

In America si è scoperto che l'intelligenza artificiale saccheggia subdolamente i giornali

Andrea Secchi a pag. 19

Un'analisi tecnica della News Media Alliance svela come si alimenta l'intelligenza artificiale

Così l'IA saccheggia i giornali Editori Usa: ingerisce contenuti protetti e ne usa pure le frasi

DI ANDREA SECCHI

«**Uso pervasivo e non autorizzato dei contenuti degli editori da parte degli sviluppatori di intelligenza artificiale generativa». È la conclusione di uno studio commissionato dalla News Media Alliance americana, l'associazione che riunisce 2.200 editori, da Hearst a Condé Nast, dalla società del New York Times a quella del Washington Post. Non semplicemente un allarme generico: l'associazione ha fatto realizzare un'analisi tecnica che documenta questa accusa e l'ha inserita in un libro bianco di 77 pagine in cui approfondisce ciò che sta accadendo nello sviluppo dell'AI e come possano essere evitati i danni per l'editoria. Il documento, infatti, analizza anche l'impatto che questa pratica potrebbe avere «sulla sostenibilità e la disponibilità di contenuti originali di alta qualità», oltre che le implicazioni legali di tale utilizzo.**

L'analisi mostra come nei set di dati che le aziende utilizzano per costruire i large language model, che alimentano poi i prodotti di intelligenza artificiale generativa come ChatGpt e gli altri, si utilizzino ampiamente contenuti informativi di giornali e media digitali generalmente senza autorizzazione o collegamento alla fonte originaria.

Non solo, però: alcuni dei modelli linguistici di grandi dimensioni danno un peso notevolmente superiore ai contenuti che arrivano dagli editori rispetto a quelli raccolti sul web di libero utilizzo: nell'addestramento dell'IA un articolo ha un peso da 5 a 100 volte superiore rispetto ai contenuti generici. Cosa che mostra il valore che ha il lavoro di editori e giornalisti in questo processo.

Il contenuto, poi, è sfruttato anche nella generazione delle risposte: i chatbot, in particolare, riproducono pezzi, frasi degli articoli, «dimostrando che i modelli conservano e possono memorizzare il contenuto espressivo delle opere di addestramento». Una «copia pervasiva», ribadisce il white paper, che viola i diritti degli editori.

C'è infine un ulterio-

re elemento che chiarisce quale sia il rapporto contenuti editoriali-IA: «La copia dell'intelligenza artificiale generativa per "addestramento" non serve a uno scopo diverso dalle opere originali perché i large language model tipicamente ingeriscono (cioè copiano) preziose notizie, riviste e contenuti web dei

media digitali per la loro espressione scritta, in modo da poter imitare quella stessa forma di espressione».

Le risposte dei modelli di intelligenza artificiale generativa «competono anche direttamente con il contenuto protetto che è stato copiato e utilizzato per addestrarli». Non si

aiutano semplicemente gli utenti a navigare verso le fonti originali: gli sviluppatori si vantano del fatto che gli utenti non hanno più bisogno di accedere o consultare tali fonti, sostiene il rapporto.

L'IA in questo modo non sta imparando, continua la News Media Alliance: non impara dai fatti rac-

contati negli articoli, non assorbe i concetti, conserva nei propri materiali di addestramento copiati le espressioni dei fatti che sono contenuti nelle opere e che il copyright protegge. Non si può, per tutto questo, invocare il fair use, l'uso senza autorizzazioni dei contenuti, non fosse altro perché gli utilizzi dei chatbot sono commerciali.

Nel documento, però, non c'è una condanna senza appello. La News Media Alliance, che ha presentato l'analisi anche all'U.S. Copyright Office, spiega che «gli sviluppatori di Gai (intelligenza artificiale generativa, ndr) e gli editori possono lavorare insieme per evitare tali risultati drammatici», ovvero il danneggiamento del settore dell'informazione e dei media digitali, cosa che si ritorcerebbe anche contro all'IA: la soluzione sarebbe una cooperazione in modo che l'intelligenza artificiale massimizzi il suo valore «in un modo che sia equo per gli editori e condivida equamente la ricchezza generata dal contenuto». «I sistemi di intelligenza

artificiale generativa dovrebbero essere ritenuti responsabili proprio come qualsiasi altra attività», ha detto Danielle Coffey, presidente e ceo dell'Alliance. «L'uso continuato e non autorizzato danneggerà i mercati esistenti che riconoscono il valore dei contenuti di qualità archiviati e in tempo reale, e col tempo gli stessi modelli Gai si deterioreranno. Ottieni ciò che inserisci. È fondamentale che le nostre protezioni sul copyright siano adeguatamente applicate e che elevati standard di qualità e responsabilità siano il fondamento di queste e di altre nuove tecnologie».

© Riproduzione riservata

- Le aziende che sviluppano IA generativa copiano in gran quantità contenuto degli editori senza autorizzazione e gli danno peso anche 100 volte superiore rispetto a ciò che trovano liberamente online
- La copia costituisce violazione del diritto d'autore: non può essere giustificata dalla dottrina del fair use
- Le risposte dei modelli di IA generativa competono direttamente con il contenuto protetto che è stato copiato, andando oltre il semplice scopo di ricerca
- La News Media Alliance invita gli sviluppatori a collaborare per massimizzare il valore dell'IA Generativa, in modo equo per gli editori e la società, condividendo la ricchezza generata dai contenuti

