Chi lo ha detto per primo

Settembre 2026: un romanzo escluso dal Goncourt in quattro giorni, sulla base di un punteggio software e di un'accusa di plagio. Otto schede usate come banco di prova dell'equazione del valore su un caso reale, per vedere quale variabile ha deciso davvero gli esiti.

CLAUDIO CAMMARANO 4 OTTOBRE 2026

FONTI	8 schede, lug 2023 – ott 2026
ARGOMENTI	Scrittura · AI · Editoria
CONCETTI	legge di Goodhart · patto di trasparenza · rilevatori di testo generato · segnale costoso · test severo
ORIGINE	https://claudiocammarano.com/writings/2026-10-04-chi-lo-ha-detto-per-primo/

TESI

Nei casi in cui un testo è stato accusato di essere scritto da una macchina, l'esito lo ha deciso la dichiarazione dell'autore, resa verificabile dalla traccia del processo, mentre la quantità di macchina nel testo non ha spostato nulla; è ciò che l'equazione del valore prevede, perché l'autenticità è una proprietà della dichiarazione e nessuna ispezione della prosa può stabilirla.

Nei due saggi di questa serie ho provato a sostituire la domanda «chi o cosa ha scritto questo testo?» con un'equazione del valore. Nella prima versione lo stack di strumenti, che lì chiamavo Σ, non compariva in nessun termine additivo di contenuto, e l'autenticità diventava una grandezza separata e binaria, che vale uno se la dichiarazione sul processo corrisponde a ciò che è stato fatto e zero altrimenti. Nella seconda un simulatore rompeva l'equazione in tre punti, la riparavo, e la tesi sopravviveva. Restava da vedere come si comporta su un caso reale, con un autore vero, un premio vero e un procedimento che nessuno aveva progettato per metterla alla prova. Il caso è arrivato a settembre, in Francia, e Davide Piacenza, su *Culture Wars*, lo ha raccontato come un triste dibattito.[1] Questa catena lo usa come banco di prova: otto schede, quattro passaggi e, alla fine, la domanda su quale variabile abbia deciso davvero gli esiti.

La domanda posta nel modo peggiore

Il romanzo è *C'était ça ou mourir* di Thélyson Orélien, scrittore haitiano e canadese. Nel settembre 2026 un account anonimo, @Pangramed, intitolato al software che sosteneva l'accusa, lo indica come scritto in larga parte da un modello linguistico; il 25 settembre si scopre che dietro l'account c'è un cronista del *Figaro* che dal 2025 lavora in un fondo che investe in startup di intelligenza artificiale. In quattro giorni il libro viene escluso dalla prima selezione del Goncourt, un premio già assegnato gli viene revocato e un tour di presentazioni viene annullato. Il giorno prima dell'esclusione, sul *Nouvel Obs*, Mohamed Mbougar Sarr, che il Goncourt lo ha vinto nel 2021 con un romanzo su uno scrittore africano accusato di plagio, scrive di ignorare se Orélien abbia usato l'intelligenza artificiale e di rifiutare comunque il modo in cui lo si sta stabilendo.[2] Nomina la dimensione razziale, che il presidente dell'Académie, Philippe Claudel, aveva evocato a sua volta il 23 settembre, e chiede soltanto che all'autore si lasci la possibilità di spiegarsi.

---

2 · Mohamed Mbougar Sarr / Le Nouvel Obs · 24 set 2026 · **STRUMENTO CONCETTUALE**

« J'éprouve une compassion profonde pour Thélyson Orélien »

Si può rifiutare un processo senza pronunciarsi sul merito, e dirlo: non so se abbia usato l'IA, e so che questo non è il modo di accertarlo.

https://fr.news.yahoo.com/j-%C3%A9prouve-compassion-profonde-th%C3%A9lyson-173009619.html

---

L'argomento di Sarr ha la forma che questa catena adotta: separa la domanda sulla verità dell'accusa da quella sulla correttezza del procedimento, e osserva che la seconda si può giudicare senza conoscere la risposta alla prima. Il procedimento aveva tre difetti visibili a chiunque. Rovesciava l'onere della prova, perché toccava all'autore dimostrare di non aver usato un modello; affidava l'accertamento a uno strumento mai validato sul materiale in questione; e si svolgeva in ventiquattro ore su una piattaforma. Il comunicato dell'Académie del 25 settembre decide «secondo ogni verosimiglianza», che è la formula con cui un'istituzione dichiara di non avere una prova e di decidere lo stesso, e la decisione non è unanime, perché Pascal Bruckner si oppone.[3]

La tesi di questa catena è che **nei casi in cui un testo è stato accusato di essere scritto da una macchina, l'esito lo ha deciso la dichiarazione dell'autore, resa verificabile dalla traccia del processo, mentre la quantità di macchina nel testo non ha spostato nulla**. È quello che l'equazione prevede, e il caso francese lo mostra per contrasto, perché è l'unico in cui alla dichiarazione mancata non ha supplito una traccia del processo, e si è cercato di ricavare la risposta dal testo.

Lo strumento, preso sul serio

Prima di giudicare l'uso del rilevatore conviene capire che cosa misura, e la letteratura è meno univoca di come la raccontano entrambe le parti. Il numero che circola di più viene da Stanford, luglio 2023: sette rilevatori applicati a saggi TOEFL di non madrelingua li classificavano come generati da una macchina nel 61,22 per cento dei casi, contro il 5,19 per cento dei temi di studenti statunitensi.[4] Il numero è vero, e va datato: riguarda strumenti di prima generazione, in inglese, che stimavano la perplessità del testo, cioè quanto le parole risultino prevedibili per un modello. 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 rilevava lo stile, e attraverso lo stile risaliva a una presunta origine.

Nel febbraio 2026 un gruppo dell'Università Carlo di Praga rifà la verifica in ceco e trova che la perplessità dei non madrelingua non è più bassa, che i rilevatori di tre famiglie diverse non mostrano un bias sistematico e che quelli contemporanei non si basano più sulla perplessità.[5] Nello stesso anno una verifica della Vrije Universiteit di Bruxelles su 1.163 tesi magistrali non trova falsi positivi per Pangram, Copyleaks e Turnitin. La difesa più usata da chi viene segnalato, cioè che lo strumento discrimina chi non scrive nella propria lingua, non regge più nella forma in cui circola.

I limiti però restano, e sono quelli che contano per il caso francese. Una misurazione indipendente dei falsi positivi sul francese letterario non esiste, e le verifiche favorevoli riguardano tesi e saggi accademici. Uno studio dell'Università di Notre Dame, dell'agosto 2026, trova che i punteggi crescono con la densità lessicale e che le discipline umanistiche vengono segnalate molto più di quelle scientifiche: la prosa colta alza il punteggio a prescindere da chi l'ha scritta.[6] Lo stesso studio misura l'altro lato del problema. Un editing leggero e dichiarato con un modello fa scattare la segnalazione fra il 38 e l'80 per cento dei casi, mentre dopo un passaggio in un servizio di «umanizzazione» meno del 4 per cento dei testi generati resta segnalato, il che porta i falsi negativi oltre il 96 per cento. Sopra tutto questo c'è un limite di principio, dimostrato nel 2023 da un gruppo dell'Università del Maryland guidato da Soheil Feizi: man mano che la distribuzione di un modello si avvicina a quella della scrittura umana, anche il miglior rilevatore possibile si avvicina alle prestazioni di una moneta lanciata.

---

6 · Jonathan A. Karr Jr, Grigorii Khvatskii, Ting Hua e Nitesh V. Chawla / University of Notre Dame · 6 ago 2026 · **DATO CHE CORREGGE**

Why AI Detection Fails for Academic Integrity

Chi usa l'AI dichiarandolo rischia la sanzione più di chi la usa e poi fa riscrivere il testo per non farsi prendere: il rilevatore premia la dissimulazione.

https://arxiv.org/abs/2608.11256

Nei termini dell'equazione la diagnosi era già scritta nel primo saggio. L'equazione separa ciò che un testo porta, cioè l'osservazione nuova, la selezione e la ricombinazione, dalla forma in cui lo porta, che lì chiamavo η, l'efficienza di trasmissione: quanta parte di quel contenuto arriva davvero a chi legge. Un rilevatore legge soltanto la forma, e da lì pretende di risalire agli strumenti usati e al contenuto. Liang e Karr trovano lo stesso meccanismo da lati opposti, la povertà lessicale punita nel 2023 e la ricchezza nel 2026, perché in entrambi i casi il punteggio reagisce alla competenza formale di chi scrive. Karr aggiunge la conseguenza che lo rende inservibile come prova: uno strumento che sanziona chi usa poco e dichiara, e assolve chi usa molto e nasconde, finisce per selezionare la capacità di dissimulare, che è la legge di Goodhart nella sua forma più pura. La frase che sopravvive a tutta questa letteratura, compresi gli studi più favorevoli, è la stessa ovunque: i punteggi non vanno usati come prova unica in decisioni ad alta posta. Nella settimana parigina sono stati la base dell'accusa pubblica.

La domanda noiosa

Il comunicato dell'Académie porta però due motivazioni, e la prima che nomina non riguarda l'intelligenza artificiale: sono «i plagi di cui l'autore si è reso colpevole nel tempo e che sono appena stati rivelati». Secondo le verifiche di *Libération* e di Radio-Canada si tratta di passaggi molto vicini a testi altrui in almeno cinque scritti del 2012 e del 2013, dalla prefazione del *Mattino dei maghi* di Pauwels e Bergier a una cronaca che sarebbe la traduzione di un articolo del *Philippine Star*, cioè anni prima che ChatGPT esistesse.[3] L'autore contesta anche queste accuse, ma sono accuse di un altro genere: si verificano confrontando due testi, con il metodo più vecchio della filologia, e la loro solidità non dipende da nessun software. C'è anche un dato che quasi nessuno ha ripreso: secondo NielsenIQ BookData, riportato da *Livres Hebdo*, nella settimana della polemica, fra il 21 e il 27 settembre, il romanzo ha venduto 18.254 copie contro le 12.506 della settimana precedente. L'esclusione ha protetto il premio e ha fatto circolare di più il libro.

---

3 · ActuaLitté · 25 set 2026 · **ELEMENTI DI SCENARIO**

Plagiat, IA : le prix Goncourt exclut le roman de Thélyson Orélien

Un'istituzione letteraria decide su una presunzione — «secondo ogni verosimiglianza» — e sovrappone due accuse di natura diversa, una verificabile e una no.

https://actualitte.com/article/134194/prix-litteraires/plagiat-ia-le-prix-goncourt-exclut-le-roman-de-thelyson-orelien

---

La domanda spettacolare, se il libro sia stato scritto da una macchina, ha prodotto una risposta contestabile; quella noiosa, se certe pagine siano state copiate, una risposta che si può controllare riga per riga. Nei termini dell'equazione la differenza ha una ragione precisa. La prima domanda chiede il valore dello stack, che non compare in nessun termine di contenuto. La seconda chiede da dove vengano l'osservazione e la selezione, cioè proprio i termini additivi che contano, e che un plagio attribuisce a chi non le ha fatte. È anche una questione di autenticità nel senso del primo saggio, perché un plagio è una dichiarazione falsa sul processo, e lo è in un modo che si può accertare dall'esterno.

Che cosa scagioni davvero lo mostra un caso di qualche mese prima. Nel maggio 2026 il racconto vincitore per i Caraibi del Commonwealth Short Story Prize, di Jamir Nazir, di Trinidad e Tobago, viene classificato da Pangram come generato al cento per cento, e il *Guardian* e *Wired* ne titolano.[7] La revisione della Commonwealth Foundation, chiusa il 22 giugno, lo scagiona, e il 30 giugno Nazir vince il premio assoluto. Lo scagionano bozze di lavoro e documenti datati, colloqui sul processo creativo e il giudizio di almeno sette giurati per racconto, senza nessun contro-software e senza una dichiarazione preventiva, che il regolamento non prevedeva. Un'etichetta apposta sul prodotto finito è una promessa e vale quanto chi la firma; un archivio di stati intermedi datati è una traccia, costa poco a chi ha lavorato davvero e molto a chi non l'ha fatto, ed è quindi un segnale costoso nel senso della catena *La firma e il segnale*. Quella catena sosteneva in linea di principio che riconoscere la macchina dallo stile è una strada chiusa; il caso Nazir mostra quale strada resta aperta.

---

7 · Brittle Paper · 26 giu 2026 · **METODO DI LAVORO**

Commonwealth Short Story Prize Clears Regional Winners of AI Use Following Month-Long Review

Scagionato dopo un punteggio del cento per cento: non da una dichiarazione né da un contro-software, ma da bozze datate — la tracciabilità del processo batte il certificato sul prodotto.

https://brittlepaper.com/2026/06/commonwealth-short-story-prize-clears-regional-winners-of-ai-use-following-month-long-review/

La variabile vera

Il controfattuale più pulito viene dal Giappone. Il 17 gennaio 2024 Rie Kudan vince il premio Akutagawa con *Tōkyō-to Dōjō-tō* e, in conferenza stampa lo stesso giorno, dichiara di aver usato un modello generativo e di averne ripreso direttamente circa il 5 per cento del testo.[8] Il premio non viene messo in discussione. Keiichiro Hirano, della giuria, spiega che la parte generata compare dove il romanzo mette in scena un personaggio che usa un modello, cioè come materia narrativa dichiarata. Quel 5 per cento era una cifra detta d'istinto e non una misura, e nell'aprile 2025 Kudan ha pubblicato su commissione un racconto generato al 95 per cento, senza scandalo, perché la regola era dichiarata in partenza. Negli Stati Uniti, nell'agosto 2022, Jason M. Allen aveva vinto il concorso d'arte digitale della Colorado State Fair con un'immagine fatta con Midjourney, iscritta a nome di «Jason M. Allen via Midjourney», e il premio gli è rimasto nonostante le proteste.

Apparato · Quattro casi, una variabile

Caso

Dove e quando

Uso della macchina

Dichiarazione prima dell'accusa

Su che cosa si è deciso

Esito

Jason M. Allen, Théâtre D'opéra Spatial

Stati Uniti, agosto 2022

Immagine generata con Midjourney

Sì, nel nome con cui iscrisse l'opera

Il giudizio della giuria su un'opera dichiarata

Premio mantenuto

Rie Kudan, Tōkyō-to Dōjō-tō

Giappone, gennaio 2024

Circa il 5% del testo, a sua stima

Sì, il giorno del premio

La dichiarazione dell'autrice e la lettura della giuria

Premio Akutagawa confermato

Jamir Nazir, racconto per il Commonwealth

Trinidad e Tobago, maggio–giugno 2026

Nessuno, secondo la revisione; 100% per Pangram

Non prevista dal regolamento

Bozze datate, colloqui sul processo, giurati

Scagionato, poi vincitore assoluto

Thélyson Orélien, C'était ça ou mourir

Francia, settembre 2026

Negato dall'autore; sostenuto da punteggi di software

Nessuna; il diniego arriva dopo

Plagi accertati per confronto su testi del 2012-2013, e una «verosimiglianza»

Escluso dal Goncourt, vendite in aumento

Messi in fila, i casi lasciano una sola variabile che si muove insieme all'esito. La quantità di macchina va dallo zero accertato di Nazir al 95 per cento dichiarato del racconto di Kudan, e non spiega niente; la dichiarazione, quando arriva prima dell'accusa, e la traccia del processo, quando la dichiarazione manca, spiegano tutto. È la forma che l'autenticità aveva nel primo saggio: una funzione che vale uno se la dichiarazione corrisponde al processo effettivo, una proprietà dei metadati che nessuna ispezione della prosa può stabilire. Il caso francese è il tentativo di stabilirla ispezionando la prosa, e il suo esito è quello che l'equazione faceva prevedere. Piacenza arriva allo stesso punto per un'altra strada quando descrive la lettura come un patto non scritto di trasparenza reciproca: l'uso non dichiarato rompe la reciprocità dello scambio, e una relazione, a differenza di un'essenza, si ripara con una dichiarazione.[1]

Controtesi 5 · Adnan Al Ali, Jindřich Helcl e Jindřich Libovický / EACL 2026

Si può obiettare che il problema è transitorio. Se i rilevatori del 2026 non mostrano bias sui non madrelingua, non si basano più sulla perplessità e non producono falsi positivi su più di mille tesi, fra qualche anno la dichiarazione diventerà superflua, perché basterà misurare.[5] Il gruppo di Praga e quello di Bruxelles descrivono uno strumento migliore di come lo raccontano i suoi critici, e la tendenza va in quella direzione.

L'obiezione ha ragione sullo strumento e lascia intatto il punto. Anche un rilevatore perfetto risponderebbe soltanto alla domanda su quali strumenti siano stati usati, e nell'equazione gli strumenti entrano in due posti soli: nella forma, dove possono migliorare la trasmissione fino a un tetto e non oltre, e nel tasso d'errore, dove una verifica accurata ne annulla l'effetto. In nessuno dei due aggiungono o tolgono contenuto. Sull'osservazione e sulla selezione, i termini su cui si è deciso davvero il caso francese, un rilevatore perfetto non direbbe nulla. Dovrebbe poi convivere con il limite di principio del 2023 e con i falsi negativi dell'umanizzazione: un rilevatore che migliora alza il prezzo della dissimulazione, e la dichiarazione resta l'unico dato che non si può ricavare dal testo. Gli stessi autori più favorevoli, del resto, scrivono che i loro strumenti non vanno usati come prova unica.

Il mercato sta arrivando alla stessa conclusione da solo, e ci arriva dal lato sbagliato. Negli Stati Uniti, dal gennaio 2025, l'Authors Guild rilascia ai propri iscritti una certificazione *Human Authored*, un marchio da apporre ai libri il cui testo è stato scritto da persone, sulla base di un'autodichiarazione. È una dichiarazione, e in questo va nella direzione giusta. Dichiara però la variabile che pesa meno, lo stack, e lo fa nella forma più debole, come etichetta sul prodotto: una promessa, per usare la distinzione del caso Nazir, dove servirebbe una traccia. La seconda versione dell'equazione concludeva che un sistema di dichiarazione

onesto dichiara ciò che si può verificare, e che ciò che si può verificare è il termine che pesa meno. Il caso Orélien aggiunge che, quando la dichiarazione manca, a decidere è chi accusa per primo, e che le bozze datate valgono più di qualunque certificato. Per chi scrive, oggi, la protezione più efficace costa pochissimo: dire come ha lavorato prima che glielo chieda qualcuno, e conservarne le prove.

Nella catena

1. Davide Piacenza / Culture Wars, *Il triste dibattito sullo scrivere con l'IA*, 2 ott 2026. Strumento concettuale.
   https://www.culturewars.it/libro-francia-intelligenza-artificiale/
   Scheda: https://claudiocammarano.com/curated/2026-10-02-piacenza-triste-dibattito-scrivere-ia/

2. Mohamed Mbougar Sarr / Le Nouvel Obs, *« J'éprouve une compassion profonde pour Thélyson Orélien »*, 24 set 2026. Strumento concettuale.
   https://fr.news.yahoo.com/j-%C3%A9prouve-compassion-profonde-th%C3%A9lyson-173009619.html
   Scheda: https://claudiocammarano.com/curated/2026-09-24-mbougar-sarr-compassione-profonda-nouvelobs/

3. ActuaLitté, *Plagiat, IA : le prix Goncourt exclut le roman de Thélyson Orélien*, 25 set 2026. Elementi di scenario.
   https://actualitte.com/article/134194/prix-litteraires/plagiat-ia-le-prix-goncourt-exclut-le-roman-de-thelyson-orelien
   Scheda: https://claudiocammarano.com/curated/2026-09-25-goncourt-esclusione-orelien-actualitte/

4. 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/

5. 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/

6. Jonathan A. Karr Jr, Grigorii Khvatskii, Ting Hua e Nitesh V. Chawla / University of Notre Dame, *Why AI Detection Fails for Academic Integrity*, 6 ago 2026. Dato che corregge.
   https://arxiv.org/abs/2608.11256
   Scheda: https://claudiocammarano.com/curated/2026-08-06-karr-perche-la-rilevazione-fallisce-arxiv/

7. Brittle Paper, *Commonwealth Short Story Prize Clears Regional Winners of AI Use Following Month-Long Review*, 26 giu 2026. Metodo di lavoro.
   https://brittlepaper.com/2026/06/commonwealth-short-story-prize-clears-regional-winners-of-ai-use-following-month-long-review/
   Scheda: https://claudiocammarano.com/curated/2026-06-26-commonwealth-nazir-falso-positivo-brittlepaper/

8. CNN, *Japanese author Rie Kudan wins prestigious Akutagawa Prize for novel partly written*

by ChatGPT, 17 gen 2024. Dato che corregge.
https://www.cnn.com/2024/01/19/style/rie-kudan-akutagawa-prize-chatgpt/index.html
Scheda: https://claudiocammarano.com/curated/2024-01-17-kudan-akutagawa-cinque-per-cento-cnn/

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/2026-10-04-chi-lo-ha-detto-per-primo/. Le parti interattive della catena sono disponibili solo online.

Questa catena fa parte della serie *L'equazione del valore*: https://claudiocammarano.com/serie/lequazione-del-valore/.

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.