

Perchè i colossi dell'AI distruggono i libri?

di Antonio Dini



L'ultima novità è di poche settimane fa e viene dal Giappone. Nel paese del Sol Levante sta infatti succedendo una cosa curiosa: alcune librerie dell'usato hanno recentemente [registrato vendite quintuplicate](#), con ordini da centinaia di volumi alla volta, tutti diretti allo stesso centro logistico nella prefettura di Okayama.

Da lì, i pallet di libri partono verso gli Stati Uniti: oltre cinquanta tonnellate in totale, secondo le notizie raccolte dalle testate giornalistiche nipponiche. Nessuno ha ancora confermato chi sia il compratore finale, ma il percorso ricalca uno schema già noto altrove.

È lo stesso schema emerso in una causa giudiziaria che sta facendo la storia: *Bartz v. Anthropic*. Gli atti del processo sono stati resi noti dalla stampa americana e raccontano la storia di

un'azienda che ha comprato milioni di libri di seconda mano, li ha fatti smontare tagliando la rilegatura, li ha fatti scansionare pagina per pagina sostanzialmente distruggendoli, e infine li ha buttati al macero.

Quello di Anthropic non è l'unico caso. Un'inchiesta della testata 404 Media [ha seguito](#), grazie a un tracciatore nascosto in un lotto di volumi rari, un processo analogo che ha portato un lotto di alcune centinaia di libri comprato all'ingrosso fino a un magazzino Amazon vicino a Las Vegas, dove, anche in questo caso, alcuni lavoratori avrebbero raccontato di aver tagliato le rilegature, scansionato le pagine e gettato i resti.

La società non ha però confermato pubblicamente che i file siano destinati all'addestramento dei suoi modelli, né ha chiarito quanti libri siano coinvolti o quale criterio usi per selezionare i volumi. Amazon ha [risposto](#) che *“acquista libri tramite i canali commerciali per contribuire allo sviluppo e al miglioramento dei prodotti e dei servizi utilizzati dai nostri clienti”*.

Questo è uno dei punti comuni. Le aziende spesso non negano l'acquisto dei libri, ma evitano di confermare pubblicamente il processo completo: acquisto, scansione, uso nei dataset, distruzione e conservazione delle versioni digitali ricavate. Segnalazioni di pratiche simili arrivano anche da Regno Unito, Irlanda e altri Paesi europei.

Il filo che lega questi episodi è più interessante della singola notizia e conduce in una direzione molto chiara: la guerra senza quartiere in corso tra i colossi dell'intelligenza artificiale per ottenere la supremazia sviluppando nuovi modelli di frontiera sempre più potenti.

Per anni la battaglia sui dati con cui si addestrano i modelli linguistici ha riguardato soprattutto le persone: foto, conversazioni e tutte le altre tracce digitali lasciate in rete. Oggi quel fronte si è spostato sulla pagina stampata, perché i grandi archivi del web [cominciano a scarseggiare](#) e i libri restano un giacimento ancora relativamente poco sfruttato.

Anthropic, per esempio, [ha sostenuto che Claude](#) viene addestrato su una combinazione di dati disponibili pubblicamente sul web, dataset acquistati commercialmente e dati prodotti dall'azienda. Ha anche dichiarato che acquistare libri è una pratica ampiamente utilizzata nel settore dell'AI, e che nessuno dei suoi programmi di acquisizione dati compra e distrugge “*libri rari o antiquari*”.

Questa dichiarazione è importante, ma va letta con precisione. Non equivale a dire che Anthropic non abbia mai usato la cosiddetta “*scansione distruttiva*”. Gli atti giudiziari del caso *Bartz v. Anthropic* descrivono invece un programma chiamato [Project Panama](#), nel quale copie cartacee venivano acquistate, private della rilegatura, scansionate e poi scartate.

La dichiarazione aziendale insiste sulla categoria “*rari o antiquari*”, lasciando aperta la questione di libri fuori catalogo, specialistici o difficili da reperire, ma non formalmente classificati come antiquari.

Gli atti e le cause americane indicano che il problema riguarda più aziende, non soltanto Anthropic. [Meta avrebbe definito l'accesso](#) a grandi quantità di dati librari “essenziale per restare competitiva”.

OpenAI è stata accusata di aver usato libri protetti da diritto d'autore, e avrebbe riconosciuto di aver scaricato materiale da piattaforme pirata come Library Genesis (LibGen) e Z-Library

(spesso menzionate insieme ad altre “shadow libraries” come Sci-Hub e Bibliotik), sostenendo di averlo cancellato prima del lancio di ChatGPT. Questi elementi non provano però che tutte le aziende citate abbiano adottato lo stesso sistema di acquisto, taglio e distruzione fisica dei libri.

La distinzione è fondamentale: ci sono prove di una corsa generale ai libri, ma non esiste ancora una prova pubblica uniforme che ogni grande azienda abbia un “Project Panama” identico.

Materia prima pregiata

C'è un motivo se l'industria editoriale è entrata nel mirino dei colossi dell'AI, e riguarda la qualità del lavoro umano specializzato. In generale, un romanzo curato da un editor, un saggio storico, un manuale specialistico offrono una scrittura di qualità superiore a gran parte di quanto circola online, caratterizzata da testi e strutture coerenti e corretti, frutto di anni di lavoro redazionale.

A finire nella lista dei desideri delle aziende sono soprattutto testi scritti prima che l'intelligenza artificiale generativa cominciasse a riempire il web di contenuti sintetici, quindi genuinamente umani.

Per un modello linguistico questa è la materia prima più pregiata che esista perché evita del tutto il rischio del [“collasso del modello”](#) (*model collapse*), che si verifica invece quando l'addestramento avviene con contenuti generati da un'altra AI (o da versioni precedenti di se stessa).

Questa forma di endogamia digitale (non a caso chiamata anche, con una certa ironia, [Habsburg AI](#), l'“AI degli Asburgo”, dinastia nota per la sua endogamia) è un processo di addestramento

paragonabile a una fotocopia di una fotocopia: a ogni passaggio il modello addestrato perde sfumature, dettagli rari e aderenza alla realtà, accumulando errori e appiattendolo le risposte.

La soluzione al “*problema Asburgo*” dell’AI è trovare “*materiale umano*” fresco, meglio se verificato, com’è il caso dell’industria editoriale pre-ChatGPT. E di questo materiale ne serve tanto.

I numeri danno l’idea della scala. [Secondo Epoch AI](#), l’insieme del testo umano di qualità disponibile online è un numero attorno ai trecentomila miliardi di token, con un esaurimento previsto tra il 2026 e il 2032. Nello stesso periodo si è formato un mercato delle licenze per dataset che vale già alcuni miliardi di dollari l’anno.

Anthropic, per esempio, nel 2026 ha pagato un miliardo e mezzo di dollari per chiudere una causa legata a circa cinquecentomila libri scaricati da archivi pirata.

Qui entra in gioco il diritto d’autore. Chi compra una copia di un libro può rivenderla, prestarla o anche distruggerla senza chiedere permesso a nessuno. È il principio che rende possibili le biblioteche e le bancarelle dell’usato. Possedere il volume, però, non significa avere anche il diritto di riprodurne liberamente il contenuto, per esempio digitalizzandolo integralmente.

Sul versante opposto, va anche segnalato che la distruzione dei libri è tutt’altro che un’anomalia: gli editori mandano da sempre al macero le copie non vendute, restituite dai librai come parte del “reso”, cioè dell’invenduto. Che è però un modo per chiudere una perdita logistica, non per estrarne valore.

Lo scandalo del macero industriale

Quindi, la distruzione delle copie in sovrappiù c'è sempre stata. Allora perché la vicenda dei libri usati comprati dai produttori e poi distrutti diventa scandalosa? Il macero tradizionale elimina copie che nessuno vuole più leggere. Le aziende di AI, invece, comprano deliberatamente libri che sono sul mercato, che hanno ancora un valore informativo, ne assorbono il contenuto e poi ne distruggono il supporto fisico, spesso senza che autori ed editori vedano un euro. Se il volume era un'edizione rara, quel testo esce per sempre dal mondo condiviso della lettura, ogni libro sparisce davvero solo quando smettiamo di parlarne e di sfogiarlo.

Tuttavia, alcuni commentatori tecnologici e di orientamento libertario sostengono che la scansione distruttiva venga presentata in modo troppo apocalittico. Per esempio, Jack Nicastro, in un commento ripreso dall'Information Technology and Innovation Foundation, argomenta che il procedimento può convertire rapidamente libri fisici in testo ricercabile, aiutare a rispettare il copyright e ampliare l'accesso alla conoscenza. Secondo questa lettura, il problema principale sarebbe la trasparenza e non la distruzione materiale in sé.

La posizione della Commissione Europea è invece opposta: lo scorso 16 settembre la Commissione ha dichiarato che la digitalizzazione di un libro fisico protetto è un atto di riproduzione e che le eccezioni non coprono, nella sua interpretazione, la digitalizzazione di opere originariamente fisiche.

Inoltre, l'European Writers Council ha condannato la scansione distruttiva di milioni di libri, sostenendo che non si tratta soltanto di copyright, ma di rispetto per il lavoro creativo e per il patrimonio culturale. Questa posizione considera il libro non un semplice contenitore di token: un volume può avere valore come

oggetto, come esemplare raro, come testimonianza di una biblioteca, come testo annotato o come unica copia ancora reperibile. Il PDF può salvare le parole, ma non conserva integralmente la storia materiale del libro.

È difficile capire quale possa essere il punto di vista corretto, soprattutto se l'obiettivo è di preservare la cultura di cui i libri sono simbolo oltre che contenitori. Un'altra analisi pubblicata da Stanford difende la scansione distruttiva in linea di principio: anche biblioteche e archivi possono smontare copie duplicate per preservare il contenuto e renderlo studiabile. La condizione, però, è che la procedura sia usata correttamente, su copie adeguate e dentro una politica di conservazione responsabile.

Da un lato, digitalizzare una copia duplicata e depositare il file in un archivio controllato può avere un valore conservativo.

Dall'altro, è facilmente dimostrabile che comprare indiscriminatamente migliaia di libri fuori catalogo, tagliarli in modo irreversibile, non rendere pubblico l'elenco dei titoli e usare il risultato per addestrare prodotti commerciali è invece molto più problematico.

Europa e Stati Uniti: stessi libri, leggi diverse

Quello che Anthropic e le altre aziende stanno facendo nasce, come detto, dal bisogno di dati freschi e dagli spazi lasciati dal diritto vigente, soprattutto negli Stati Uniti: la legge per ora permette più di quanto il senso comune sia disposto a concedere.

Negli Stati Uniti il giudice che ha istruito il caso Anthropic ha separato con cura i libri scaricati da biblioteche pirata dai libri comprati legalmente e poi scansionati. Ha ritenuto legittimo l'uso dei testi per addestrare i modelli e la digitalizzazione delle copie

acquistate, mentre ha escluso che il *fair use* potesse giustificare l'acquisizione e la conservazione di libri ottenuti da archivi pirata. Si tratta di una sentenza di merito legata a quella singola fattispecie, ma è stata interpretata come un lasciapassare generale.

Chi si oppone a questa interpretazione, autori ed editori inclusi, nota però che il proprietario di una copia comprata legalmente può altrettanto legalmente rivenderla o distruggerla, ma questo non gli concede automaticamente il diritto di riprodurre il contenuto su scala industriale.

In Europa il quadro è più prudente e non ricalca quello americano. Le eccezioni per l'estrazione di testi e dati previste dalla normativa comunitaria permettono in alcuni casi di analizzare opere a cui si abbia accesso lecito, ma gli autori possono riservarsi esplicitamente i diritti per gli usi commerciali.

Per questo le segnalazioni arrivate da librerie del Regno Unito, dell'Irlanda e di almeno sette Paesi del continente (non risulta ad oggi che l'Italia sia tra questi) mostrano che il fenomeno non è confinato agli Stati Uniti. La conclusione di un tribunale californiano non si trasferisce da sola nel diritto europeo: davanti a un giudice del Vecchio continente il risultato con tutta probabilità sarebbe molto diverso.

Anthropic, peraltro, non è sola in questa corsa. Ci sono Meta e OpenAI, probabilmente anche Amazon. Le aziende hanno fornito dichiarazioni parziali: confermano acquisti commerciali o descrivono in modo generale le fonti dei dati, ma raramente pubblicano liste complete dei libri, procedure di conservazione, criteri per escludere opere rare e informazioni verificabili sul percorso dei file.

Nessuno di questi elementi dimostra infatti che tutte le aziende seguano lo stesso protocollo di acquisto, taglio e distruzione fisica, ma conferma che la caccia ai libri è un fenomeno parte di un mercato ormai strutturato, non un'anomalia isolata.

Resta, sullo sfondo, un deficit di trasparenza che accomuna quasi tutti i protagonisti della vicenda. In Giappone l'associazione degli editori ha chiesto chiarimenti al grande distributore Nippan, il cui nome compare negli atti del contenzioso americano, senza ottenere finora risposte sulle singole transazioni.

Librai e artisti contro l'AI

Questa storia si inserisce in una vicenda più ampia che da alcuni decenni vede contrapposta l'industria culturale alle multinazionali tecnologiche. Dalle [guerre a Napster](#) alla fine degli anni Novanta fino agli scontri di oggi fra artisti e colossi dello streaming, tutto ruota intorno alla ridistribuzione dei profitti: artisti pagati pochi centesimi per ascolti che generano fatturati miliardari.

È la stessa guerra che ha mandato in [sciopero per 148 giorni](#) gli sceneggiatori di Hollywood, preoccupati che l'AI potesse scrivere al posto loro togliendo lavoro futuro (alla fine hanno ottenuto la garanzia che nessun testo generato da un modello potesse valere come materiale letterario), ma finora nessuno si era posto il problema, enorme, della sostituzione anche fisica del libro in quanto strumento più antico e potente per la trasmissione della cultura delle nostre società.

Adesso, il libro, oggetto apparentemente innocuo su uno scaffale, è diventato il terreno di uno “scontro finale” tra chi crea e chi addestra. Tuttavia, le storie che circolano online secondo le quali i grandi dell'intelligenza artificiale “*stanno distruggendo tutti i libri*

del mondo”, “*ogni volume giapponese finisce in Anthropic*”, “*tutte le aziende bruciano opere uniche*” vanno ridimensionate.

Il fenomeno non è una leggenda nata sui social. La vera novità però è che la trasformazione del libro in materia prima industriale per l’AI sta diventando abbastanza vasta da produrre effetti visibili nei mercati dell’usato, nei rapporti fra editori e distributori e nel dibattito sulla conservazione culturale.