

La formula dell'autenticità

Perché in tempi di LLM l'autenticità di un testo non è la grandezza che stiamo cercando. Tentativo di trovare l'equazione del valore di un testo, per rispondere alle domande che contano davvero.

Claudio Cammarano 23 agosto 2026

ARGOMENTI AI · Epistemologia · Scrittura

CONCETTI A Mathematical Theory of Communication · Ashby, W. Ross · delega epistemica · editoria · guadagno epistemico · industrie creative · Landauer, Rolf · legge della varietà richiesta · LLM come attante zero · neghentropia · principio di Landauer · scrittura · Shannon, Claude E.

ORIGINE <https://claudiocammarano.com/writings/2026-08-23-la-formula-dellautenticita/>

Atene, la statua di Platone davanti all'Accademia Nazionale. Oltre a essere l'unico filosofo classico di cui deteniamo l'opera integrale, Platone si è distinto non solo nella ricerca della verità, ma anche nella ricerca delle adeguate forme espressive.

0. La tesi

Il ricorso agli LLM sta diventando sempre più pervasivo, e questo naturalmente genera anche una forma di rigetto culturale, di cui possiamo tutti trovare testimonianza sui social media e anche nelle università (specie quelle anglosassoni). Senza entrare eccessivamente in casistiche di questo o quel tipo, con la loro aneddotica collegata, si può affermare senza troppo sbagliare che questo atteggiamento nasca come una forma di compensazione rispetto al potere generativo della tecnologia, il cui fine culturale è per l'appunto quello di ridurre la velocità del cambiamento sociale, che come tutti sappiamo ha sempre un costo e la cui entità è spesso difficile da valutare in anticipo.

Nel vasto mondo delle industrie creative, il dibattito ha preso la forma di una *querelle* sull'autenticità dei testi scritti con l'AI — a proposito della quale ritengo di avere una posizione più aperta della moda, per quanto abbia in questo un interesse molto ben definito, visto che, come ho dichiarato più volte, senza gli LLM questo sito non esisterebbe. D'altra parte, i miei interessi sono complessivamente molto compositi, se si considera che lavoro in una casa editrice: che è per definizione il luogo in cui si dà — giustamente — enorme valore all'autorialità. Ho quindi cercato di formulare una formalizzazione che cercasse di argomentare le mie posizioni con il maggiore rigore di cui sono capace, tenendo insieme il rigore argomentativo, un certo spirito divulgativo e la sempre indispensabile onestà intellettuale. E, tanto per essere chiari, sarò come

già in altri casi supportato da un LLM per quanto riguarda il ricorso alla notazione matematica legata alla formalizzazione.

La questione su cui vorrei misurarmi è il concetto di **autenticità**. Sul piano strettamente personale, è un valore che non mi entusiasma: e lo dico pur essendo una sorta di fanatico del realismo metafisico, ossia una persona che mediamente mette mano alla pistola quando sente parlare un sostenitore del postmoderno. Tuttavia, mi pare che il concetto, almeno nella sua accezione nel linguaggio comune, celi più di quanto non riveli, e che quindi vada esaminato con attenzione, come il veleno di un serpente o la coda di uno scorpione.

Vorrei quindi definire in qualche modo il significato del termine, capire quanto la stesura di un testo con un LLM possa intaccare la genuinità di un testo e, soprattutto, se l'autenticità sia un parametro adatto a definire la qualità e la quantità di innovazione che un testo può introdurre entro un dato sistema culturale. In particolare, **la mia tesi è che la domanda “chi o cosa ha scritto questo testo” sia formalmente mal posta, e che le domande ben poste su un testo siano al contrario quattro:**

- **quanta osservazione nuova entra,**
- **chi ha operato la selezione,**
- **quale tasso d'errore non corretto viaggia nel testo,**
- **e che cosa accade alla varietà del sistema quando molti testi vengono prodotti così.**

Trattarle tutte al momento è abbastanza al di là dei miei sforzi. Di queste quattro, lo strumento — cioè l'AI — tocca la terza e la quarta, e non tocca affatto le prime due. D'altra parte, il discorso pubblico mi pare non ne discuta nessuna. La matematica che scomoda per la mia argomentazione ha un valore ovviamente ostensivo: non c'è pretesa di originalità in questo contributo, se non appunto un contributo alla chiarezza per chi abbia la pazienza di prendere in considerazione le formule.

Quello che segue è per l'appunto il tentativo di scrivere per esteso l'equazione che descrive il fenomeno.

1. Primitivi e notazione

Fisso un dominio di configurazioni possibili Ω — proposizioni su un certo argomento, stati del mondo, ipotesi in competizione. Non serve che sia enumerabile nella pratica: serve che abbia senso parlare di alternative in competizione. Se la questione è l'andamento del mercato del libro nei prossimi cinque anni, Ω contiene tutti gli scenari che qualcuno potrebbe ragionevolmente sostenere.

Un **ricevente** R è caratterizzato da una distribuzione p_R su Ω : le sue attese, il suo prior, ciò che considera più o meno probabile. La sua entropia è massima quando la distribuzione è uniforme, che è precisamente il caso in cui non sappia che pesci pigliare (la condizione che per Kahneman, ricordo, ha il maggiore costo emotivo, e che attiva quasi immediatamente le euristiche del Sistema 1).

$$H(p_R) = - \sum_{\omega \in \Omega} p_R(\omega) \log p_R(\omega)$$

Introduco subito qui un punto cruciale. Adotto qui il senso termodinamico e cibernetico della parola **entropia**, quello che va da Schrödinger a Brillouin a Wiener: entropia come

stato di equiprobabilità, indifferenza fra le alternative, piattume. L'informazione coincide con la neghentropia, cioè con la distanza da quello stato.

La forma dell'equazione, non sfuggirà a chi ha studiato un po' di matematica alle scuole superiori, è quella dell'entropia di Shannon: per ogni configurazione ω del dominio si moltiplica la sua probabilità per il logaritmo della stessa probabilità, si sommano i contributi e si cambia segno. Il segno meno serve solo perché i logaritmi di numeri fra 0 e 1 sono negativi, quindi rende positiva una quantità che altrimenti lo sarebbe al contrario.

Il punto che a me preme parecchio è che la **stessa identica espressione** è l'entropia di Gibbs in termodinamica statistica:

$$S = -k_B \sum_i p_i \ln p_i$$

Le uniche differenze sono la costante di Boltzmann davanti, che fissa le unità fisiche (joule per kelvin invece di bit), e la base del logaritmo, che è naturale invece che 2. È esattamente questa coincidenza di forma — notata da von Neumann e poi sfruttata da Brillouin — che ci autorizza a usare “entropia” nel senso termodinamico senza uscire dal formalismo di Shannon: massima quando tutte le configurazioni sono equiprobabili, cioè il piattume; nulla quando una sola configurazione ha probabilità 1.

Aggiungo una precisazione che serve a disinnescare l'obiezione di chi ha studiato Shannon e sente puzza di bruciato quando legge «entropia uguale assenza di informazione». Nel formalismo shannoniano convivono due oggetti che portano lo stesso nome: l'entropia di una *sorgente*, che misura quanta informazione essa può emettere, e l'entropia di uno *stato*, che misura quanto quello stato sia indifferenziato. Quando qui dico entropia parlo sempre del secondo. E la grandezza che sto chiamando informazione ha, in Shannon, un nome esatto: è la ridondanza, $R = 1 - H/H_{\max}$, cioè la distanza normalizzata dalla distribuzione piatta. I due lessici dicono la stessa cosa da due lati diversi, e non c'è nulla da sanare.

Ancora, un **testo** T è un operatore che porta p_R in $p_{R|T} \equiv p'_R$. Nient'altro: nella teoria un testo lungo, elegante, documentato e privo di effetti sulle attese di chi lo legge è l'operatore identità. Devo ammettere che nella mia testa sto pensando a un testo saggistico; ma nulla vieta di pensare a un prodotto letterario: del resto l'innovazione dei formati narrativi o dei linguaggi serve precisamente la funzione informatica che stiamo cercando di definire.

Un **campo** o pubblico è una misura μ sullo spazio dei riceventi $\mathcal{R}$. L'autore stesso è un punto di quello spazio, R_A , e nulla nella teoria gli assegna uno statuto privilegiato. È una scelta che va argomentata, non assunta, ma desacralizzare il ruolo della produzione culturale denota sicuramente una postura intellettuale più feconda, per chi voglia elaborare una teoria.

Un **corpus** $\mathcal{C}$ è la distribuzione generatrice dei testi già esistenti, con densità $p_{\mathcal{C}}$: l'insieme di ciò che è già stato scritto su una questione, considerato come distribuzione di probabilità sulle cose dicibili.

Uno **stack di strumenti** Σ è l'insieme dei mezzi impiegati nel processo: alfabeto, schedario, dizionario, motore di ricerca, correttore ortografico, LLM. Tratto Σ come una variabile ordinaria, da collocare nell'equazione dove i conti la mettono, senza deciderne in anticipo la posizione — che è poi tutto il punto di questo esercizio.

2. Due grandezze distinte, che il dibattito confonde in una

Un testo produce due effetti misurabili e indipendenti.

Spostamento — quanto le attese si muovono:

$$D(T, R) = D_{KL}(p'_R \parallel p_R) = \sum_{\omega} p'_R(\omega) \log \frac{p'_R(\omega)}{p_R(\omega)}$$

Lo spostamento registra quanto le attese del lettore si sono mosse. Si prende ogni configurazione possibile, si guarda il rapporto fra la probabilità che il lettore le assegna dopo aver letto e quella che le assegnava prima, se ne prende il logaritmo — così che raddoppiare una probabilità pesi quanto dimezzarla, con segno opposto — e si fa la media di questi scarti pesandoli con le probabilità *finali*. Il peso finale è la scelta che rende la grandezza asimmetrica, e l'asimmetria è voluta: contano gli stati che il testo ha reso credibili, mentre quelli che ha reso trascurabili escono di scena. Il risultato vale zero quando il lettore, dopo, crede esattamente ciò che credeva prima, e cresce senza limite superiore man mano che il testo lo porta a considerare probabile ciò che considerava impossibile.

Struttura — quanta indifferenza si riduce:

$$\Delta H(T, R) = H(p_R) - H(p'_R)$$

La struttura registra tutt'altro: quanta indifferenza fra le alternative si è consumata. Si calcola l'entropia prima, l'entropia dopo, e si sottrae. È positiva quando il lettore esce dalla lettura con meno alternative in gioco di quante ne avesse entrando, negativa quando ne esce con più.

Le due grandezze non coincidono, e la loro non coincidenza è il primo risultato utile. Vale l'identità

$$\Delta H = D_{KL}(p' \parallel p) - [C(p', p) - H(p)], \quad C(p', p) = - \sum_{\omega} p'(\omega) \log p(\omega)$$

dove il termine fra parentesi quadre misura quanto il modello precedente fosse mal tarato sugli stati che il testo rende probabili — chiamo **rigidità** quel termine. **Un testo produce struttura nella misura in cui lo spostamento che impone eccede la rigidità del ricevente.**

Perché non coincidono. Riprendiamo il caso del mercato del libro, su cui la nostra lettrice o il nostro lettore stanno cercando di farsi un'idea. Un lettore può essere spostato moltissimo e restare più disorganizzato di prima. Riduciamo l'andamento del mercato del libro nei prossimi cinque anni a quattro scenari in competizione. Il lettore è un operatore del settore, convinto che non stia succedendo granché; il testo lo convince che sotto la superficie piatta stia avvenendo un riassetto.

scenario	prima	dopo
il mercato cresce	0,05	0,10
il mercato resta stabile	0,85	0,10
il mercato cala	0,05	0,10

scenario	prima	dopo
il mercato si riassetta: cala il trade, cresce l'ascolto	0,05	0,70

Lo spostamento vale circa 2,56 bit, che è moltissimo: con quattro alternative l'entropia massima possibile — il piattume totale, quattro scenari equiprobabili — vale 2 bit, e la divergenza li supera perché non ha tetto, cresce senza limite quando il testo rende probabile ciò che il lettore aveva quasi escluso. Quasi tutto quel valore viene da una sola riga, la quarta, la cui probabilità è stata moltiplicata per quattordici.

La struttura, sulla stessa coppia di distribuzioni, vale meno 0,51 bit. Il lettore ha cambiato idea e insieme ha perso certezza, perché ha barattato una convinzione ferma (0,85) con una convinzione tiepida (0,70): esce dalla lettura con più alternative in gioco di quante ne avesse entrando.

Il secondo lettore è chi del settore non sa nulla e assegna 0,25 a ciascuno scenario. Lo stesso identico testo lo porta alla stessa distribuzione finale, e su di lui le due grandezze coincidono esattamente, 0,64 bit entrambe. Stesso testo, due lettori, effetti diversi per grandezza e per segno.

L'identità. Il fatto che nel secondo caso coincidano e nel primo no ha una causa unica, ed è ciò che l'identità isola. Lo spostamento e la struttura differiscono per un termine che confronta l'entropia incrociata con l'entropia del prior: l'entropia incrociata misura quanto male le vecchie attese prevedevano proprio gli stati che il testo rende probabili, l'entropia del prior misura quanto male prevedevano se stesse. La differenza fra le due — la **rigidità** — dice se il testo sta puntando verso una zona che il lettore aveva già considerato o verso una che aveva scartato.

Detta in parole povere: **la struttura che un testo produce è lo spostamento che impone, meno la rigidità che deve vincere.** Con un lettore piatto la rigidità è nulla per costruzione e ogni bit di spostamento diventa un bit di struttura. Con un lettore già orientato altrove la rigidità è alta e una parte dello spostamento si consuma nello smontare la convinzione precedente. E il termine può anche essere negativo — nel caso in cui il testo concentri il lettore su ciò che già riteneva più probabile — e allora la struttura eccede lo spostamento: è la forma esatta di ciò che chiamiamo conferma.

Da qui si leggono i due modi di fallire. Un testo con D alto e $\Delta H \leq 0$ muove senza organizzare: disorienta, moltiplica le possibilità aperte, e a lungo andare produce piattume nonostante l'apparenza di novità — è il regime della provocazione e della polemica. Un testo con ΔH alto e $D \approx 0$ organizza senza muovere: conferma, ribadisce, consola (forse radicalizza). **Il valore richiede entrambe le grandezze positive, e il dibattito sull'AI non ne discute per nulla.**

3. Il guadagno verso il vero

Struttura e correttezza sono indipendenti l'una dall'altra. La nostra teoria deve dirlo, altrimenti finisce per premiare il testo peggiore che si possa immaginare.

Il caso da temere è infatti quello di un saggio scritto benissimo che organizza le attese del lettore attorno a una tesi falsa. Nei termini della sezione precedente, un testo falso produce struttura alta: riduce l'indifferenza, produce ordine, lascia chi legge con la sensazione lucida di aver finalmente capito. È neghentropico in senso stretto e distruttivo in senso epistemico, e — dettaglio che dovrebbe inquietare chiunque scriva per mestiere — è tanto più efficace quanto migliore è la prosa.

Per escluderlo introduco la distribuzione di riferimento q , cioè il mondo, o un suo surrogato operativo di cui dirò al §13, e definisco il **guadagno epistemico**:

$$G(T, R) = D_{KL}(p_R \parallel q) - D_{KL}(p'_R \parallel q)$$

Si legge, letteralmente, come un accorciamento di distanza. Prima della lettura il ricevente sta a una certa distanza dalla distribuzione vera; dopo la lettura sta a un'altra distanza; il guadagno è di quanto quella distanza si è ridotta. Positivo quando il testo avvicina il ricevente al mondo, negativo quando ne allontana, e nullo in un terzo caso che nel dibattito non compare mai e che potremmo definire la sindrome di Siddharta, il personaggio di Herman Hesse che per cercare la verità le prova letteralmente tutte: l'eroe della storia, e in questo caso il lettore, si sposta lateralmente, sostituisce un errore con un altro errore ugualmente distante, pur avendo la sensazione di aver imparato qualcosa.

Il caso interessante è dunque quello con $\Delta H > 0$ e $G < 0$: struttura positiva, guadagno negativo. Ogni teoria del valore fondata sulla sola riduzione di entropia premia questo caso, e per questo la sola riduzione di entropia non basta. Assumo G , e non ΔH , come funzionale primario del valore.

La scelta ha una conseguenza immediata sul punto in cui l'automazione entra nell'equazione, e conviene anticiparla perché per le posizioni attuali del dibattito è controintuitiva: la produzione automatica non riduce affatto la capacità di produrre struttura. Potrebbe persino aumentarla, quando migliora la forma di partenza (pensate al caso in cui vi aiuta a scrivere un brief o un paper, casi in cui un'adeguata strutturazione del testo è condizione necessaria alla veicolazione di un'idea). E agisce invece sul tasso d'errore con cui quella struttura viene prodotta. Ci arrivo al §8.

4. Il sistema di riferimento

Il guadagno è definito su un ricevente, mai in astratto. Questa non è una limitazione tecnica della teoria, ma la sostanza. Al di là della forma linguistica che possa usare, per esigenze di sintesi, in questo testo, **non esiste la quantità di informazione «di un testo»**. Esiste soltanto la differenza che quel testo fa a qualcuno che aveva certe attese.

Il valore di un testo rispetto a un pubblico è quindi un integrale sullo spazio dei riceventi:

$$G(T, \mu) = \int_{\mathcal{R}} G(T, R) d\mu(R)$$

che è il modo compatto di dire una cosa semplice: si sommano i guadagni ottenuti dai singoli lettori, pesando ciascuno per quanti lettori di quel tipo il testo effettivamente raggiunge. Lo stesso identico testo ha valori diversi rispetto a pubblici diversi, e può valere molto per uno e zero per un altro senza che una virgola sia cambiata. (Potrebbe sortire effetti molto diversi anche sulla base dell'ampiezza della sua audience: ma è un elemento che per ora non esploreremo.)

Ne segue immediatamente la risposta al caso limite. Prendiamo un testo con $G(T, R_A) = 0$ e $G(T, \mu) > 0$: nulla per l'autore, molto per il pubblico. Ha valore positivo, e la teoria non dispone di alcun termine con cui chiamarlo inautentico. Il costo di questa posizione va pagato per intero, e lo pago: **sotto questo criterio l'origine di un testo non è una variabile del suo valore**.

Restano però tre termini che l'argomento sull'origine manca e che l'equazione registra. Sono i §§ 7, 8 e 9. Ma prima conviene guardare da vicino da dove viene, dentro il processo di scrittura, la varietà che il testo immette — perché è lì che si decide tutto il resto.

5. Da dove viene la varietà

Scompongo il guadagno di contenuto secondo la fonte della varietà che il testo immette:

$$G_{\text{contenuto}} = G_{\text{oss}} + G_{\text{sel}} + G_{\text{ric}}$$

Tre addendi, tre operazioni distinte.

G_{oss} è l'**osservazione**: dati, documenti, misure, esperienza diretta, interviste, la lettura di una fonte che nessuno aveva letto. È varietà che entra nel sistema dall'esterno del corpus, e ha una proprietà che la distingue dalle altre due: è l'unica che aumenta la varietà complessivamente disponibile invece di ridistribuirla.

G_{sel} è la **selezione**: la scelta della domanda, del taglio, della scala di analisi, di ciò che viene scartato. Nella forma della legge di Ashby, la varietà del regolatore deve eguagliare quella del sistema regolato: selezionare è l'operazione che rende un corpus adeguato a un problema. È anche l'operazione meno visibile nel prodotto finito, per una ragione tautologica — di tutto ciò che è stato scartato il lettore non vede nulla — e questa invisibilità avrà conseguenze pesanti al §6.

G_{ric} è la **ricombinazione**: trasporto di una struttura da un dominio a un altro, analogia, applicazione fuori sede. Non aggiunge varietà al sistema, ne aumenta l'accessibilità: porta in un campo la struttura che un altro campo aveva già. Questo intero saggio, che applica termodinamica e teoria dell'informazione a una questione di scrittura, è un esercizio di G_{ric} e nient'altro.

A queste si aggiunge, separata, l'**espressione**: la realizzazione superficiale, sintassi, lessico, ritmo, ordine dell'esposizione. Va tenuta separata perché non è un termine additivo. Agisce come efficienza di trasmissione $\eta \in [0, 1]$, un coefficiente che governa quanta parte del guadagno di contenuto arriva effettivamente al ricevente:

$$G_{\text{consegnato}} = \eta \cdot G_{\text{contenuto}}$$

Detto a parole: una forma pessima disperde — il lettore abbandona a metà, fraintende, si annoia, e il guadagno potenziale resta sulla pagina senza mai raggiungerlo. **Una forma perfetta consegna tutto il contenuto e non un bit di più.** Il limite superiore di η è uno, e questo è l'intero spazio in cui la qualità della scrittura, umana o assistita, può operare. Se il contenuto è nullo, non c'è prosa che lo moltiplichi in qualcosa.

Per fare un esempio che sembra una battuta, quanto guadagno di contenuto avremmo avuto in più, se Hegel avesse saputo scrivere! Viceversa, e mi scuso con i suoi fan, fra i quali mi annovero anch'io: se Nietzsche avesse scritto molto male, forse oggi ne parleremmo molto meno (anche perché alcune delle sue intuizioni le intravediamo già in Spinoza, che peraltro nella sua *Etica* ha uno stile molto diverso fra dimostrazioni e scoli).

6. Lo spazio delle posizioni

A questo punto la teoria dispone di tre operazioni di contenuto e di un coefficiente, e diventa possibile chiedersi che aspetto abbia un sistema di dichiarazione costruito su queste coordinate. Ne uso uno concreto, che è quello che adotto per i testi che pubblico su questo sito. Il motivo per cui lo faccio è che voglio dare una fondazione teorica – opinabile, per carità – alle scelte che ho fatto nella realizzazione di questo sito. Il quale, come ho scritto in più passaggi, senza gli LLM non potrebbe esistere, per quanto mi sforzi di non essere dipendente dagli LLM nel suo mantenimento.

La notazione che propongo ha due assi. Il primo, orizzontale e sempre presente, dichiara il **grado di generazione della prosa** con quattro valori ordinali ed esclusivi: 00 nessun intervento, DL dialogo – discussione, brainstorming, ricerca e verifica, con il testo scritto interamente da me –, ED editing di un testo interamente mio, WR prima stesura generata su mie indicazioni e poi riscritta. (Non esiste, nel mio caso, uno scenario in cui faccio scrivere testi all'LLM e li pubblico direttamente – per quanto disponga di un set di skill piuttosto sofisticate.) Il criterio, dichiarato in colophon, è che la scala misuri *quanta della superficie pubblicata sia passata dal modello*, e non il peso intellettuale del contributo.

Il secondo asse, verticale e facoltativo, dichiara gli **apparati** trattati, con due modificatori cumulabili e non ordinali: FM per la costruzione assistita di apparati tecnici – formule, modelli, codice, tabelle – che entrano nel testo, e TR per la traduzione o l'adattamento. Le combinazioni possibili sono quattro: nessuno, FM, TR, FM·TR.

Quattro valori sul primo asse per quattro combinazioni sul secondo danno una griglia discreta di sedici posizioni potenziali.

apparati	prosa 00	prosa DL	prosa ED	prosa WR
nessuno	00	DL	ED	WR
FM	00·FM	DL·FM	ED·FM	WR·FM
TR	00·TR	DL·TR	ED·TR	WR·TR
FM·TR	00·FM·TR	DL·FM·TR	ED·FM·TR	WR·FM·TR

Fin qui è una legenda. Letta con l'equazione in mano diventa qualcosa di diverso, perché le sedici caselle smettono di essere un elenco di casi e diventano lo spazio degli stati della teoria: ogni posizione è un vettore che dice quali termini sono rimasti all'autore e quali sono passati allo strumento. E appena la si legge così, dice quattro cose che una legenda non direbbe.

L'asse orizzontale dichiara il coefficiente. I quattro valori orizzontali misurano, per costruzione e per dichiarazione esplicita, quanta superficie è passata dal modello: misurano cioè un'operazione che agisce su η . L'asse più visibile della notazione è quello che nell'equazione non compare in nessun termine additivo. Non è un difetto della notazione – è la sua onestà: dichiara ciò che si può verificare guardando il testo, e si astiene dal dichiarare ciò che peserebbe di più e che nessuno potrebbe controllare.

Il modificatore che sembra tecnico è l'unico che tocca il contenuto. FM dichiara la costruzione assistita di un apparato formale, e un apparato formale non è superficie: è struttura, entra in G_{ric} , e in certi casi anche in G_{oss} , quando la formalizzazione produce derivazioni che nessuno aveva fatto prima. È il caso di questo articolo, per esempio. Nella griglia FM occupa la posizione modesta del modificatore facoltativo, mentre nell'equazione è il solo codice della notazione capace di spostare un termine additivo.

Chi legge la matrice come una scala di intervento crescente da sinistra a destra sta guardando l'asse sbagliato. E chi scrive questa nota, avendo dichiarato in apertura di essersi fatto assistere proprio sulla notazione matematica, sta dichiarando il codice che pesa di più.

Il modificatore TR è la dimostrazione empirica del §4. Una traduzione ha guadagno epistemico rigorosamente nullo per chi ha già letto l'originale: nessuno spostamento, nessuna struttura nuova, l'operatore identità. E ha guadagno pieno per un pubblico che a quell'originale non aveva accesso. È esattamente il caso limite di cui parlavo — un testo non neghentropico rispetto a un sistema e neghentropico rispetto a un altro, senza che nel testo sia cambiato nulla — con la differenza che qui il caso limite si scopre essere l'operazione più ordinaria del mestiere editoriale, praticata da qualche millennio senza che a nessuno sia mai venuto in mente di chiamarla inautentica. Chi rifiuta il criterio relazionale del valore ha l'onere di spiegare in che senso tradurre sarebbe una frode.

Nella realtà, come spesso dichiaro, il lavoro che faccio sulle traduzioni non è mai una banale trasposizione in lingua, ma anche una forma di adattamento culturale: i post sul mio Substack *The Abstract* sono pensati in una prospettiva europea; pertanto una semplice traduzione linguistica non avrebbe senso. Inoltre, sul sito non mi pongo vincoli di brevità, mentre su Substack lo faccio eccome, per non far scappare la gente. Questo mi fa dire che in realtà uso il digramma TR in maniera potremmo dire garantista verso il lettore e la lettrice: un contributo di contenuto mio c'è sempre, ma preferisco approssimare il ruolo dell'LLM per eccesso.

Le posizioni vuote sono informative quanto quelle occupate. Delle sedici, alcune sono strutturalmente povere di occupanti, e la ragione rivela un'asimmetria fra i due modificatori che in fase di progettazione non avevo visto. FM è genuinamente ortogonale all'asse orizzontale: si può costruire un apparato in dialogo con il modello e scrivere la prosa senza toccarlo, ed è il caso $00 \cdot FM$. TR lo è molto meno, perché tradurre significa produrre superficie: se la traduzione è assistita, la prosa d'arrivo è passata dal modello, e il codice orizzontale corretto diventa ED o WR . La casella $00 \cdot TR$ resta abitabile solo se TR viene inteso come marcatore di *provenienza* — questo pezzo è l'adattamento di un mio testo precedente — invece che come marcatore di *delega*. Sono due variabili distinte tenute insieme sotto un codice solo, e questo rende la notazione meno informativa di quanto potrebbe essere.

L'asse mancante è quello che l'equazione considera decisivo. La griglia non ha alcun codice per la selezione. Dichiaro chi ha prodotto la prosa e chi ha costruito l'apparato, e tace su chi ha scelto la domanda, il taglio, ciò che resta fuori. Dal punto di vista della teoria è l'omissione più grave possibile, perché G_{sel} è il termine additivo maggiore. Dal punto di vista della pratica l'omissione ha una difesa, che viene peraltro dalla teoria stessa: un codice il cui valore non varia mai non trasporta informazione — è l'entropia della sorgente del §1, applicata questa volta ai metadati invece che al testo — e se la selezione resta invariabilmente all'autore, quel codice avrebbe entropia nulla e occuperebbe spazio senza dire niente. La difesa regge finché la notazione descrive un autore solo, con il suo metodo. Nel momento in cui venisse adottata da altri, con approcci diversi, la casella mancante diventerebbe la sola che conta davvero.

Resta un'ultima osservazione, che fa da ponte verso il paragrafo seguente. Nessuna delle sedici posizioni è una posizione di valore. La griglia è ordinata su un asse, da 00 a WR , e ogni lettore sarà tentato di leggere quell'ordine come una scala di purezza decrescente. L'equazione dice il contrario, e lo dice in modo verificabile: poiché Σ entra solo come coefficiente saturante e come termine d'errore azzerabile per verifica, **una**

posizione nella matrice non è una posizione nel valore. Un WR·FM con osservazione nuova e selezione forte vale più di un 00 che non ha niente da dire, e la matrice non ha modo di dirlo perché non è quello il suo mestiere.

7. Un teorema sul modello, e il suo scolio

Un modello linguistico è addestrato a minimizzare l'entropia incrociata con il corpus. Vale la pena leggere questa frase con il §2 in mano, perché l'entropia incrociata è la stessa grandezza che lì misurava quanto male un prior prevedeva gli stati che un altro rende probabili: addestrare un modello significa esattamente ridurre quanto male esso prevede il testo già scritto, cioè rendere la distribuzione p_M il più vicino possibile a p_C . Non è un'interpretazione critica dell'addestramento, è la sua definizione operativa.

Da lì segue una conseguenza che merita di essere scritta. Per un ricevente il cui prior coincide con il corpus, il guadagno atteso di un testo campionato da p_M è

$$\mathbb{E}_{T \sim p_M} [D_{KL}(p'_R \parallel p_R)] \rightarrow 0 \quad \text{per} \quad D_{KL}(p_C \parallel p_M) \rightarrow 0$$

che a parole si legge così: se il modello riproduce fedelmente la distribuzione di ciò che è già stato scritto, allora ciò che il modello scrive è già stato scritto, e a chi ha letto tutto non dice niente. **Quanto migliore è il modello, tanto minore è l'informazione che il suo campionamento non condizionato può trasportare verso un ricevente che conosca il corpus.**

Il guadagno residuo deriva interamente dalla distanza $D_{KL}(p_R \parallel p_C)$, cioè dall'ignoranza del singolo lettore rispetto a ciò che è già scritto. È un servizio di compressione e recupero, ha valore reale e immediato — chiunque abbia chiesto a un modello di spiegargli un campo che non conosce lo sa —, ma ha la struttura economica di una rendita su un'asimmetria, e le asimmetrie di questo tipo si chiudono man mano che i riceventi si aggiornano.

Lo scolio conta quanto il teorema, e va detto con la stessa nettezza, perché senza di esso il teorema diventa un argomento contro l'uso dei modelli, che non è quello che sostengo. Il risultato vale per il campionamento non condizionato. Condizionando su un prompt che porti G_{oss} e G_{sel} dall'esterno, l'uscita può cadere lontanissimo dalla regione modale — ma la varietà che ce la porta è varietà dell'autore, immessa attraverso il condizionamento. **Il modello moltiplica la varietà che gli viene data e non ne crea di propria:** nel primo caso Σ opera su η , nel secondo su nulla.

Da cui il risultato operativo di tutta la costruzione, e il punto in cui la griglia del paragrafo precedente mostra il proprio limite: **lo strumento diventa distruttivo esattamente nel punto in cui gli si delega la selezione, e resta neutro nel punto in cui gli si delega la produzione.** Delegare η non costa niente. Delegare G_{sel} azzerava il termine più grande dell'equazione. Che è l'inverso di ciò che il dibattito sorveglia, e l'inverso di ciò che qualunque sistema di dichiarazione — incluso il mio — riesce oggi a rendere visibile.

8. Errore e correzione

Ogni testo trasporta un tasso di affermazioni errate ϵ . L'errore non riduce semplicemente il guadagno: lo inverte, perché la struttura falsa organizza il ricevente attorno a configurazioni lontane da q , e — come si è visto al §3 — organizza tanto meglio quanto migliore è la forma. Il costo va quindi pesato con un fattore Λ asimmetrico e maggiore

dell'unità, che tiene conto della propagazione dell'errore nei testi successivi: un dato sbagliato citato tre volte diventa un fatto, e ripulire il campo costa molto più di quanto sia costato inquinarlo.

Qui la funzione dell'autore diventa formalmente descrivibile, e smette di essere una rivendicazione di categoria: **l'autore opera nel canale come codice di correzione d'errore**. Un testo che attraversa il suo sistema cognitivo arriva con un ε stimato — non necessariamente basso, ma stimato, e questa è la differenza che conta. Un testo che non lo attraversa arriva con un ε ignoto. La firma dichiara al ricevente proprietà di rumore del canale, e apporla su un testo non verificato dichiara proprietà che il canale non possiede.

Si noti che questo non è un argomento sull'origine. Un testo generato e poi verificato integralmente ha ε stimato quanto un testo scritto a mano. Un testo scritto a mano da un autore distratto ha ε alto. **La variabile è la verifica, e la verifica è indipendente dallo strumento di produzione** — comparirà infatti nell'equazione finale come intensità ν , accanto a Σ , proprio per mostrare che l'una compensa l'altro.

Uno degli aspetti se si vuole consolanti della teoria è che non tocca la sacralità dell'autore che citavo all'inizio: questa specifica autorialità è più assimilabile all'*intentio auctoris* e al ruolo svolto dal *campo* occupato dalla filiera editoriale nel suo complesso (autore/autrice, agente, editor, correttore di bozze...) che all'ineffabile intuizione creativa. D'altra parte, in questo esercizio cercare di formalizzarla non è utile, e lascia all'umano uno spazio suo specifico che resta centrale.

Un corollario, infine, riguarda i sistemi di rilevazione automatica dei testi generati, watermark statistici compresi. Quei sistemi leggono la superficie, cioè esattamente e soltanto l'asse orizzontale della griglia del §6. Sono ciechi al dialogo e alla formalizzazione, che nella prosa non lasciano traccia, e leggono una traduzione assistita come testo integralmente generato anche quando l'originale è dell'autore. Un rilevatore, in altri termini, misura η e da lì inferisce sui termini di contenuto: esegue in automatico, e su scala industriale, esattamente l'errore che questo saggio cerca di diagnosticare.

9. Il livello di sistema

Fin qui il singolo testo. Il termine che manca si vede solo sulla popolazione, ed è a mio giudizio il solo effetto dell'AI generativa che meriti davvero allarme.

Sia $\mathcal{P}_t$ la distribuzione dei prior degli autori attivi in un campo al tempo t . La capacità del sistema di produrre guadagno futuro dipende dalla varianza di $\mathcal{P}_t$: se tutti gli autori condividono lo stesso prior, nessun testo può spostare nessuno, e il campo diventa internamente muto pur continuando a pubblicare. È la legge di Ashby vista dal lato della produzione invece che della regolazione — solo varietà distrugge varietà.

Ogni pubblicazione aggiorna il prior pubblico, quindi **il guadagno è autoconsumante**: un testo riduce la distanza che lo rendeva informativo, e ripubblicarlo non produce due volte lo stesso effetto. Nel caso normale la riduzione è compensata dall'immissione di varietà nuova via G_{oss} . Nel caso del campionamento modale su larga scala la compensazione non avviene, e $\text{Var}[\mathcal{P}_t]$ decresce: il sistema converge verso la media del corpus. È produzione di entropia in senso proprio, a livello ecologico, ed è l'unico contesto in cui la parola piattume descrive un fatto misurabile invece che un giudizio di gusto.

Aggiungo il termine come costo:

$$- \kappa \cdot \Delta Var [\mathcal{P}]$$

Il rischio reale non è il singolo testo inautentico, è la contrazione della varianza fra gli autori. Che è misurabile, mentre l'autenticità del singolo testo non lo è. Questo non esclude, naturalmente, effetti sociali positivi: per esempio la costruzione del nuovo pubblico del Romantasy nasce anche dalle caratteristiche facilmente decostruibili e riproducibili del genere, associate peraltro (non me ne vogliano le TikTokker) a un certo livello di semplificazione del linguaggio. Questo, associato al marketing delle piattaforme e al contagio delle audience, porta a possibilità combinatorie inesauribili e una grande facilità di riproduzione tecnica. Fino al punto che il pubblico di riferimento non avrà un gusto sviluppatissimo per la qualità della scrittura (merce rarissima sempre, comunque), ma ha ormai le competenze narratologiche di un Vladimir Propp.

Vale la pena aggiungere una lettura temporale che la griglia del §6 suggerisce e che l'equazione da sola non vede. Le sedici posizioni non sono equivalenti nel tempo. Quelle che delegano la fase in cui l'autore impara — il momento in cui si scopre, scrivendo, di non aver capito — producono lo stesso testo oggi e un autore diverso domani. Il termine ecologico è la somma di questi effetti su molti autori e molti anni: la varianza fra i prior non cala soltanto perché i testi si somigliano, cala perché i sistemi cognitivi che li producono smettono di divergere.

L'altra considerazione da fare è che, all'interno del sistema, il “successo” (in termini quasi neoevoluzionistici, sulla scia dei meme di Dawkins) di un testo ha sempre un costo opportunità: che è precisamente l'occultamento di altri testi, che sono rispetto ad esso in rapporto di concorrenza paradigmatica. Il costo, in termini hjelmsleviani di “selezione” è un dato strutturale: non può non esserci, anche a costi marginali di distribuzione pari a zero, perché l'attenzione è una risorsa scarsa. Ma se a vincere è un testo di propaganda no vax, per esempio, il danno al sistema è per questa ragione ancora più rilevante. In termini morali, vincere implica una forma di responsabilità epistemica. Nella notazione del §8 è il fattore Λ a registrarlo: un testo ad ϵ alto che vince non sbaglia soltanto, occupa la posizione da cui l'errore si propaga, ed è la ragione per cui quel fattore va tenuto maggiore dell'unità.

10. Ancoraggio termodinamico

Il legame con la termodinamica non è metaforico e va reso esplicito, perché è quello che autorizza tutto il vocabolario adottato dal §1.

Per il principio di Landauer, la cancellazione di un bit costa almeno $k_B T \ln 2$ — una quantità piccolissima alle temperature ordinarie, ma diversa da zero, e questo è l'unico punto che conta: esiste un costo fisico minimo, e riguarda la cancellazione. Per Bennett, la computazione può essere resa logicamente reversibile, e il passaggio irriducibilmente dissipativo è appunto la cancellazione, cioè la selezione fra stati.

Scrivere è selezionare: ogni testo esiste in quanto scarto di tutti i testi che non sono stati scritti. **Il costo di un testo è sempre stato nella cancellazione delle alternative, mai nella produzione della stringa.** Gli strumenti generativi abbassano di ordini di grandezza il costo della produzione e lasciano intatto il costo della selezione, che è dove la termodinamica colloca il lavoro. Chi conclude che scrivere sia diventato gratuito ha misurato il termine sbagliato — e il fatto che questo risultato coincida con quello del §7, ottenuto per una via del tutto indipendente, è la ragione principale per cui credo che la costruzione stia in piedi.

11. L'equazione, e dove sta Σ

$$V(T; \mu) = \eta(\Sigma, F) \cdot \int_{\mathcal{R}} [G_{\text{oss}} + G_{\text{sel}} + G_{\text{ric}}] d\mu(R) - \Lambda \varepsilon(\Sigma, \nu) - \kappa \Delta \text{Var}[\mathcal{P}]$$

con F la competenza formale e ν l'intensità della verifica.

Letta da sinistra a destra: il valore di un testo per un pubblico è la somma di ciò che il testo porta — osservazione, selezione, ricombinazione — pesata sui lettori che raggiunge e moltiplicata per quanto bene è scritto; meno il danno degli errori che non sono stati corretti; meno la quota di appiattimento che quel testo contribuisce a produrre nel campo in cui esce.

Lo stack Σ compare tre volte, e tutte e tre le volte fuori dai termini di contenuto. Compare dentro η , dove è limitato superiormente da uno e satura. Compare dentro ε , dove è interamente compensabile aumentando ν , cioè verificando. Compare nel termine ecologico, dove agisce su scala di popolazione e non di testo, e dove nessun autore singolo può compensarlo. **In nessun termine additivo di contenuto Σ compare.**

Questa è la forma precisa dell'affermazione da cui sono partito. Chiedere se un testo sia stato scritto con l'AI significa chiedere il valore di una variabile che entra nell'equazione come coefficiente moltiplicativo saturante e come termine d'errore azzerabile per verifica. È una domanda legittima e piccola, e la griglia del §6 è il modo che ho scelto per risponderle senza pretendere che sia la domanda grande. Le domande grandi riguardano G_{oss} , G_{sel} e $\text{Var}[\mathcal{P}]$, e nessuna delle tre viene mai posta.

12. Autenticità

Resta da collocare la parola da cui il dibattito prende le mosse, e da cui è partito anche questo saggio.

Definisco un vettore di dichiarazione d — strumenti impiegati, punto del processo in cui sono intervenuti, verifica effettuata — e pongo

$$A(T) = \mathbf{1} [d = \text{processo effettivo}]$$

dove la scrittura $\mathbf{1}[\cdot]$ indica la funzione che vale uno quando ciò che sta fra parentesi è vero e zero quando è falso. In italiano: l'autenticità vale uno se la dichiarazione corrisponde a ciò che è stato fatto, e zero altrimenti.

Ne seguono tre proprietà, e sono le risposte alle domande che mi ero posto in apertura. **L'autenticità è una proprietà dei metadati, non del testo:** nessuna ispezione della prosa può stabilirla, il che spiega perché il dibattito sia condannato a girare a vuoto finché si ostina a cercarla dentro le pagine. È binaria: una dichiarazione o corrisponde o no, e non esistono gradi di autenticità. Ed è verificabile a costo quasi nullo, mentre il valore richiede un panel di lettori e qualche anno di attesa.

Soprattutto, non compare in V . Valore e autenticità sono grandezze ortogonali: esistono testi preziosi e non dichiarati, e testi dichiaratissimi e nulli. La confusione fra le due è, credo, l'errore che genera l'intero dibattito — e questo risponde alla terza domanda che mi ero posto: **no, l'autenticità non è un parametro adatto a misurare l'innovazione che un testo introduce in un sistema culturale.** Non lo è per una ragione strutturale e non

per una debolezza contingente della definizione: misura la sincerità della dichiarazione sul canale, e la sincerità di una dichiarazione non ha alcun rapporto con la quantità di neghentropia che il messaggio trasporta.

Ne segue anche il perché operativo di un colophon lungo, e la griglia delle sedici posizioni ne è la forma compatta: è il luogo dove d viene reso pubblico, ed è tutto ciò che la teoria richiede sul versante dell'autenticità. Il resto lo giudica chi legge.

13. Operazionalizzazione

Una teoria che non si lascia falsificare non serve. Le stime praticabili con la tecnologia disponibile oggi:

Grandezza	Stimatore
spostamento D	perplessità del testo rispetto a un modello calibrato sul corpus del ricevente
struttura ΔH	riduzione della dispersione fra ipotesi in un panel, prima e dopo la lettura
guadagno G	regole di punteggio proprie (Brier) sulle predizioni implicite del testo, risolte a posteriori
errore ε	tasso di errore su campione di affermazioni atomiche verificabili
varianza $Var[\mathcal{P}]$	varianza nello spazio degli embedding del corpus di un campo, serie storica

La prima riga rende quantitativa la domanda «per chi è nuovo questo testo», che oggi si discute solo per impressioni ed è invece la sola domanda che il §4 consideri sensata. La terza è il surrogato di q promesso al §3: non disponendo del mondo, si usano le predizioni che il testo implica e si aspetta di vedere quali si avverano. La quarta è già prassi corrente in redazione, sotto il nome di fact-checking, e la teoria si limita a spiegare perché costi tanto e valga tanto.

14. Limiti

Quattro, dichiarati.

I prior dei riceventi non sono distribuzioni di probabilità in senso proprio, e Ω è definibile con rigore solo per domini ristretti: il formalismo è esatto su questioni circoscritte e diventa analogico su testi ampi — incluso, va detto, questo.

La distribuzione q non è accessibile, e ogni stima di G deve surrogarla con un consenso posteriore o con predizioni risolte, il che introduce una dipendenza dal tempo che la notazione tace.

I coefficienti Λ e κ non sono derivati, sono postulati: la teoria sostiene che quei termini esistono e ne fissa il segno, non ne fissa il peso.

E la griglia del §6 dichiara due dei quattro luoghi in cui lo strumento può intervenire, tacendo proprio sul terzo — la selezione — per una ragione difendibile ma locale, cioè che in questa pratica non varia mai.

Nessuno dei quattro limiti tocca il risultato centrale, che è di struttura e non di misura: **Σ non compare in nessun termine additivo di contenuto, e un dibattito che lo tratta come la variabile principale sta leggendo l'equazione dalla parte del coefficiente.**

Bibliografia

- ASHBY, W. Ross, *An Introduction to Cybernetics*, Londra, Chapman & Hall, 1956.
- BENNETT, Charles H., “The Thermodynamics of Computation – a Review”, *International Journal of Theoretical Physics*, vol. 21, n. 12, 1982, pp. 905-940.
- BRILLOUIN, Léon, *Science and Information Theory*, New York, Academic Press, 1956.
- DAWKINS, Richard, *The Selfish Gene*, Oxford, Oxford University Press, 1976 (trad. it. *Il gene egoista*, Milano, Mondadori, 1992).
- ECO, Umberto, *Opera aperta*, Milano, Bompiani, 1962.
- HESSE, Hermann, *Siddhartha*, Berlino, S. Fischer, 1922 (trad. it. *Siddharta*, Milano, Adelphi, 1975).
- HJELMSLEV, Louis, *Prolegomena to a Theory of Language*, Baltimora, Waverly Press, 1953 (ed. or. *Omkring sprogteoriens grundlæggelse*, Copenhagen, Munksgaard, 1943; trad. it. *I fondamenti della teoria del linguaggio*, Torino, Einaudi, 1968).
- KAHNEMAN, Daniel, *Thinking, Fast and Slow*, New York, Farrar, Straus and Giroux, 2011 (trad. it. *Pensieri lenti e veloci*, Milano, Mondadori, 2012).
- LANDAUER, Rolf, “Irreversibility and Heat Generation in the Computing Process”, *IBM Journal of Research and Development*, vol. 5, n. 3, 1961, pp. 183-191.
- PROPP, Vladimir Ja., *Morfologia della fiaba*, Torino, Einaudi, 1966 (ed. or. *Morfologija skazki*, Leningrado, Academia, 1928).
- SCHRÖDINGER, Erwin, *What Is Life? The Physical Aspect of the Living Cell*, Cambridge, Cambridge University Press, 1944 (trad. it. *Che cos'è la vita?*, Milano, Adelphi, 1995).
- SHANNON, Claude E., “A Mathematical Theory of Communication”, *The Bell System Technical Journal*, vol. 27, 1948, pp. 379-423 e 623-656.
- STILL, Susanne, David A. Sivak, Anthony J. Bell e Gavin E. Crooks, “Thermodynamics of Prediction”, *Physical Review Letters*, vol. 109, n. 12, 2012.
- WIENER, Norbert, *Cybernetics: or Control and Communication in the Animal and the Machine*, Cambridge (Mass.), MIT Press, 1948 (trad. it. *La cibernetica*, Milano, il Saggiatore, 1968).

Intervento AI: DL·FM (Dialogo · Formalizzazione). 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/2026-08-23-la-formula-dellautenticita/>. Le parti interattive del saggio sono disponibili solo online.

Questo saggio 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>.