



Riconoscere la AI

ANDREA INDIANO INTELLIGENZA ARTIFICIALE 09.09.2025

I modelli linguistici AI hanno un vizio di stile che li rende facili da identificare

Uno studio italiano mostra come gli LLM non si limitino a inventare fatti: ripetono ossessivamente la formula “non X, ma Y”, trasformando un artificio retorico in marchio di fabbrica

Quando si parla di modelli linguistici, l'attenzione si concentra quasi sempre sulle allucinazioni: date inventate, citazioni mal esistite, riferimenti errati. Un nuovo paper del ricercatore italiano Filippo Lubrano (fondatore della startup di realtà virtuale e aumentata Metaphora e membro di H-Farm) sposta lo sguardo su un piano diverso, quello dello stile. L'analisi dimostra che gli LLM non si limitano a commettere errori fattuali, ma tendono a riprodurre un ritmo linguistico ricorrente che li tradisce: si tratta della figura retorica epanortosi enfatica, la correzione immediata che prende la forma “non X, ma Y”.

Nel testi umani la figura ha un valore preciso: serve a ribaltare un'aspettativa, intensificare un concetto o introdurre un colpo di scena retorico. Nel testi generati dai modelli con AI, invece, diventa una sorta di tic narrativo, presente tre volte più che negli scritti umani. I numeri sono eloquenti: nelle risposte di ChatGPT e Claude si contano in media 27 occorrenze ogni mille frasi, contro le 5 dei testi umani e le 9 riscontrate nel dataset di addestramento. Una tendenza imperfetta che rischia di aumentare i contenuti approssimativi rilasciati dal bot e, di conseguenza, abbassare il livello delle informazioni presenti sul web.

Cos'è l'epanortosi enfatica

In generale, il funzionamento della epanortosi è semplice, quasi meccanico. Alcuni esempi di frasi così caratterizzate sono: Non è solo un caffè, è energia liquida, oppure Non è un blackout tecnico, è il segnale di una rete fragile. La struttura dà l'impressione di chiarezza e incisività, ma proprio per questo tende a essere abusata. Nel testi umani compare per sottolineare un'eccezione o una sfumatura; nei modelli linguistici si trasforma in una scorciatoia stilistica che appare in quasi ogni risposta di una certa lunghezza, soprattutto nei passaggi di transizione tra argomenti. Una regressione logistica condotta sugli output ha confermato che l'epanortosi è un predittore significativo di testo generato: quando la frequenza cresce oltre una certa soglia, è molto probabile che il contenuto non sia di origine umana.

L'epanortosi (dal greco antico ἐπανόρθωσις *epanóρθōsis*, «correzione») è una figura retorica che consiste nel modificare o anche nel capovolgere un'affermazione precedente (e risponde a calcolati effetti retorici, a differenza della correzione di un *lapsus* involontario, detta correctio).

La spiegazione sta nell'incontro tra **addestramento e ottimizzazione**. Nel dataset che alimentano gli LLM, le formule di negazione seguite da rilancio positivo sono comuni in settori come marketing, self-help e comunicazione politica. Sono testi che puntano a essere memorabili e persuasivi. A questo, spiega Lubrano, si aggiunge l'effetto del ***reinforcement learning from human feedback (RLHF)***: gli annotatori, incaricati di valutare la qualità delle risposte, tendono a premiare passaggi percepiti come chiari, assertivi e rafforzativi. Così l'epanortosi enfatica diventa non solo frequente, ma anche rinforzata artificialmente dal processo di training.

Lubrano definisce questo fenomeno *sloganoid effect*: un'apparente chiarezza che in realtà riduce la varietà espressiva. Invece di fornire nuove sfumature, i modelli ripropongono lo stesso schema binario, replicando il linguaggio della pubblicità e dei titoli acchiappaclick.

"**AI slop**", often simply "slop", is a term for low-quality media, including writing and images, made using generative artificial intelligence technology, characterized by an inherent lack of effort, being generated at an overwhelming volume. Coined in the 2020s, the term has a pejorative connotation similar to "spam".

Il rischio della lingua piatta

La questione non è estetica ma culturale. Se gli utenti leggono continuamente testi che funzionano secondo lo schema "non... ma...", finiscono per **interiorizzarlo e usarlo a loro volta**. Questo alimenta un circolo vizioso: i modelli lo apprendono dal web, lo amplificano, lo rilanciano agli utenti che a loro volta lo riversano nei contenuti online, rinforzando ulteriormente i dataset futuri. Il risultato è una lingua digitale più piatta, dominata da **dualismi e slogan**, che riduce le possibilità espressive e schiaccia le alternative retoriche.

Lo studio nota come questa dinamica non sia limitata all'Inglese: campionamenti preliminari mostrano che lo stesso eccesso compare anche in spagnolo, francese, mandarino e arabo. È un segnale che l'epanortosi potrebbe diventare una sorta di universale artificiale della scrittura generata. Per contestualizzare, Lubrano richiama la **retorica classica**, da Quintiliano a Lausberg, in cui l'epanortosi era definita come strumento di intensificazione e chiarificazione, da usare in modo mirato. Ora, per colpa dell'AI, può diventare un fenomeno diffuso che contribuisce a diminuire l'efficacia di ciò che leggiamo su Internet.

Attenzione all'AI slop

Secondo Lubrano, il fenomeno può essere contenuto. Tra le strategie proposte: inserire nel dataset più testi che usano forme diverse di enfasi (metafore, comparazioni, clausole concessive), modificare i *reward models* penalizzando la ripetizione del pattern e spingere i prompt engineer a richiedere esplicitamente soluzioni alternative. Non si tratta di cancellare l'epanortosi — resta una figura retorica utile — ma di riportarla a scelta consapevole, invece che a default algoritmico.

Il messaggio finale del paper è chiaro: concentrarsi solo sulle allucinazioni significa guardare a metà del problema. Gli LLM non modellano soltanto i contenuti, ma anche le forme. E quando una forma retorica diventa eccesso strutturale, rischia di modificare come leggiamo i testi artificiali, nonché il modo in cui scriviamo noi stessi. È in questo contesto che si inserisce anche il dibattito culturale attorno al termine **AI slop**, ovvero i pasticci delle Intelligenze artificiali sempre più frequenti online: un'etichetta che denuncia l'appiattimento linguistico prodotto da questi automatismi e che oggi diventa chiave di lettura per comprendere il rapporto tra AI e stile. Perché i modelli possono aiutarci a scrivere, ma non sempre il loro apporto è positivo.

Come cercare documentazione in rete sui tool di rilevamento AI (*AI detection tools*)



Giacomo Ciarlini  **Autore**

1 giorno

Head of Content & Education at Datapizza 🍕

Se volete letteratura rilevante

"Testing of Detection Tools for AI-Generated Text"

"Can AI-Generated Text be Reliably Detected?"

"MAGE: Machine-generated Text Detection in the Wild"

"SeqXGPT: Sentence-Level AI-Generated Text Detection"

"EAGLE: A Domain Generalization Framework for AI-generated Text Detection"

"Detecting AI-Generated Text: Factors Influencing Detectability with Current Methods"

"On the Possibilities of AI-Generated Text Detection"

novembre 2024

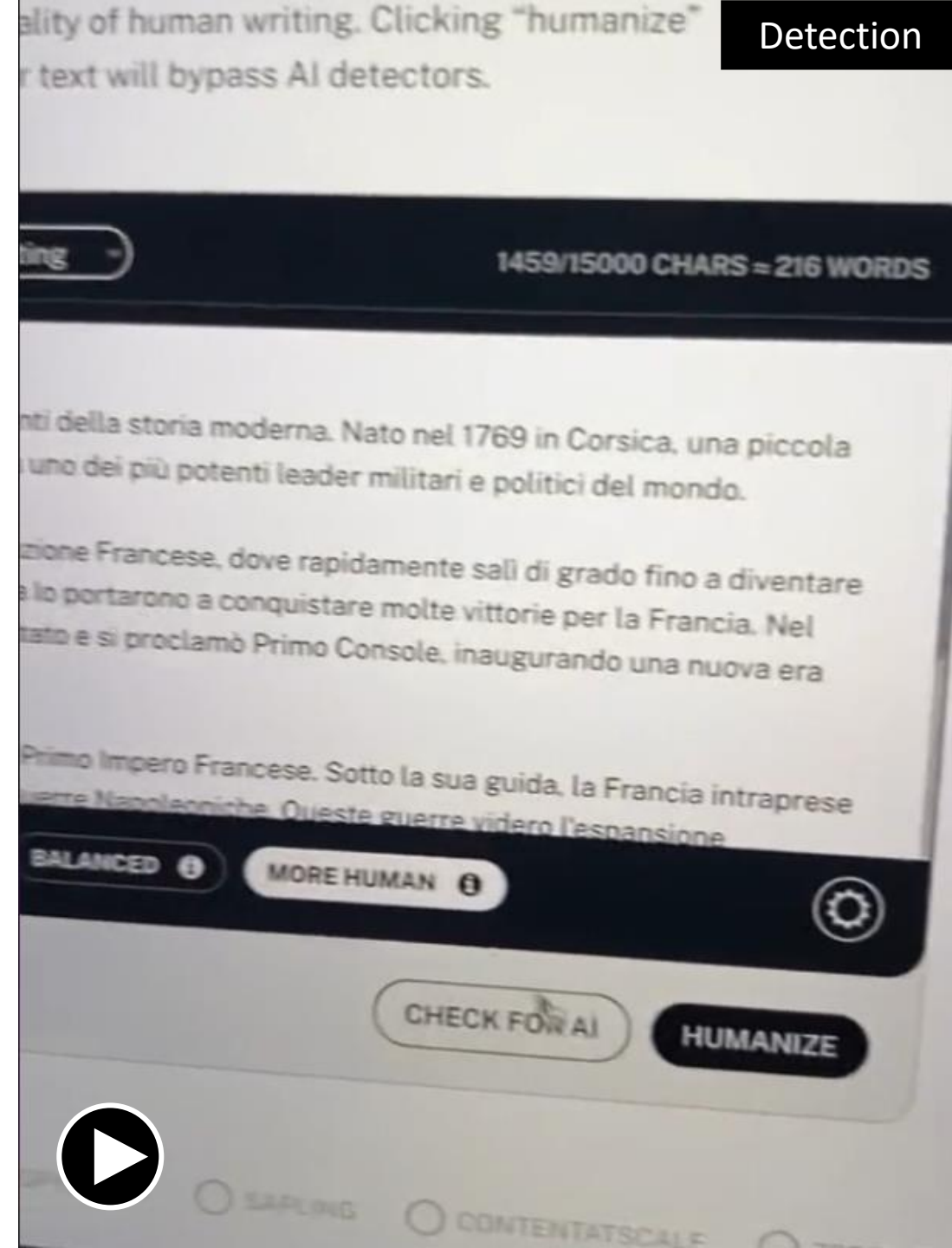
Come umanizzare un testo prodotto da un LLM?

AI Humanizer and Detector

Link al tool
«Undetectable AI»



Piccole modifiche - quali eliminare i suggerimenti finali e le emoticon, oppure riformattare il testo - sono sufficienti ad ingannarlo



University Used AI to Detect AI and Wrongly Accused 6,000 Students

ottobre 2025

Detection



Image: ABC News: Danielle Bonica)

Australian Catholic University (ACU) used AI tools to flag nearly **6,000 students** for academic misconduct in 2024, most for alleged use of generative AI. Many were later cleared, revealing how flawed detection systems can upend real lives when used without oversight or context.

How did this affect staff and students?

- **Scale of accusations:** About 90% of misconduct cases involved AI, though many students had done nothing wrong.
- **Student impact:** Some students lost internships and job offers while transcripts were marked "results withheld."
- **Faulty tools:** ACU relied on Turnitin's AI detector, which later admitted its reports "may misidentify" human writing.
- **Course correction:** The university dropped the detector and introduced training on ethical AI use for students and staff.

AI detectors are notoriously unreliable, often flagging human writing as machine output including content from the **U.S. Constitution** and the **Bible**. Turnitin's AI feature uses deep-learning models to spot content "likely" written or paraphrased by LLMs. The next step for universities is not policing algorithms with more algorithms, but teaching digital literacy, transparency, and fairness.

Come riconoscere le foto realizzate con l'intelligenza artificiale

1/3

WIRED

Dagli errori visivi ai software di rilevamento: una guida per smascherare le immagini sintetiche create con l'AI

ANDREA INDIANO

LA GUIDA 13.01.2025

Le immagini generate dall'intelligenza artificiale sono **sempre più realistiche**, ma nascondono dettagli che possono tradirne l'origine. Scoprire come riconoscerle è diventata una competenza cruciale in un mondo digitale in rapida evoluzione. Ritratti iper-realistici, panorami dettagliati e manipolazioni visive sofisticate mettono in crisi la capacità di distinguere tra ciò che è reale e ciò che è **sintetico**. Eppure, anche le immagini più convincenti lasciano tracce della loro origine con intelligenza artificiale: per ora, **fino alla prossima evoluzione** dell'innovativa tecnologia, ci sono dei consigli pratici da seguire per capire quali immagini siano vere ed evitare la diffusione di pericolose fake news correlate.

Analizzare i dettagli visivi: mani, occhi e luci

Uno degli approcci più immediati per **smascherare immagini create con AI** è l'analisi dei dettagli visivi. Le mani rappresentano uno degli elementi più problematici per le intelligenze artificiali: dita con forme distorte, proporzioni innaturali o persino un numero sbagliato di falangi sono segnali comuni. Anche i volti possono tradire anomalie, in particolare con occhi asimmetrici, riflessi incoerenti o una mancanza di naturalezza nello sguardo.



Immagine creata con intelligenza artificiale (© Lummi) LUMMI

Un altro punto debole per le AI è la luce. Le immagini sintetiche spesso presentano una **distribuzione delle ombre innaturale** o riflessi mal posizionati rispetto alla sorgente luminosa. In alcuni casi, l'intera illuminazione dell'immagine può apparire artificiale, con una qualità uniforme o troppo perfetta che non si verifica in contesti reali. Bisogna prestare particolare attenzione alle scene che sembrano sovraesposte o con contrasti sospettosamente bilanciati: la perfezione, paradossalmente, è spesso il segno rivelatore. Le **scritte presenti** nell'immagine, come loghi o cartelli, possono fornire indizi preziosi: è il caso della foto copertina di questo articolo, nella quale il cartello giallo a destra ha delle lettere non comprensibili. Le AI tendono a creare lettere distorte, errate o parole prive di senso. Un'insegna commerciale con un testo incomprensibile o un mix di caratteri strani dovrebbe immediatamente far scattare l'allarme.

Consigli pratici per riconoscere le immagini fatte con AI

Un elemento chiave per distinguere un'immagine reale da una generata con intelligenza artificiale è l'analisi del **rumore digitale**. Questo termine si riferisce alla grana visibile o alle variazioni casuali nei pixel di una foto, un effetto comune nelle immagini scattate con una fotocamera, soprattutto in condizioni di scarsa illuminazione o a sensibilità ISO elevata. Il rumore deriva dalle limitazioni dei sensori fotografici nel catturare la luce e si manifesta come punti di colore casuali che spezzano l'uniformità dell'immagine. Le **immagini generate da AI non replicano questo rumore** in modo naturale. Al contrario, spesso presentano un aspetto fin troppo uniforme, privo di quella casualità che caratterizza le foto reali. Nei casi in cui le AI tentano di simulare il rumore, il risultato tende a essere artificioso, con pattern ripetitivi o incoerenti, soprattutto nelle aree più scure o prive di dettagli significativi. L'esempio qui sopra, preso dal database di contenuti artistici con intelligenza artificiale [Lummi](#), dimostra la falsità del rumore nella foto.



Immagine creata con AI (© Lummi) • LUMMI

Gli strumenti di rilevamento: software e metodi avanzati

Anche se la **capacità di osservazione umana** è fondamentale, gli strumenti tecnologici possono offrire un vantaggio decisivo. I software analizzano le caratteristiche delle immagini pixel per pixel, individuando schemi innaturali o segni distintivi delle immagini create con AI. Questi programmi utilizzano reti neurali addestrate per rilevare il linguaggio visivo artificiale, con un'accuratezza che si avvicina al 90%. Uno strumento utile è [FotoForensics](#), che consente di **esaminare i metadati** di un file. Questi dati possono rivelare il software utilizzato per la creazione dell'immagine o dettagli sulla compressione, spesso distintivi dei generatori AI. Tuttavia, bisogna considerare che i metadati possono essere manipolati o rimossi, rendendo necessaria un'analisi più approfondita. Per i più esperti, programmi come [Glaze](#) aggiungono un livello di protezione per le opere artistiche, rendendo **più difficile per le AI replicare** determinati stili o caratteristiche uniche.

E quando si tratta di video?

Anche i **video creati con AI** possono essere smascherati analizzando dettagli specifici. In molte produzioni sintetiche, i movimenti risultano innaturali, e i volti possono cambiare leggermente tra un fotogramma e l'altro. Inoltre, la sincronizzazione tra audio e video potrebbe non essere perfetta, creando discrepanze che rivelano l'origine artificiale. Strumenti come Deepware AI Scanner analizzano questi contenuti frame-by-frame, **rilevando anomalie** nei movimenti o nel pattern di compressione. L'analisi dell'audio può anche evidenziare segnali di sintesi, come tonalità robotiche o cambiamenti improvvisi che contraddistinguono i deepfake.



Il valore del sesto senso umano

Nonostante la crescente sofisticazione delle immagini AI, gli esseri umani mantengono un'incredibile capacità intuitiva di riconoscere contenuti sintetici. Questo *sesto senso* visivo è spesso guidato da una percezione sottile di dettagli imperfetti o elementi che non sembrano del tutto naturali, anche se non si riesce a spiegare razionalmente il motivo. Unendo questa capacità ai consigli pratici, è possibile riconoscere i contenuti sintetici sul web.

Digital forensics reporter breaks down how to spot AI-generated "people" | ABC News Verify

KEVIN NGUYEN
REPORTER, ABC NEWS VERIFY



La **capacità di riconoscere immagini e video creati con AI** non è solo una questione tecnica, ma anche una responsabilità etica. La **disinformazione e l'uso improprio** di queste tecnologie possono avere implicazioni sociali profonde: è fondamentale sviluppare capacità personali e strumenti condivisi che siano facilmente accessibili a tutti, aumentando la consapevolezza generale.