

Anche NVIDIA avrebbe piratato dei libri per addestrare la sua IA

L'idea originaria di un internet egualitario e senza padroni è stata, negli anni, fagocitata e annichilita da una realtà dominata da poche grandi aziende tecnologiche capaci di imporsi in modo sempre più incontrollato. E, talvolta, questa assimilazione è più letterale di quanto si possa immaginare. Secondo gli atti di un procedimento legale, **Anna's Archive** - l'enorme archivio di libri, articoli scientifici, fumetti e documenti provenienti dalle principali "biblioteche ombra" del web libero - avrebbe finito per soddisfare le ambizioni di **NVIDIA**, gigante tech che avrebbe richiesto accesso a file di ogni tipo, compresi **materiali illeciti**, pur di alimentare i propri modelli di intelligenza artificiale.

La notizia, riportata inizialmente da [TorrentFreak](#), evidenzia come l'aggiornamento di una **class action** depositata agli inizi del 2024 presso la Corte federale statunitense abbia portato alla luce preziosi retroscena riguardanti una delle aziende più influenti al mondo. Non si tratta infatti di un semplice aggiustamento procedurale, bensì dell'inserimento di una serie di documenti interni che, qualora confermati come autentici, dimostrerebbero esplicitamente come alcuni dirigenti avessero richiesto esplicitamente materiali che sapevano essere stati **diffusi online in maniera illegale**.

La natura stessa di [Anna's Archive](#) è al centro di accesi dibattiti legali. Il progetto mira a costruire una **biblioteca universale** che sia capace di raccogliere ogni tipo di libro, rivista o pubblicazione scientifica, rendendola disponibile al maggior numero possibile di persone. Un obiettivo che incarna ideali di libero accesso alla conoscenza e che, nella pratica, rappresenta anche uno strumento potente contro censure e oscurantismi di Stato. Tuttavia, per sua stessa struttura, l'approccio divulgativo della no profit si sovrappone in modo marcato alla **violazione del diritto d'autore**, una promiscuità che si accattiva regolarmente le ire formali di molteplici governi.

Nonostante la natura legalmente ambigua del progetto, NVIDIA avrebbe comunque deciso di attingere ai file condivisi da Anna's Archive, pur dopo essere stata avvertita dagli stessi gestori che gli archivi richiesti contenevano **materiale piratato**. Uno scambio di email mostra infatti come il team di data strategy dell'azienda sia entrato in contatto con i responsabili del portale per ottenere nuovi documenti con cui **addestrare il proprio modello di intelligenza artificiale**, NeMo. Avvertiti della natura illecita dei contenuti, i tecnici avrebbero comunque fornito l'autorizzazione a procedere. In questo modo NVIDIA avrebbe ottenuto l'accesso a un archivio di circa **500 GB** che, secondo l'accusa, sarebbe stato successivamente affiancato da ulteriori fonti pirata, tra cui gli archivi LibGen, Sci-Hub e Z-Library.

Dal canto suo, NVIDIA [sostiene da tempo](#) che l'esfiltrazione di dati dal web per addestrare sistemi di intelligenza artificiale rientri nel perimetro del "**fair use**". Tuttavia, recenti

Anche NVIDIA avrebbe piratato dei libri per addestrare la sua IA

sviluppi giudiziari suggeriscono che questa posizione potrebbe non reggere, soprattutto quando viene dimostrato l'impiego di materiale ottenuto in palese violazione del diritto d'autore. Lo [scorso giugno](#), un giudice californiano ha stabilito che Anthropic non aveva infranto il copyright nell'utilizzare, **senza il consenso di autori ed editori**, numerosi libri per l'addestramento della propria IA. Secondo la corte, il processo di trasformazione operato dai modelli di intelligenza artificiale era sufficientemente profondo da **escludere la dimensione del plagio**.

Allo stesso tempo, il tribunale ha precisato che quella decisione non esonerava comunque l'azienda dalle responsabilità legate all'aver ottenuto i documenti in modo illegale, ricorrendo a pratiche del tutto simili a quelle che oggi vengono contestate a NVIDIA. La Big Tech, però, non è l'unica destinata a subire le conseguenze di queste rivelazioni: Anna's Archive si trova sotto pressioni crescenti. Il sito è finito al centro di un caso di portata eccezionale quando, [a dicembre](#), è riuscito a copiare gran parte dell'archivio del servizio musicale **Spotify**, replicando una quantità di dati che rappresenterebbe il 99,6% degli ascolti totali della piattaforma. Pochi giorni dopo, a inizio gennaio, il principale dominio della no profit — annas-archive.org — [è stato oscurato](#) dal suo fornitore a seguito di un [ordine restrittivo](#) richiesto dalle principali case discografiche statunitensi.

*Aggiornamento 22/01/25: in fase di pubblicazione di questo articolo sono emerse le carte giudiziarie che contestualizzano la sospensione di alcuni domini di Anna's Archive. L'informazione è stata integrata nel testo.*



## Walter Ferri

Giornalista milanese, per *L'Indipendente* si occupa di analisi nel campo della tecnologia, dei diritti informatici, della privacy e dei nuovi media, indagando le implicazioni sociali ed etiche delle nuove tecnologie. È coautore e curatore del libro *Sopravvivere nell'era dell'Intelligenza Artificiale*.

Anche NVIDIA avrebbe piratato dei libri per addestrare la sua IA



## ***Vuoi approfondire l'argomento?***

***Ventitré esperti di livello internazionale  
selezionati da L'Indipendente, affrontano  
con chiarezza e rigore i principali aspetti  
sociali, individuali e tecnologici del futuro  
che ci attende con la diffusione dell'IA.***

**Acquista ora**