

Così l'intelligenza artificiale voleva distruggere tutti i libri del mondo



Potrebbe essere il plot di una nuova versione di “Fahrenheit 451”, riveduta e molto, molto corretta. Lì, nel romanzo di Ray Bradbury, la conoscenza era bandita e ai “pompieri” veniva affidato il compito di bruciare qualsiasi tipo di libro. Qui la conoscenza è stata trasformata a ritmo frenetico per sfamare la bestia insaziabile dell’intelligenza artificiale, acquisendo milioni di libri, tagliarne il dorso, smembrarli e scansionare le pagine su scanner ad alta velocità. Comunque sia, anche in questo caso per eliminarli.

La folle corsa all’accaparramento è diventata cronaca di giornale quando un gruppo di autori ha deciso di fare causa ad

Anthropic per violazione delle leggi americane sul copyright, con la richiesta di 183 miliardi di dollari di danni: un'enormità. Ad agosto la compagnia di San Francisco ha accettato di pagarne un miliardo e mezzo. Tutto sommato un affare, e sembrava finita lì.

Solo che, una decina di giorni fa, la scelta di un giudice californiano di desecretare parte degli incartamenti depositati ha permesso di aggiungere elementi su come i signori dell'intelligenza artificiale stanno costruendo il loro dominio. Tutto comincia all'inizio del 2024, quando i dirigenti della startup fondata tre anni prima da alcuni transfughi di OpenAI varano il Progetto Panama.

Sembra l'incipit di una spy-story e, in effetti, c'è tutta la volontà di mantenere segreto l'obiettivo da raggiungere: «Scansionare in modo distruttivo tutti i libri del mondo». Nell'anno successivo Anthropic spende una montagna di denaro per rastrellare i volumi da dare in pasto al suo chatbot. Si chiama Claude e per molti è già più avanti di ChatGPT, soprattutto nella scrittura e nell'editing. Premessa: dietro l'IA generativa come è conosciuta da noi utenti finali – faccio una richiesta anche complessa e in breve ottengo una risposta – c'è una realtà più articolata, almeno fino al prossimo salto tecnologico sul quale in questi giorni, ironia della sorte, ha lanciato l'allarme proprio il ceo di Anthropic, Dario Amodei, con [un saggio di 38 pagine](#) in cui si preoccupa di quella che chiama “Intelligenza artificiale potente”, capace di sfuggire al controllo umano prima di quanto oggi si immagini.

Nell'attesa non è ancora una vera intelligenza, ma un imitazione del comportamento umano basata su calcoli statistici e capacità di elaborare enormi quantità di dati. Una scimmia ammaestrata ipertecnologica che necessita per scrivere meglio di miliardi di parole da assimilare. Di parole di senso compiuto e cosa meglio dei libri? Di tutti i libri del mondo. Ecco, dal processo californiano emerge che nel

gennaio del 2023 proprio Amodei teorizzò che per addestrare al meglio i modelli di Claude era più utile farlo sui libri invece di imitare il «linguaggio di bassa qualità reperibile dai testi sul web». Con due problemi da affrontare: i libri avevano autori ed editori da pagare; i concorrenti di Anthropic avevano lo stesso obiettivo e la stessa fretta.

Il caso Anthropic non è isolato: è solo quello giunto prima a conclusione e fa parte di un'ondata di cause legali intentate contro le aziende di intelligenza artificiale, da Meta a Google, a OpenAI o Microsoft. Protagonisti di «una corsa frenetica, a volte clandestina, per acquisire le opere complete dell'umanità» ha scritto [il Washington Post](#), che sul caso ha molto investigato.

Pur non essendo di per sé criminoso rastrellare milioni di copie di libri e distruggerli, emerge dalle carte messe insieme dai giudici che i signori dell'IA abbiano ritenuto utile e sbrigativo farlo senza attendere le autorizzazioni o stipulare contratti, anzi all'insaputa degli autori.

In una prima fase, le compagnie cominciarono a scaricare i libri digitalizzati da alcune "biblioteche ombra" come Pirate Library Mirror o LibGen che, in maniera programmatica, si facevano vanto di violare le normative sul copyright e che per questo avevano perso delle cause negli Stati Uniti. Dalla messaggistica di Meta, agli atti in un altro processo in California, per esempio, si apprende che in diverse occasioni i manager si dissero preoccupati dallo scaricare online milioni di titoli piratati, ma già a fine 2023 in un'altra mail interna si affermava che la pratica era stata approvata dopo una «escalation a MZ», con quell'MZ che fa pensare alle iniziali di Mark Zuckerberg, il padrone del colosso di Facebook, Instagram e Whatsapp.

I tipi di Anthropic avevano preferito la scansione distruttiva proprio per non essere etichettati come pirati del web dediti a un *torrenting* di proporzioni colossali, assoldando anche un

esperto della materia: l'ex dirigente di Google che nei primi anni Duemila aveva condotto la discussa impresa di Google Libri (la creazione del più grande database di testi mai esistito fino ad allora), naufragata dopo una guerra legale durata 12 anni e comunque vinta dall'azienda di Mountain View.

Quella vittoria oggi fa sperare gli avvocati dei signori dell'intelligenza artificiale, perché anche due recenti sentenze hanno ritenuto prevalente il ["fair use"](#), istituto proprio del diritto americano. Cioè: copiare materiale protetto dal diritto d'autore è legale se fatto per un uso limitato e "trasformativo", se non fosse che definire "trasformativo" è costato tempo e milioni di dollari già diverse volte nel passato. Tutto o quasi è affidato all'interpretazione del giudice che nella prima delle due sentenze ha addirittura paragonato le tecniche di addestramento a quelle di un benemerito insegnante che «addestra i suoi studenti a scrivere meglio».

Riuscire, poi, a dimostrare che l'intelligenza artificiale avesse danneggiato la vendita dei libri (questa l'obiezione contenuta nella seconda delle due sentenze) sarebbe stato quasi impossibile.

A conti fatti, Anthropic è l'unica finora ad aver sottoscritto un accordo miliardario, e ha acquistato copie cartacee, non scaricato gratis dal web. Quasi una beffa. Interpellata dal Washington Post, la startup ha affidato la risposta a uno dei suoi consulenti legali: «La questione su cui ci siamo accordati riguardava il modo in cui alcuni materiali venivano acquisiti, non se potessimo utilizzarli per sviluppare modelli di intelligenza artificiale».

La transazione raggiunta dalla società di Amodei e la causa ancora in corso che coinvolge Meta non sono isolate. A New York nomi di primo piano come David Baldacci, Michael Connelly, Jonathan Franzen, John Grisham o George R.R. Martin

sono in causa con OpenAI e Microsoft, mentre due editori del calibro di Hachette e Cengage chiedono di essere ammessi a una class action contro Google. Da due anni va avanti la disputa che vede il New York Times confrontarsi (anche qui) con Microsoft e OpenAI dopo la denuncia depositata alla corte federale di Manhattan, e la scorsa settimana Universal Music ha di nuovo citato in giudizio Anthropic con l'accusa di avere scaricato illegalmente almeno 20 mila canzoni.

Il confine tra le ragioni dell'innovazione tecnologica e la tutela della creatività umana è lungi dall'essere tracciato, nel mentre la natura predatoria dell'intelligenza artificiale è sempre più evidente, e la rincorsa all'accaparramento delle proprietà intellettuali ne è un aspetto non trascurabile. L'immagine che restituisce il clima resta quella di Donald Trump nel giorno dell'insediamento, attorniato dagli oligarchi digitali: da normative più o meno restrittive sul diritto d'autore, sulla privacy o sulla concorrenza dipende gran parte del loro predominio. Con i Borgiani – l'efficace definizione coniata da Giuliano Da Empoli nel suo "L'ora dei predatori" – bisognerà sempre più fare i conti.