraintendimenti. Ad esempio, un post satirico in una lingua straniera potrebbe essere erroneamente interpretato come una minaccia violenta dopo una traduzione letterale.

Le conseguenze di questi bias sono gravi: vanno dalla profilazione ingiusta di individui o gruppi, all'errata attribuzione di attacchi informatici, fino all'adozione di decisioni strategiche basate su un'Intelligence fondamentalmente viziata, con potenziali ripercussioni di natura legale e reputazionale.

Le "allucinazioni" dell'IA

I modelli di IA generativa, come i Large Language Models (LLM), hanno una tendenza intrinseca a produrre informazioni completamente false o fabbricate, presentandole con la stessa sicurezza e coerenza stilistica di un'informazione veritiera. Questo fenomeno, noto come "allucinazione", non è un semplice bug da correggere, ma una caratteristica fondamentale del loro funzionamento probabilistico, basato sulla predizione della parola più probabile in una sequenza. Molti esperti ritengono che questo problema sia, in una certa misura, "non risolvibile" in modo definitivo, e ciò rappresenta una minaccia diretta e persistente all'integrità di qualsiasi attività OSINT.

Un'allucinazione può innescare una catena catastrofica di decisioni errate. Un esempio lampante, anche se al di fuori dell'OSINT, ha visto un avvocato statunitense utilizzare ChatGPT per la sua ricerca legale. Il modello ha "inventato" con grande disinvoltura una serie di

precedenti giudiziari inesistenti, citando casi e sentenze mai avvenuti. L'avvocato, fidandosi dell'output, ha presentato questi "fatti" in tribunale, subendo gravi sanzioni professionali e un enorme danno reputazionale. Trasposto nel mondo dell'Intelligence, un LLM potrebbe "allucinare" un collegamento tra un individuo e un gruppo terroristico, inventare dettagli su un piano di attacco o creare un profilo aziendale fittizio, portando gli analisti e i decisori su una strada completamente sbagliata.
Inoltre, questa vulnerabilità può essere sfruttata attivamente da attori ostili. Attraverso tecniche di "avvelenamento dei dati" (data poisoning), un avversario può deliberatamente inserire informazioni false nelle fonti aperte che sa essere monitorate dai sistemi di IA, con l'obiettivo di ingannarli e indurli a produrre Intelligence errata.

L'erosione del Pensiero Critico: Il rischio della dipendenza

Forse la debolezza più insidiosa e a lungo termine dell'IA-OSINT non è di natura tecnologica, ma umana: l'erosione del pensiero critico dell'analista. Uno studio condotto da ricercatori della Carnegie Mellon University e di Microsoft ha rivelato un pattern preoccupante: maggiore è la fiducia che i professionisti ripongono negli strumenti di IA generativa, minore è il loro sforzo cognitivo e la loro tendenza a pensare in modo critico.

Gli analisti rischiano quindi di trasformarsi in semplici "operatori di automazione". Invece di formulare ipotesi proprie, le chiedono all'IA. Invece di verificare meticolosamente le fonti, assumono che l'IA lo abbia già fatto. Invece di costruire un quadro complesso da frammenti di informazione, si limitano a modificare e integrare il riassunto fornito dal modello.

Ciò può creare un circolo vizioso pericoloso: l'analista si fida del riassunto, smette di consultare le fonti grezze e, di conseguenza, perde la capacità di individuare le sottigliezze, le sfumature, le incongruenze o gli errori che l'IA, inevitabilmente, commette.
L'arte dell'Intelligence, che si fonda sullo scetticismo metodico, sulla verifica incrociata e sul giudizio contestuale, rischia così di atrofizzarsi e l'eccessiva dipendenza dall'IA può portare a una perdita di integrità del processo analitico. I fallimenti in questo scenario non derivano da cattive intenzioni, ma da un eccesso di fiducia in strumenti che sono "abbastanza buoni da sembrare affidabili, e abbastanza fallaci da essere pericolosi". In definitiva, la più grande debolezza dell'IA-OSINT non risiede nel potenziale fallimento del modello, ma nella fallibilità dell'analista che si fida troppo acriticamente del modello stesso.

Ciò sposta il focus del problema ***dalla "correzione dell'IA" alla "formazione dell'analista", rendendo l'investimento in "alfabetizzazione all'IA" e in training sul pensiero critico tanto importante quanto l'investimento in nuove tecnologie.*** *Le organizzazioni di Intelligence devono instillare una cultura in cui l'output dell'IA è trattato come una fonte grezza, non verificata, e mai come un oracolo infallibile.*

Torna all'indice

OPPORTUNITÀ: NUOVI ORIZZONTI PER L'INTELLIGENCE DA FONTI APERTE

Nonostante le sue debolezze intrinseche, l'Intelligenza Artificiale apre frontiere operative straordinarie per l'OSINT, creando opportunità senza precedenti per affrontare alcune delle sfide più complesse del nostro tempo. Queste opportunità spaziano dal contrasto alla disinformazione alla gestione delle crisi umanitarie, dalla lotta alla criminalità globale alla democratizzazione dell'accesso a capacità investigative avanzate.

Contrasto alla disinformazione e ai Deepfake

L'IA si presenta come un'arma a doppio taglio nella lotta alla disinformazione: così come può essere usata per creare contenuti falsi, può anche essere lo strumento più potente per rilevarli. Algoritmi avanzati di *Computer Vision e Natural Language Processing* (NLP) sono in grado di analizzare testi, immagini e video su larga scala per identificare manipolazioni, contenuti sintetici (noti come deepfake) e le

tracce di campagne di influenza coordinate. Nel rilevamento dei deepfake, l'IA può individuare artefatti e incongruenze spesso impercettibili all'occhio umano. Questi includono anomalie nell'illuminazione, riflessi innaturali nelle pupille degli occhi, movimenti facciali non del tutto fluidi, o piccole imperfezioni nel rendering dei capelli o dei denti. Sebbene nessun singolo strumento sia infallibile, l'analisi automatizzata può segnalare contenuti sospetti per un'ulteriore verifica umana.

Ancora più promettente è l'emergere di strumenti di analisi narrativa. Questi sistemi "*narrative-aware*" vanno oltre la semplice analisi superficiale del linguaggio. Sono addestrati a mappare le strutture narrative complesse, a tracciare le "personas" degli account (valutando se un'identità sembra autentica o fabbricata), a identificare la ripetizione di storyline simili su più piattaforme e a decodificare i riferimenti culturali per capire se una campagna è progettata per risuonare con un pubblico specifico. Ciò permette di passare dal rilevamento del singolo contenuto falso all'identificazione dell'intera campagna

di disinformazione orchestrata. Nonostante la crescente sofisticazione dei falsi, ci sono stati importanti successi. Il famoso video deepfake del Presidente ucraino Volodymyr Zelensky che ordinava la resa delle sue truppe all'inizio dell'invasione russa, sebbene diffuso ampiamente, è stato rapidamente smascherato dalla comunità OSINT e dalle piattaforme stesse, grazie alla sua scarsa qualità e all'immediata reazione di verifica. Gruppi investigativi come Bellingcat testano e utilizzano attivamente strumenti per l'analisi e la verifica di immagini, contribuendo a stabilire le migliori pratiche. Tuttavia bisogna essere consapevoli che l'efficacia di questi strumenti può variare notevolmente a seconda della qualità e della compressione delle immagini.

Gestione delle crisi e risposta umanitaria

L'OSINT potenziato dall'IA potrebbe rivelarsi uno strumento importante, se non un vero e proprio "salvavita", nella gestione dei disastri e nella risposta umanitaria. In situazioni di crisi, dove le informazioni tempestive e accurate

sono fondamentali, l'IA può fornire una consapevolezza situazionale in tempo quasi reale.
Durante disastri naturali come incendi, terremoti o uragani, i sistemi di IA possono analizzare simultaneamente migliaia di post sui social media, articoli di notizie e immagini satellitari. Ciò permette di mappare le aree più colpite, valutare l'entità dei danni, identificare infrastrutture critiche danneggiate (come strade e ponti) e persino localizzare persone che chiedono aiuto, guidando le squadre di soccorso in modo più efficiente. In zone di conflitto, l'AI-OSINT offre capacità di monitoraggio senza precedenti per le organizzazioni non governative (ONG) e le agenzie internazionali. Può essere utilizzata per monitorare i movimenti delle truppe, verificare e documentare potenziali crimini di guerra analizzando video e immagini geolocalizzate e valutare le necessità umanitarie della popolazione civile analizzando le comunicazioni online.

L'opportunità più grande in questo ambito non è tanto l'analisi di eventi passati, quanto la capacità di prevedere e prevenire crisi future, passando da una costosa risposta reattiva a una molto più efficace attività di prevenzione proattiva.

Lotta alla criminalità finanziaria

Nel settore finanziario, l'IA-OSINT è diventata uno strumento indispensabile per le indagini sui crimini finanziari, in particolare per l'antiriciclaggio (Anti-Money Laundering - AML) e il contrasto al finanziamento del terrorismo (Counter-Terrorism Financing - CTF). L'*Enhanced Due Diligence* (EDD) è un processo in cui le istituzioni finanziarie devono indagare a fondo sui clienti ad alto rischio. L'IA automatizza e potenzia questo processo, scansionando in pochi secondi un'enorme quantità di fonti aperte — registri aziendali, archivi di notizie, social media, forum — per identificare "*adverse media*" (notizie negative), collegamenti non dichiarati con persone politicamente esposte (PEP) e altri rischi reputazionali o di conformità.

L'**analisi di rete per la finanza** è un'altra applicazione chiave. Strumenti di visualizzazione e analisi di grafi, alimentati da algoritmi di IA, possono mappare reti societarie complesse e flussi finanziari internazionali. Ciò permette di scoprire strutture opache come le società di comodo (*shell companies*), identificare i beneficiari effettivi (*ultimate beneficial owners*) e tracciare transazioni sospette che, prese singolarmente, potrebbero sembrare legittime, ma che, viste nel loro insieme, rivelano schemi di riciclaggio.

Infine, il monitoraggio del dark web è un altra area di estrema rilevanza. L'IA può scandagliare costantemente forum criminali, marketplace illegali e canali di comunicazione criptati per tracciare la vendita di dati finanziari rubati, strumenti di hacking, ecc..., fornendo un'Intelligence proattiva che può prevenire frodi e attacchi.

Democratizzazione dell'Analisi e *citizen journalism*

Una delle opportunità più significative a livello sociale è la capacità dell'IA di rendere strumenti di analisi OSINT, un tempo riservati alle agenzie di Intelligence governative, accessibili a una gamma molto più ampia di attori. Giornalisti investigativi, ricercatori accademici, organizzazioni per i diritti umani e persino singoli cittadini possono oggi sfruttare queste tecnologie per le loro indagini.

Svariati collettivi hanno dimostrato la straordinaria potenza dell'OSINT collaborativo per indagare su crimini di guerra, disastri aerei e operazioni di Intelligence statali. Riconoscendo la sfida di trovare e utilizzare gli strumenti giusti, Bellingcat ha recentemente lanciato un nuovo "Online Investigations Toolkit" che integra un'interfaccia di ricerca in linguaggio naturale basata su tecnologia OpenAI.

Uno strumento che permette ai ricercatori, anche a quelli meno esperti, di descrivere semplicemente il loro obiettivo investigativo ("trovare la posizione di una foto", "verificare un video") e ricevere suggerimenti sugli strumenti più adatti, democratizzando di fatto l'accesso alla ricerca OSINT.
Questa democratizzazione sta creando un nuovo ecosistema di Intelligence "ibrido", in cui attori statali, aziende private e la società civile collaborano, competono e si controllano a vicenda. Le agenzie governative si affidano sempre più a partnership con il settore privato per l'analisi dei dati, mentre le ONG usano l'IA per monitorare le azioni e i governi e delle grandi corporation. Ciò crea un panorama dell'Intelligence più frammentato, dinamico e meno centralizzato, in cui la collaborazione inter-settoriale diventa essenziale per affrontare sfide globali come la disinformazione o il cambiamento climatico.

Torna all'indice

MINACCE: LA STRUMENTALIZZAZIONE DELL'IA E L'INDEBOLIMENTO DELLA FIDUCIA

L'enorme potenziale dell'Intelligenza Artificiale nell'OSINT è intrinsecamente legato a minacce altrettanto significative. La stessa tecnologia che potenzia la difesa e l'analisi può essere, e viene, strumentalizzata da attori ostili. Questa dualità, unita all'impatto sulla privacy e a un quadro normativo ancora incerto, rappresenta una delle sfide più critiche per la sicurezza e la stabilità nell'era digitale.

L'avversario potenziato dall'IA: Una minaccia asimmetrica

Come anticipato in apertura, l'IA è una "lama a doppio taglio". Ogni progresso tecnologico che migliora le capacità difensive di OSINT è immediatamente disponibile anche per gli avversari. Attori statali, gruppi terroristici, organizzazioni criminali e persino singoli individui malintenzionati possono sfruttare l'IA per rendere i loro attacchi più efficienti, scalabili, mirati e difficili da rilevare.

La **ricognizione automatizzata** è una delle applicazioni più immediate. Gli avversari possono utilizzare strumenti AI-OSINT per condurre una ricognizione approfondita sui loro obiettivi, che si tratti di individui, aziende o infrastrutture critiche. Possono profilare le vittime per attacchi di social engineering altamente personalizzati e convincenti, identificare vulnerabilità nei sistemi informatici e raccogliere informazioni strategiche per pianificare operazioni offensive. Report di Intelligence di fonti autorevoli come Google e Microsoft hanno confermato che attori statali affiliati a paesi come la Cina (con gruppi noti come Charcoal Typhoon e Salmon Typhoon), l'Iran (Crimson Sandstorm) e la Corea del Nord (Emerald Sleet) stanno già utilizzando attivamente i Large Language Models (LLM) per supportare le loro attività malevole. Questi usi includono la ricerca di vulnerabilità software, la generazione di codice per strumenti di hacking, la traduzione e la creazione di contenuti per campagne di influenza e la comprensione di tecnologie di difesa.

Un aspetto particolarmente preoccupante è l'abbassamento della soglia di ingresso. In passato, condurre operazioni di spionaggio o di influenza complesse richiedeva risorse significative e competenze tecniche specialistiche. L'IA generativa e altri strumenti AI-OSINT abbassano drasticamente questa soglia, mettendo capacità offensive sofisticate nelle mani di un numero molto più ampio di attori, aumentando così la superficie complessiva della minaccia.
L'Intelligenza Artificiale generativa ha rivoluzionato la produzione di disinformazione. È ora possibile creare contenuti falsi — testi, articoli, immagini, audio e video deepfake — su una scala di massa, a costi irrisori e con un livello di realismo che li rende sempre più difficili da distinguere dalla realtà. Queste campagne non mirano solo a diffondere singole falsità, ma perseguono obiettivi strategici più ampi: erodere la fiducia nelle istituzioni democratiche (governi, media, sistema giudiziario), polarizzare le società, fomentare disordini e influenzare l'esito di processi elettorali.

L'uso di deepfake e contenuti generati dall'IA è stato ampiamente documentato in contesti geopolitici critici. Durante le elezioni a Taiwan, in India e negli Stati Uniti, sono stati utilizzati per creare messaggi audio e video falsi di candidati, diffamare avversari politici o tentare di dissuadere gli elettori dal votare. Nel conflitto in Ucraina, la Russia ha utilizzato deepfake per impersonare il presidente Zelensky e altri funzionari, nel tentativo di minare il morale e creare confusione. Sebbene l'impatto catastrofico previsto per l'anno elettorale 2024 sia stato più limitato del previsto — grazie a una maggiore consapevolezza pubblica e alle contromisure delle piattaforme — la minaccia rimane estremamente alta e in continua evoluzione. Peraltro, il vero problema della disinformazione basata sull'IA non è rappresentato dal singolo "deepfake perfetto" che inganna tutti, ma da un fenomeno più sottile e pervasivo: l'erosione generale della fiducia nell'intero ecosistema informativo. Questo fenomeno, noto come il "dividendo del bugiardo", si verifica quando

la semplice consapevolezza che i deepfake esistono porta le persone a dubitare di tutto, comprese le informazioni autentiche. Gli attori ostili possono sfruttare questo scetticismo diffuso per screditare notizie e prove reali, etichettandole semplicemente come "false" o "generate dall'IA". *L'obiettivo finale non è far credere a una bugia specifica, ma* ***far sì che non si creda più a nulla, generando un caos informativo che paralizza il dibattito pubblico e il processo decisionale democratico.***

Sorveglianza di massa e crollo della privacy

La capacità dell'IA di raccogliere, aggregare e analizzare dati personali da una miriade di fonti aperte su una scala di massa e senza precedenti solleva enormi e fondate preoccupazioni per la privacy individuale e collettiva. Anche se i dati sono "pubblicamente disponibili", la loro aggregazione sistematica può creare profili incredibilmente dettagliati di individui, rivelando le loro abitudini,

affiliazioni, opinioni politiche, stato di salute e relazioni personali, spesso senza che questi ne abbiano la minima consapevolezza o abbiano fornito un consenso esplicito per tale analisi. Ciò può portare a una forma di "sorveglianza invisibile" che erode progressivamente l'anonimato, l'autonomia personale e la libertà di espressione. Il timore di essere costantemente monitorati potrebbe altresì indurre le persone ad autocensurarsi limitando la loro partecipazione al dibattito pubblico. Altro aspetto rilevante riguarda il principio spesso invocato secondo cui "se è pubblico, è lecito"; principio in realtà sia eticamente che legalmente molto problematico. La disponibilità pubblica di un singolo dato (un post, una foto) non equivale al consenso per la sua raccolta sistematica, la sua correlazione con altri dati e la sua analisi per scopi di Intelligence o commerciali, specialmente quando ciò può avere conseguenze negative tangibili sulla vita di una persona (es. negazione di un lavoro). Vds OSINT: strumenti legali per aziende e investigatori privati

Il vuoto normativo, etico e di responsabilità

Lo sviluppo e l'implementazione della tecnologia AI-OSINT stanno procedendo a una velocità vertiginosa, superando di gran lunga la capacità dei quadri legali, etici e normativi di tenere il passo. Ciò crea un pericoloso "vuoto" che lascia molte questioni critiche senza risposta.

Dal punto di vista legale, l'AI ACT cerca di definire regole comuni per i sistemi di IA ad alto rischio cosi come il Regolamento Generale sulla Protezione dei Dati (GDPR) dell'Unione Europea impone principi rigorosi come la liceità, la correttezza, la trasparenza, la limitazione delle finalità e la minimizzazione dei dati. Tuttavia, questi principi sono spesso in conflitto con la natura "affamata di dati" dei modelli di IA e con la pratica della raccolta informativa su larga scala da fonti eterogenee. Applicare le norme alle attività AI-OSINT è una sfida complessa e ritengo ancora irrisolta.

Dal punto di vista della responsabilità (accountability), il problema è ancora più grave. Quando un sistema AI-OSINT commette un errore — ad esempio, identifica erroneamente un individuo come un terrorista o fornisce un'analisi errata che porta a una decisione sbagliata — stabilire chi sia responsabile è estremamente difficile. La colpa è dello sviluppatore del modello, del fornitore dei dati di addestramento, dell'organizzazione che ha implementato lo strumento o dell'analista umano che si è fidato del suo output? Questa "diffusione della responsabilità" rende quasi impossibile per le vittime ottenere giustizia e riparazione per i danni subiti e ostacola anche la creazione di meccanismi di controllo efficaci. Questo vuoto normativo crea un "selvaggio west" digitale, dove gli attori con meno scrupoli etici e legali e considerazioni etiche sulla privacy, altri attori operano senza tali restrizioni, sfruttando appieno la potenza dell'IA per la sorveglianza e lo spionaggio.

Questa asimmetria crea uno squilibrio nella competizione internazionale anche in ambito Intelligence, con potenziali gravi conseguenze per la sicurezza nazionale delle democrazie.

Torna all'indice

RACCOMANDAZIONI STRATEGICHE: GOVERNARE LA RIVOLUZIONE IA-OSINT

L'integrazione dell'Intelligenza Artificiale nell'Open-Source Intelligence rappresenta una delle trasformazioni più profonde e ambivalenti nel campo della sicurezza e dell'analisi informativa. La tecnologia offre una potenza senza precedenti, ma introduce parallelamente rischi esistenziali per l'accuratezza, l'etica e la stabilità democratica. La sfida principale non è più di natura tecnologica, ma di governance: come imbrigliare questa potenza, massimizzandone i benefici e mitigandone i pericoli.

Sintesi dell'Analisi SWOT: Bilanciare potenza e pericoli

L'analisi condotta in questo report evidenzia un quadro complesso di forze contrapposte. Da un lato, i punti di forza dell'IA — velocità, scala, riconoscimento di pattern e capacità predittiva — stanno rivoluzionando l'efficacia dell'OSINT.

Dall'altro lato, i punti di debolezza — l'opacità della "black box", i bias algoritmici, le allucinazioni e l'erosione del pensiero critico — minano le fondamenta stesse della fiducia e dell'affidabilità. Le opportunità sono immense, spaziando dal contrasto alla disinformazione alla gestione delle crisi e alla lotta contro la criminalità globale. Tuttavia, le minacce sono altrettanto gravi, con l'uso malevolo da parte di avversari, la sorveglianza di massa potenziale e un pericoloso vuoto normativo che rischia di favorire attori senza scrupoli etici.

Tabella 2 - Risultati chiave dell'analisi SWOT.

Punti di Forza	Punti di Debolezza
Automazione su larga scala: Processa volumi di dati insostenibili per l'uomo, superando il sovraccarico informativo.	**Problema della "Black Box":** I processi decisionali dei modelli complessi sono opachi, minando fiducia e accountability.
Riconoscimento di pattern: Identifica correlazioni, reti e anomalie nascoste, rivelando insight invisibili all'analisi umana.	**Bias algoritmico:** Eredita e amplifica i pregiudizi presenti nei dati, portando a risultati distorti e discriminatori.
Analisi Predittiva: Sposta l'Intelligence da un approccio reattivo a uno proattivo, anticipando minacce e crisi.	**Allucinazioni dell'IA:** Genera informazioni false con apparente sicurezza, minacciando l'integrità delle analisi.
Velocità e tempo reale: Riduce drasticamente il tempo tra la raccolta dei dati e la produzione di Intelligence azionabile.	**Erosione del Pensiero Critico:** L'eccessiva dipendenza dall'IA rischia di atrofizzare le capacità di analisi e verifica degli operatori umani.
Opportunità	**Minacce**
Contrasto alla Disinformazione: Sviluppo di strumenti avanzati per il rilevamento di deepfake e campagne di influenza.	**Uso avversario dell'IA:** Attori ostili (statali, criminali) sfruttano l'IA per la ricognizione e per attacchi più sofisticati.
Gestione delle crisi: Fornisce consapevolezza situazionale in tempo reale per la risposta a disastri e conflitti.	**Disinformazione su scala globale:** Produzione a basso costo e su larga scala di contenuti falsi per destabilizzare le società.
Lotta alla criminalità: Potenzia le indagini su riciclaggio di denaro, finanziamento del terrorismo e frodi complesse.	**Sorveglianza di massa ed erosione della Privacy:** La raccolta e analisi massiva di dati pubblici erode i diritti fondamentali.
Democratizzazione dell'Analisi: Rende strumenti investigativi avanzati accessibili a giornalisti, ONG e società civile.	**Vuoto normativo ed etico:** La tecnologia si sviluppa più velocemente delle leggi, creando un'area grigia priva di regole e responsabilità.

L'imperativo della spiegabilità (XAI): ricostruire la fiducia

Per superare la crisi di fiducia generata dal problema della "black box", è fondamentale investire massicciamente nella ricerca, nello sviluppo e nell'adozione di sistemi di Explainable AI (XAI). Le organizzazioni che utilizzano l'IA per l'OSINT devono pretendere un livello minimo di trasparenza dai fornitori di tecnologia. L'obiettivo dovrebbe essere quello di integrare modelli che siano "interpretabili per progettazione" (interpretability by design), piuttosto che affidarsi a spiegazioni parziali e applicate a posteriori (post-hoc). L'XAI non è un lusso accademico, ma un prerequisito essenziale per garantire la responsabilità legale, la fiducia operativa e la capacità degli analisti di validare e contestualizzare i risultati dell'IA.

Verso un quadro normativo globale: stabilire regole comuni

Il già citato "selvaggio west" digitale, creato dal rapido sviluppo dell'IA, richiede una risposta normativa coordinata. È necessaria una cooperazione internazionale tra governi, industria tecnologica e società civile per sviluppare standard legali ed etici condivisi. Questi quadri normativi devono trovare un equilibrio tra le legittime esigenze di sicurezza nazionale e la protezione dei diritti fondamentali, in particolare la privacy e la libertà di espressione. Le leggi esistenti devono essere aggiornate e interpretate per affrontare specificamente le sfide poste dall'IA. Inoltre, è fondamentale istituire meccanismi di audit e supervisione indipendenti per verificare la conformità dei sistemi AI a tali regole e per fornire un percorso di ricorso per coloro che potrebbero essere danneggiati da decisioni algoritmiche errate. In questo senso un passo significativo è la ISO/IEC 42001, che rappresenta il primo standard internazionale per la gestione dei sistemi di intelligenza artificiale.

Il ruolo insostituibile dell'Analista: promuovere l'Intelligenza aumentata

La conclusione più importante di questa analisi è che il futuro dell'Intelligence non risiede nella sostituzione dell'uomo con la macchina, ma in una profonda e sinergica collaborazione. L'analista deve rimanere saldamente al centro del processo decisionale, come supervisore e validatore finale. Per realizzare questo modello di "intelligenza aumentata", è necessario un cambiamento radicale nella formazione e nelle competenze degli analisti. Le nuove competenze chiave includono:

- Alfabetizzazione all'IA: Una comprensione profonda di come funzionano i modelli di IA, dei loro limiti intrinseci, delle loro vulnerabilità e dei tipi di bias a cui sono soggetti.
- Pensiero Critico e metodico scetticismo: La capacità di trattare l'output di un sistema di IA non come una conclusione

definitiva, ma come un punto di partenza per l'indagine. La verifica umana incrociata di ogni informazione critica rimane un'abilità fondamentale e non negoziabile.

- Prompt engineering: L'abilità di interrogare e guidare efficacemente i modelli di IA per ottenere risultati pertinenti, accurati e utili, minimizzando al contempo i rischi di allucinazioni e bias.

L'obiettivo finale è un ecosistema in cui l'IA gestisce la scala, la velocità e la complessità dei dati, mentre l'essere umano fornisce il contesto, il giudizio etico, la creatività investigativa e, in ultima analisi, la decisione strategica. Solo governando attivamente questa rivoluzione sarà possibile sfruttarne l'immenso potenziale per un mondo più sicuro e informato, senza sacrificare i principi fondamentali su cui si basano le società aperte.

Tabella 3- Strategie di mitigazione per affrontare i principali rischi identificati.

Rischio Identificato	Strategia di mitigazione tecnologica	Strategia di mitigazione procedurale/organizzativa	Strategia di mitigazione formativa
Bias Algoritmico	Utilizzare dataset di addestramento diversificati e rappresentativi. Implementare strumenti per misurare e correggere il bias.	Condurre audit regolari sui bias dei modelli. Adottare procedure che impongano l'uso di fonti multiple e culturalmente diverse.	Formazione continua sulla consapevolezza dei diversi tipi di bias (selezione, piattaforma, culturale) e su come riconoscerli.
Allucinazioni dell'IA	Impiegare architetture come RAG (Retrieval-Augmented Generation) che basano le risposte su fonti verificabili. Integrare sistemi XAI per tracciare le fonti.	Istituire un obbligo di verifica umana incrociata per ogni output critico. Implementare protocolli di fact-checking a più livelli.	Addestrare gli analisti a non fidarsi ciecamente delle risposte dell'IA, a riconoscere i segnali di allucinazione e a verificare sempre le fonti primarie.
Erosione del Pensiero Critico	Progettare sistemi "human-in-the-loop" che richiedano l'intervento e l'approvazione umana nei punti critici del processo.	Sviluppare procedure che richiedano all'analista di documentare la propria catena di ragionamento, separata da quella dell'IA.	Investire in formazione continua sul pensiero critico, sulla logica investigativa e sull'Osint tradizionale, applicandola all'ambiente digitale.
Violazione della Privacy	Adottare tecniche di anonimizzazione e pseudonimizzazione dei dati. Implementare principi di "data minimization by design".	Adottare framework di valutazione d'impatto sulla privacy (es. OPIF). Garantire la stretta aderenza a normative come il GDPR e AI ACT.	Formazione obbligatoria su etica, protezione dei dati e sui quadri legali pertinenti per tutti gli operatori OSINT.