Guida: Come rilevare i contenuti generati dall'IA

Maria Cattini | 30/03/2026 | Open source intelligence

I tool sbagliano fino al 30-40% dei casi: ecco cosa guardano davvero gli investigatori digitali. Nel marzo 2026, il problema della verifica dei contenuti è diventato più complesso che mai. Sebbene gli strumenti di rilevamento siano aumentati, è emersa una nuova sfida: la nascita di **siti di "fact-checking" sponsorizzati dagli Stati** (come il GFCN russo, legato all'agenzia TASS) che utilizzano l'apparenza della verifica per diffondere narrazioni governative. Oggi, la manipolazione politica è la ragione principale dietro la creazione di falsi generati dall'IA.

Per navigare in questo scenario, non basta affidarsi alla tecnologia; è necessario un approccio che combini strumenti automatici e intuito umano.

1. La Regola d'Oro: Il "Rilevatore" non è un "Investigatore"

Le fonti sottolineano che circa il **60-70% del lavoro di verifica rimane umano**. Gli strumenti automatici (detectors) come Hive possono fornire indizi, ma spesso operano come "scatole nere" e possono fornire risultati contrastanti o errati su uno stesso file.

• Strategia: Considera il risultato di un software di rilevamento solo come una prova "A" o "B", non come un verdetto finale. Il tuo ruolo è quello del detective, non solo dell'utente di un software.

2. Strategie di Verifica Visiva e Forense

Per smascherare immagini o video contraffatti, segui questi passaggi:

• Cerca anomalie fisiche: Oltre al classico problema delle dita (che possono essere sette invece di cinque), osserva attentamente l'illuminazione, le ombre e la coerenza degli oggetti (ad esempio, un laptop che non potrebbe chiudersi correttamente o ombre che puntano in direzioni sbagliate).
• Analisi fotogramma per fotogramma: Non limitarti a guardare il video nel suo insieme; analizzalo frame dopo frame per individuare incongruenze nei loghi o nei dettagli dello sfondo.
• Ricerca della massima risoluzione: Prima di analizzare un'immagine, cerca sempre la versione con la risoluzione più alta possibile. Spesso, i dettagli che rivelano la falsità sono visibili solo nell'originale di alta qualità.
• Falsificazione delle ipotesi: Non cercare solo prove a favore della tua tesi. Chiediti: "Se questa immagine fosse reale, cosa dovrebbe accadere?". Ad esempio, un anello che scompare in un video potrebbe non essere IA, ma un difetto tecnico di rifrazione di alcuni modelli di smartphone.

3. Usare l'IA come Alleata (ma con cautela)

L'IA non è solo il nemico; può essere usata per generare piste investigative (leads).

• Domande Neutre: Quando chiedi a un chatbot (come ChatGPT o Claude) di analizzare un documento sospetto, usa un linguaggio neutro e privo di aggettivi. Non dire "dimmi cosa c'è di sbagliato", ma chiedi "descrivi la funzione di ogni frase".
• Scegliere il modello giusto: Le fonti suggeriscono che modelli più "freddi" e precisi come Claude sono spesso più utili per individuare tracce di testi generati da IA in documenti ufficiali.

• Espansione per la geolocalizzazione: Una tecnica innovativa consiste nell'usare l'IA per "zoomare fuori" (outpainting) da una foto di bassa qualità. L'IA può ricostruire l'ambiente circostante in modo logico, permettendo poi di effettuare una ricerca per immagini inversa più efficace per trovare il luogo reale.

4. Strumenti Specifici e Watermarking

• SynthID: È lo strumento di Google per inserire e rilevare filigrane invisibili nelle immagini e nei video generati dalle proprie IA. Sebbene utile, ricorda che rileva solo i contenuti creati tramite prodotti Google.
• Analisi Satellitare: Strumenti come "Wayback Imagery" combinati con l'IA possono aiutare a monitorare i cambiamenti nel tempo di un'area geografica, aiutando a verificare se un edificio mostrato in una foto esiste davvero o è stato rimosso.

Il rilevamento dell'IA richiede di **mettere costantemente in discussione se stessi** e le proprie conclusioni. In un'epoca di strumenti economici e accessibili per clonare voci o creare crash simulati in pochi secondi, la verifica incrociata (check, check e double check) rimane l'unica difesa efficace.

E se vuoi approfondire davvero:

Iscriviti alla newsletter: https://coondivido.substack.com

Entra nella community Telegram: https://t.me/osintaipertutti
I tool sbagliano fino al 30-40% dei casi: ecco cosa guardano davvero gli investigatori digitali. Nel marzo 2026, il problema della verifica dei contenuti è diventato più complesso che mai. Sebbene gli strumenti di rilevamento siano aumentati, è emersa una nuova sfida: la nascita di **siti di "fact-checking" sponsorizzati dagli Stati** (come il GFCN russo, legato all'agenzia TASS) che utilizzano l'apparenza della verifica per diffondere narrazioni governative. Oggi, la manipolazione politica è la ragione principale dietro la creazione di falsi generati dall'IA.

Per navigare in questo scenario, non basta affidarsi alla tecnologia; è necessario un approccio che combini strumenti automatici e intuito umano.

1. La Regola d'Oro: Il "Rilevatore" non è un "Investigatore"

Le fonti sottolineano che circa il **60-70% del lavoro di verifica rimane umano**. Gli strumenti automatici (detectors) come Hive possono fornire indizi, ma spesso operano come "scatole nere" e possono fornire risultati contrastanti o errati su uno stesso file.

• Strategia: Considera il risultato di un software di rilevamento solo come una prova "A" o "B", non come un verdetto finale. Il tuo ruolo è quello del detective, non solo dell'utente di un software.

2. Strategie di Verifica Visiva e Forense

Per smascherare immagini o video contraffatti, segui questi passaggi:

• Cerca anomalie fisiche: Oltre al classico problema delle dita (che possono essere sette invece di cinque), osserva attentamente l'illuminazione, le ombre e la coerenza degli oggetti (ad esempio, un laptop che non potrebbe chiudersi correttamente o ombre che puntano in direzioni sbagliate).
• Analisi fotogramma per fotogramma: Non limitarti a guardare il video nel suo insieme; analizzalo frame dopo frame per individuare incongruenze nei loghi o nei dettagli dello sfondo.
• Ricerca della massima risoluzione: Prima di analizzare un'immagine, cerca sempre la versione con la risoluzione più alta possibile. Spesso, i dettagli che rivelano la falsità sono visibili solo nell'originale di alta qualità.
• Falsificazione delle ipotesi: Non cercare solo prove a favore della tua tesi. Chiediti: "Se questa immagine fosse reale, cosa dovrebbe accadere?". Ad esempio, un anello che scompare in un video potrebbe non essere IA, ma un difetto tecnico di rifrazione di alcuni modelli di smartphone.

3. Usare l'IA come Alleata (ma con cautela)

L'IA non è solo il nemico; può essere usata per generare piste investigative (leads).

• Domande Neutre: Quando chiedi a un chatbot (come ChatGPT o Claude) di analizzare un documento sospetto, usa un linguaggio neutro e privo di aggettivi. Non dire "dimmi cosa c'è di sbagliato", ma chiedi "descrivi la funzione di ogni frase".
• Scegliere il modello giusto: Le fonti suggeriscono che modelli più "freddi" e precisi come Claude sono spesso più utili per individuare tracce di testi generati da IA in documenti ufficiali.
• Espansione per la geolocalizzazione: Una tecnica innovativa consiste nell'usare l'IA per "zoomare fuori" (outpainting) da una foto di bassa qualità. L'IA può ricostruire l'ambiente circostante in modo logico, permettendo poi di effettuare una ricerca per immagini inversa più efficace per trovare il luogo reale.

4. Strumenti Specifici e Watermarking

• SynthID: È lo strumento di Google per inserire e rilevare filigrane invisibili nelle immagini e nei video generati dalle proprie IA. Sebbene utile, ricorda che rileva solo i contenuti creati tramite prodotti Google.
• Analisi Satellitare: Strumenti come "Wayback Imagery" combinati con l'IA possono aiutare a monitorare i cambiamenti nel tempo di un'area geografica, aiutando a verificare se un edificio mostrato in una foto esiste davvero o è stato rimosso.

Il rilevamento dell'IA richiede di **mettere costantemente in discussione se stessi** e le proprie conclusioni. In un'epoca di strumenti economici e accessibili per clonare voci o creare crash simulati in pochi secondi, la verifica incrociata (check, check e double check) rimane l'unica difesa efficace.

E se vuoi approfondire davvero:

Iscriviti alla newsletter: https://coondivido.substack.com

Entra nella community Telegram: https://t.me/osintaipertutti