

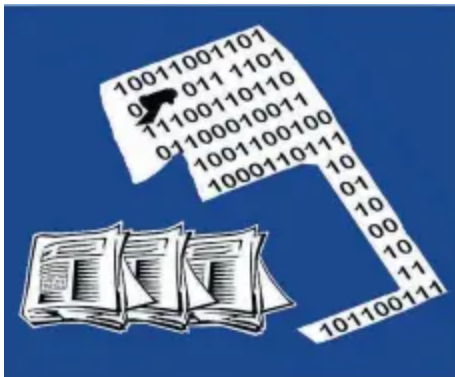
L'INTELLIGENZA ARTIFICIALE E LA PROPRIETÀ INTELLETTUALE

Diritto d'autore e furti I dati e le informazioni sono le enormi fondamenta (e le risorse) della rivoluzione digitale in corso

Corriere della Sera · 31 Dec 2023 · 30 · Di Roberto Battiston

La denuncia del New York Times contro OpenAI e Microsoft, rispettivamente creatore e principale utilizzatore commerciale di ChatGPT, è il primo esempio importante di causa per danni economici relativamente all'utilizzo dell'intelligenza artificiale generativa. Vengono in mente le class action contro la potente industria del tabacco o del petrolio. Il cuore del problema riguarda cosa sia la proprietà dei dati e dell'informazione nel contesto delle tecnologie digitali generative. Si tratta di una questione concreta: ChatGPT è un potente rastrello digitale, stupido come un rastrello, ma allo stesso tempo efficacissimo nel trattare le informazioni presenti nell'enorme quantità di dati oggi disponibili in rete a vent'anni dalla nascita del World Wide Web. Senza dati, le capacità dell'intelligenza artificiale generativa scompaiono istantaneamente: solo con data base di enormi dimensioni e della massima qualità si ottengono risultati interessanti. La denuncia del NYT mostra in modo documentato come una parte sostanziale dell'addestramento di ChatGPT sia stata ottenuta utilizzando articoli della prestigiosa testata statunitense o della rete di servizi di informazione ad essa collegati. Il testo della denuncia di 69 pagine mostra ad esempio come tra i data base usati per l'allenamento dell'algoritmo, il New York Times sia al quarto posto dopo le pagine, pubbliche, dei brevetti e quelle di Wikipedia. Ma le pagine del New York Times sono il risultato di uno sforzo professionale di grandissime proporzioni: si tratta di centinaia di migliaia di articoli di qualità, protetti da copyright, grazie a cui la testata continua a crescere in un periodo di profonda crisi della stampa. Grazie a domande precise, fatte proprio a ChatGPT, gli avvocati mostrano come l'algoritmo risponda spesso fornendo lunghe parti di articoli del NYT: non è un caso che questa sorgente di informazioni di qualità sia stata utilizzata sistematicamente da OpenAI e Microsoft. La qualità delle risposte dell'AI generativa è strettamente legata alla qualità delle informazioni disponibili. Inevitabile l'accusa di sistematico furto di diritti d'autore. Ci sono vari modi per sottrarre la ricchezza altrui, l'intelligenza generativa è semplicemente uno dei più tecnologici ed efficaci. La denuncia mostra inoltre come, in alcuni casi, le allucinazioni tipiche di questi algoritmi abbiano portato ad attribuire al NYT informazioni false o addirittura articoli non esistenti, con i relativi danni di immagine. Ma queste considerazioni non si applicano solo alle testate giornalistiche, si applicano anche alle pagine dei brevetti, mantenute con fondi pubblici o alle pagine di Wikipedia che vive di donazioni collettive: si tratta di prodotti dell'intelletto umano utilizzati gratuitamente per sostenere la lucrativa attività degli algoritmi generativi. Considerazioni simili si possono applicare, ad esempio, anche ai milioni di bloggers, che in cambio del loro contributo, non ricevono neanche un like ma contribuiscono a dare parvenza di

intelligenza ad un algoritmo che produce massicci ritorni economici per le grandissime imprese che lo hanno sviluppato: nel corso di meno di un anno il valore di OpenAI ha raggiunto 80 miliardi di dollari, 80 volte l'investimento fatto 8 anni fa.



Fino a poco tempo fa, la differenza tra ciò che è mio e ciò che è tuo, era regolamentata dalle leggi sulla proprietà privata per gli oggetti fisici e dalle leggi sulla proprietà intellettuale per le creazioni del pensiero. Il denaro rappresentava la necessaria mediazione che regolamentava lo scambio di questi beni. L'avvento del web e della diffusione digitale ha sconvolto questo scenario, introducendo servizi forniti gratuitamente, che però producono informazioni sulle nostre abitudini, pensieri, desideri, dati che a loro volta determinano le strategie commerciali del mondo delle imprese. È buona norma diffidare di ciò che è troppo facile, resistere al fascino del paese dei balocchi digitale in cui tutto sembra gratuito. Ci abbiamo messo un po' di tempo a capirlo, ma, con l'avvento di ChatGPT, questo aspetto balza agli occhi. Se questo rastrello digitale ha qualche forma di intelligenza, essa non è altro che il prodotto di centinaia di milioni di persone che giorno dopo giorno producono, nei propri contesti, senso, creatività, originalità, informazione. Non siamo ancora abbastanza coscienti del valore commerciale di ogni nostro singolo atto che si traduca in dati digitali. Questa immensa, diffusa ma fino a poco tempo fa poco sfruttata risorsa rappresenta un tessuto alla base delle relazioni sociali. I dati e le informazioni sono le fondamenta della rivoluzione digitale in corso. Le opportunità collegate all'intelligenza artificiale generativa sono enormi, così come le sfide. Ma in un mondo digitalizzato è necessario risolvere alla radice il problema di che cosa è mio e di che cosa è tuo e di come si gestiscono le transazioni economiche collegate allo scambio di informazioni e dati. Altrimenti pochissimi beneficerebbero di enormi proventi a spese di un'infinità di persone che perderanno una parte significativa del loro benessere, se non, addirittura, il proprio lavoro.