La firma tolta

Propaganda e rilevazione di testo generato condividono lo stesso punto debole: riconoscono una firma — lo stile, l'accento, l'origine — e falliscono contro chi quella firma la toglie. Otto schede dalla Finlandia alla Casa Bianca.

CLAUDIO CAMMARANO 6 OTTOBRE 2026

FONTI	8 schede, lug 2023 – ott 2026
ARGOMENTI	AI · Geopolitica · Giornalismo
CONCETTI	controllo riflessivo · ecologia dei media · legge di Goodhart · riciclaggio informativo · rilevatori di testo generato
ORIGINE	https://claudiocammarano.com/writings/2026-10-06-la-firma-tolta/

TESI

Ogni difesa che funziona riconoscendo una firma — lo stile di una prosa, l'accento di una voce, la nazionalità di una fonte — smette di funzionare contro chi quella firma la rimuove deliberatamente; e l'asimmetria fra testo e immagine nelle elezioni americane del 2026 mostra dove la rimozione riesce meglio, cioè dove nessun rilevatore sa ancora cosa cercare.

Oleksandr Liemienov e Sofiia Maksymiv, del think tank ucraino StateWatch, pubblicano il 16 settembre 2026 un'analisi sulla propaganda russa in Finlandia che va letta sapendo chi la ospita: UNITED24 Media è il braccio editoriale della piattaforma lanciata dal presidente Zelensky.[1] È una parte in guerra che descrive i metodi dell'avversario, e questo non rende false le sue affermazioni, ma obbliga a dividerle in due classi: quelle documentate altrove, che reggono da sole, e quelle che restano un'accusa. Nella prima classe c'è il fatto che conta per questa catena: la Finlandia ha, secondo gli osservatori europei, la difesa più efficace contro la disinformazione russa, fondata sul riconoscimento di una firma — fonti note, stile riconoscibile, tracce di origine. Il caso che Liemienov e Maksymiv documentano è quello in cui quella firma viene tolta.

La difesa migliore d'Europa

MV-lehti, portale finlandese fondato nel 2014 da Ilja Janitskin e da tempo classificato come veicolo di disinformazione russa, non si presenta più come una fonte straniera. Gli autori descrivono un passaggio dal riciclaggio informativo visibile — un articolo russo tradotto e ripubblicato, riconoscibile come tale — a un'operazione che adotta il linguaggio, i riferimenti e persino le lamentele locali di chi scrive davvero in Finlandia, finché la differenza fra una voce domestica scontenta e un'operazione di origine estera smette di essere rilevabile dallo stile.

1 · Oleksandr Liemienov e Sofiia Maksymiv / UNITED24 Media · 16 set 2026 · **STRUMENTO CONCETTUALE**

How Russian Propaganda Learned to Look Finnish

Una difesa che riconosce una firma fallisce contro chi quella firma la toglie. La Finlandia è il caso perché ha la difesa migliore d'Europa.

https://united24media.com/world/how-russian-propaganda-learned-to-look-finnish-22477

Il punto non è che la Finlandia abbia smesso di difendersi bene; è che la difesa su cui ha costruito la propria reputazione — riconoscere una firma — vale solo finché l'avversario non impara a toglierla. È la stessa struttura, spostata dalla geopolitica alla prosa, che l'archivio incontra nei rilevatori di testo generato: funzionano riconoscendo uno stile, e uno stile riconosciuto è uno stile che si può imparare a non avere più.

Il caso francese dell'ottobre 2026 mostra il rovescio della stessa difficoltà. Dopo violenze e saccheggi in Francia, in Italia e in Belgio, Laura Dubois e Sam Jones, sul *Financial Times*, raccolgono gli indizi: account filorussi e di estrema destra che rilanciano video violenti, uno dei quali, visto quasi sei milioni di volte su X, veniva in realtà dalle proteste di Tbilisi del 2024.[7] Le fonti di intelligence però sono nette: gli attori stranieri amplificano una rabbia autentica, non la creano. Quando l'operazione si innesta su una rimostranza locale, non c'è una firma da togliere, perché il contenuto non l'ha mai avuta.

Lo stile come firma

Il numero più citato sui rilevatori di testo generato viene da uno studio di Stanford del luglio 2023: sette strumenti applicati a saggi TOEFL di non madrelingua li classificavano come scritti da una macchina nel 61,22 per cento dei casi, contro il 5,19 per cento dei temi di studenti americani.[2] L'esperimento più istruttivo dello studio è però un altro: arricchendo il lessico dei saggi TOEFL i falsi positivi scendevano all'11,77 per cento, e impoverendo quello dei madrelingua salivano al 56,65. Lo strumento non rilevava l'origine dell'autore; rilevava uno stile, e trattava quello stile come se fosse una firma stabile di provenienza — un accento, appunto, più che un contenuto.

Tre anni dopo quella firma si è già spostata. *the Economist*, su un corpus di 55.940 frasi, trova che l'em-dash è superato come marker — solo Claude ne usa ancora più degli scrittori umani, ChatGPT ne usa meno di chiunque — e che «delve» e «tapestry» sono spariti con gli aggiornamenti recenti.[3] Il segno reale, secondo il giornale, è la «pretentious diction» descritta da Orwell: polisillabi latini, nominalizzazioni, un'assenza di punteggiatura variata — i modelli hanno adottato i vizi della cattiva prosa accademica, non le virtù di quella buona. **Ogni volta che un marker diventa noto, cessa di essere un marker**, perché sia chi usa il modello sia chi

lo addestra può correggerlo, e la firma che resta identificabile oggi è quella che nessuno ha ancora imparato a togliere.

Nessun accento da riconoscere

Linda Hourani, per il *Kyiv Independent*, smaschera con metodologia OSINT i due conduttori anonimi di *Russians With Attitude*, podcast ultranazionalista russo con 422.000 follower e un pubblico composto per il 27,6 per cento da americani.[4] «Kirill» è un attivista di estrema destra vissuto a lungo in Germania; «Nikolay» è un blogger di Ekaterinburg. L'identità viene ricostruita incrociando account multipli, email collegate e database di consegne trapelati — cioè con lo stesso lavoro filologico, non statistico, con cui *Libération* avrebbe accertato i plagi di Orélien un mese dopo.

---

4 · Linda Hourani / Kyiv Independent · 6 apr 2026 · **ELEMENTI DI SCENARIO**

Unmasking the anonymous hosts of 'Russians With Attitude,' a pro-war podcast popular with US far right

L'architettura concreta della guerra dell'informazione russa verso la destra americana: due blogger ultranazionalisti, anonimato operativo, piattaforme americane monetizzate e raccolta fondi per unità paramilitari sanzionate.

https://kyivindependent.com/exclusive-investigation-reveals-hosts-of-us-popular-far-right-podcast/

L'architettura che Hourani rivela è la versione più pulita del problema di questa catena: due voci russe, monetizzate su piattaforme americane, che raccolgono fondi per unità paramilitari sanzionate senza mai esibire un tratto che un rilevatore automatico possa marcare come straniero. Non c'è un accento da sentire né uno stile da classificare, perché il contenuto è in inglese idiomatico, prodotto da chi quella lingua la parla davvero. La firma che manca non è linguistica; è istituzionale, ed è per questo che a toglierla il lavoro giornalistico, non il software.

Lo stesso vale per le fonti. Philip Di Salvo, su *Nieman Lab*, ha intervistato dieci giornalisti che lavorano su dati sottratti, dai documenti delle istituzioni russe agli archivi pubblicati dai gruppi ransomware, e trova un mestiere ancora senza regole comuni, in cui la verifica non basta: bisogna dichiarare che cosa non si sa della fonte, perché dietro un hacktivista può esserci uno Stato.[8] «Non sappiamo chi sia la fonte, potrebbe essere il GRU», dice uno di loro. È la firma tolta nella forma più pura, e la risposta che le redazioni stanno trovando non è un rilevatore ma una dichiarazione d'incertezza.

Dove la firma ancora regge

Lawrence Norden e il Brennan Center for Justice hanno sottoposto i principali chatbot a teorie del complotto elettorali nelle settimane prima delle elezioni di midterm 2026.[5] Sul testo, i modelli si sono mostrati resistenti: alimentati con il linguaggio reale degli *election denier* e spinti ripetutamente a confermare, hanno continuato a correggere l'informazione falsa.

---

5 · Lawrence Norden / The Washington Post · 25 ago 2026 · **DATO CHE CORREGGE**

My team fed chatbots election lies. Here's what happened.

L'asimmetria strutturale tra testo (dove i chatbot tengono) e immagine (dove sono pericolosi) è il dato che corregge la narrativa in entrambe le direzioni.

https://www.washingtonpost.com/opinions/2026/08/25/ai-chatbots-may-be-next-voter-guide-election-denialism-beware/

L'asimmetria che lo studio documenta — il testo tiene, l'immagine è il lato pericoloso — è la stessa catena che questa pagina ha seguito, letta al contrario. I chatbot resistono sul testo perché quella è l'area in cui l'addestramento ha costruito una firma difensiva riconoscibile, ripetibile, verificata: il modello «sa» di dover segnalare la disinformazione elettorale perché gliel'hanno insegnato con lo stesso tipo di addestramento esplicito che, nella catena *«Il rilevatore prima dell'oggetto»*, produce risposte sulla propria coscienza. Dove quell'addestramento manca — l'immagine, la voce sintetica, il video — non c'è firma da riconoscere né da togliere, perché non è mai stata costruita.

Controtesi 6 · Adnan Al Ali, Jindřich Helcl e Jindřich Libovický / EACL 2026

Si può obiettare che il problema di Liang sia ormai superato. Un gruppo dell'Università Carlo di Praga rifà la verifica nel febbraio 2026 e trova che la perplessità dei non madrelingua non è più bassa, che tre famiglie di rilevatori diversi non mostrano un bias sistematico, e che una verifica della Vrije Universiteit su 1.163 tesi magistrali non trova falsi positivi per i tre strumenti più usati.[6] Se gli strumenti migliorano a questa velocità, la firma tornerà riconoscibile prima che l'avversario la tolga di nuovo.

L'obiezione descrive uno strumento migliore, non un vantaggio stabile. Il caso finlandese e quello di Hourani mostrano un avversario che si adatta appena la firma precedente viene catalogata — MV-lehti ha impiegato anni a smettere di sembrare straniero, ma lo ha fatto; «Kirill» e «Nikolay» non hanno mai esibito l'accento che un rilevatore avrebbe potuto imparare a riconoscere, perché hanno scritto in una lingua che non è la loro firma. E l'asimmetria di Norden mostra la via più semplice per vincere la corsa: non migliorare la contraffazione dove la verifica è forte, ma spostarsi dove la verifica non esiste ancora. I rilevatori di testo possono migliorare quanto si vuole; il problema che risolvono resta quello di ieri, mentre la propaganda, come la disinformazione elettorale, si è già spostata sul canale dove nessuno ha ancora costruito una firma da riconoscere.

Nella catena

1. Oleksandr Liemienov e Sofiia Maksymiv / UNITED24 Media, *How Russian Propaganda Learned to Look Finnish*, 16 set 2026. Strumento concettuale.
   https://united24media.com/world/how-russian-propaganda-learned-to-look-finnish-22477
   Scheda: https://claudiocammarano.com/curated/2026-09-16-liemienov-maksymiv-propaganda-smette-di-sembrare-straniera-united24/
2. Weixin Liang, Mert Yuksekgonul, Yining Mao, Eric Wu e James Zou / Patterns, *GPT detectors are biased against non-native English writers*, 10 lug 2023. Riferimento tecnico.
   https://doi.org/10.1016/j.patter.2023.100779
   Scheda: https://claudiocammarano.com/curated/2023-07-10-liang-rilevatori-non-madrelingua-patterns/
3. The Economist, *How to spot AI writing*, 30 lug 2026. Dato che corregge.
   https://www.economist.com/culture/2026/07/30/how-to-spot-ai-writing
   Scheda: https://claudiocammarano.com/curated/2026-07-30-economist-ai-writing-

detection/

4. Linda Hourani / Kyiv Independent, *Unmasking the anonymous hosts of 'Russians With Attitude,' a pro-war podcast popular with US far right*, 6 apr 2026. Elementi di scenario.
https://kyivindependent.com/exclusive-investigation-reveals-hosts-of-us-popular-far-right-podcast/
Scheda: https://claudiocammarano.com/curated/2026-04-06-hourani-russians-with-attitude-kyivindependent/

5. Lawrence Norden / The Washington Post, *My team fed chatbots election lies. Here's what happened.*, 25 ago 2026. Dato che corregge.
https://www.washingtonpost.com/opinions/2026/08/25/ai-chatbots-may-be-next-voter-guide-election-denialism-beware/
Scheda: https://claudiocammarano.com/curated/2026-08-25-norden-chatbot-election-lies-wapo/

6. Adnan Al Ali, Jindřich Helcl e Jindřich Libovický / EACL 2026, *Different Time, Different Language: Revisiting the Bias Against Non-Native Speakers in GPT Detectors*, 5 feb 2026. Dato che corregge.
https://arxiv.org/abs/2602.05769
Scheda: https://claudiocammarano.com/curated/2026-02-05-al-ali-bias-rilevatori-rivisitato-eacl/

7. Laura Dubois e Sam Jones / Financial Times, *EU probes pro-Kremlin disinformation around French riots*, 21 ott 2026. Dato che corregge.
https://www.ft.com/content/d95609cd-1bb9-4995-937b-875277de3a42
Scheda: https://claudiocammarano.com/curated/2026-10-21-dubois-jones-disinformazione-rivolte-francia-ft/

8. Philip Di Salvo / Nieman Lab, *When hackers become sources: Inside the new craft of reporting on hacked data*, 20 ott 2026. Metodo di lavoro.
https://www.niemanlab.org/2026/10/when-hackers-become-sources-inside-the-new-craft-of-reporting-on-hacked-data/
Scheda: https://claudiocammarano.com/curated/2026-10-20-disalvo-hacker-fonti-niemanlab/

Intervento AI: WR (Scrittura). Notazione descritta in https://claudiocammarano.com/notazione/.

© Claudio Cammarano. Licenza CC BY-NC-ND 4.0: si può condividere citando l'autore e la fonte, non a fini commerciali e senza modifiche.

Versione originale e aggiornata: https://claudiocammarano.com/writings/2026-10-06-la-firma-tolta/. Le parti interattive della catena sono disponibili solo online.

Questa catena fa parte della serie *La misura e il bersaglio*: https://claudiocammarano.com/serie/la-misura-e-il-bersaglio/.

Scrivo anche *The Abstract*, in inglese, una volta al mese, sull'economia dei contenuti: https://claudiocammarano.com/newsletter/. C'è anche il feed: https://claudiocammarano.com/feed.xml.