

Diritto d'autore e Intelligenza Artificiale: L'azione legale del *New York Times* contro *Open AI* solleva un problema che non possiamo più ignorare?

📅 12/01/2024

📌 PROPRIETÀ INTELLETTUALE, PRIVACY E CYBERSECURITY, SOCIETÀ, PROSPETTIVE.

Jacopo Piemonte
Giulia Beneduci
Federico Aluigi

L' intelligenza artificiale ("IA") è da diverso tempo uno dei maggiori argomenti di discussione nel dibattito pubblico.

Negli ultimi mesi si è assistito a intensi scambi sul piano continentale, in un quadro dove l'Unione Europea si accinge a fornire la prima regolamentazione a livello mondiale su tale fenomeno. Inoltre, il prossimo semestre del G7, presieduto dall'Italia, avrà come tema l'impatto che l'intelligenza artificiale potrà avere sul mercato del lavoro (soprattutto nell'ottica delle possibili perdite

occupazionali che le nuove tecnologie potranno determinare).

Non sarà ora sfuggito ai più attenti osservatori che, verso la fine dello scorso anno, in data 27 dicembre 2023, la nota casa editrice *The New York Times Company* ("NYT") ha intentato un'azione legale contro *OpenAI*, la società proprietaria di ChatGPT, e *Microsoft*, accusandoli di violazione del diritto d'autore¹.

Questo sviluppo apre un nuovo fronte "caldissimo" e potrebbe portare a uno stravolgimento nell'ambito dei rapporti tra proprietà intellettuale e Intelligenza Artificiale, inaugurando una battaglia

¹ Per consultare l'articolo pubblicato dallo stesso *New York Times*, si veda il seguente [LINK](#).

legale con possibili esiti “*game changing*” sull'uso non autorizzato di opere pubblicate per l'addestramento dell'IA.

L'Azione di *The New York Times* Company contro *Open AI* per violazione del diritto d'autore

Veniamo dunque ai dettagli specifici dell'azione intentata dal NYT. Secondo quanto riferito dalla stessa società giornalistica, milioni dei suoi articoli sono stati utilizzati per addestrare i *chatbot*, software che simulano ed elaborano conversazioni umane scritte o parlate, ponendoli ora in competizione con il giornale come fonte affidabile di informazioni. La rilevanza di questa questione si estende oltre la singola controversia, poiché avrà un impatto significativo sull'intera industria dell'IA e sui produttori di contenuti, che dovranno considerare attentamente le implicazioni di questo evento nei loro futuri programmi e strategie.

Il documento contenente la denuncia² si basa sulla tesi per cui ChatGPT avrebbe utilizzato milioni di articoli della casa editrice, i quali avrebbero “alimentato” il *chatbot* rendendolo in grado di competere direttamente con le produzioni giornalistiche del NYT³. Viene peraltro allegato come l'utilizzo indiscriminato e gratuito di detta proprietà intellettuale abbia portato ad un guadagno, per *Open AI*, di circa 90 miliardi di dollari⁴.

La denuncia, oltre i meri fatti, rivela la sua importanza storica e giuridica nelle argomentazioni, mettendo in luce le “epocali” problematicità nascenti dall'impatto tra attività umana ed emulazione/produzione dell'IA. La casa editrice evidenzia infatti, nelle prime battute del documento, le proprie radici storiche nella sua attività di giornalismo indipendente all'interno della società civile statunitense nonché internazionale; sottolinea la “*professional expertise*” degli innumerevoli collaboratori e la complessità di tutta la catena di produzione, e soprattutto afferma che “*per produrre un giornalismo di livello mondiale, il Times investe un'enorme quantità di tempo, denaro, competenze e talenti, sia nella redazione che nei team di prodotto, tecnologia e di supporto*”⁵; ed ancora, che “*sono necessarie enormi risorse per pubblicare, in media, più di 250 articoli originali al giorno. Molti di questi articoli richiedono mesi, e a volte più tempo, per essere pubblicati. Questa produzione è frutto del lavoro di circa 5.800 dipendenti a tempo pieno del Times (al 31 dicembre 2022), di cui circa 2.600 direttamente coinvolti nelle operazioni giornalistiche del Times*”⁶.

Su queste premesse, viene successivamente argomentato come *Open AI* osti alla sopravvivenza di dette attività, e più ampiamente, al giornalismo nel suo complesso. La casa editrice afferma di dipendere da sempre dai suoi diritti esclusivi di riproduzione, adattamento, pubblicazione, esecuzione

² Per consultare il documento ufficiale, si veda il seguente [LINK](#).

³ Si veda il punto I.2. del documento, per cui “... *Defendants' unlawful use of The Times's work to create artificial intelligence products that compete with it threatens The Times's ability to provide that service. Defendants' generative artificial intelligence (“GenAI”) tools rely on large-language models (“LLMs”) that were built by copying and using millions of The Times's copyrighted news articles, in-depth investigations, opinion pieces, reviews, how-to guides, and more. While Defendants engaged in widescale copying from many sources, they gave Times content particular emphasis when building their LLMs—revealing a preference that recognizes the value of those works. Through Microsoft's Bing Chat (recently rebranded as “Copilot”) and OpenAI's ChatGPT, Defendants seek to free-ride on The Times's massive investment in its journalism by using it to build substitutive products without permission or payment ...*”.

⁴ Si veda il punto I.6. del documento, per cui “... *Using the valuable intellectual property of others in these ways without paying for it has been extremely lucrative for Defendants. Microsoft's deployment of Times-trained LLMs throughout its product line helped boost its market capitalization by a trillion dollars in the past year alone. And OpenAI's release of ChatGPT has driven its valuation to as high as \$90 billion ...*”.

⁵ Si veda il punto IV.32. del documento.

⁶ Si veda il punto IV.38. del documento.

ed esposizione ai sensi della legge sul diritto d'autore, e di aver registrato il *copyright* della sua edizione stampata quotidianamente per oltre 100 anni, mantenendo un *paywall* ed implementando termini di servizio che fissino limiti alla copia e all'uso dei contenuti. Da qui, per utilizzare i contenuti del NYT a fini commerciali, la parte interessata dovrebbe dunque prima contattare la casa editrice per un accordo di licenza. Il NYT richiede a terzi di ottenere il consenso prima di utilizzare i suoi contenuti e marchi per scopi commerciali, e da decenni concede in licenza i propri contenuti in base ad accordi di licenza negoziati, i quali contribuiscono a garantire che il NYT controlli come, dove e per quanto tempo appaiono i suoi contenuti ed il suo marchio, e che riceva un equo compenso per l'utilizzo da parte di terzi⁷.

Tutto quanto sopra, appare in pieno contrasto con una serie di allegazioni, nelle pagine seguenti, in cui i denunciati, attraverso l'utilizzo della piattaforma ChatGPT, pongono una serie di domande al *chatbot*, il quale in risposta fornisce degli scritti identici ad articoli pubblicati nel corso degli anni dal giornale stesso. La Società afferma, di conseguenza, che *“la riproduzione e l'esposizione non autorizzata dei contenuti Times da parte dei convenuti è intenzionale”*. Ancora, si osserva che *“i convenuti erano intimamente coinvolti nell'addestramento, nella messa a punto e nel collaudo dei modelli GPT. Gli imputati sapevano o avrebbero dovuto sapere che queste azioni comportavano la copia non autorizzata di opere del Times su scala massiccia durante l'addestramento, che si risolvevano nella*

*codifica non autorizzata di un numero enorme di tali opere nei modelli stessi e avrebbero inevitabilmente portato alla visualizzazione non autorizzata di tali opere che i modelli avevano memorizzato o avrebbero presentato agli utenti sotto forma di risultati di ricerca sintetici”*⁸.

Infine, viene messo in luce il danno subito dalla Società, posto che *“i convenuti hanno utilizzato quasi un secolo di contenuti protetti da copyright, per i quali non hanno pagato al Times un equo compenso”*⁹.

Le reazioni di Open AI

Non sono mancate comunque le prime reazioni “a caldo” da parte di *Open AI*. Fonti giornalistiche¹⁰ rivelano che la Società contesterebbe *in toto* argomentazioni come quelle avanzate da NYT.

Il fulcro della difesa prospettata da *Open AI* si baserebbe sulla circostanza per cui sarebbe “impossibile” creare servizi come ChatGPT se gli fosse impedito di fare affidamento su tutte le opere protette dal diritto d'autore. Più specificamente, si ritiene che dal momento in cui il *copyright* copre virtualmente ogni tipo di espressione umana, inclusi post di blog, fotografie, interventi nei forum, frammenti di codice software e documenti governativi, semplicemente non si potrebbero addestrare i modelli di intelligenza artificiale di punta senza

⁷ Si vedano i punti IV.49. e IV.50. del documento.

⁸ Si veda il punto IV.124. del documento.

⁹ Si veda il punto IV.156. del documento.

Più approfonditamente, viene affermato che *“... The Times strictly limits the content it makes accessible for free and prohibits the use of its material (whether free or paid for) for commercial uses absent a specific authorization. Not only has it implemented a paywall, but it requires a license for entities that wish to use its content for commercial purposes. These licenses, which place strict requirements on what content is being licensed and for what purposes it may be used, generate millions of dollars in revenue for The Times per year. Here, by contrast, Defendants have used almost a century's worth of copyrighted content, for which they have not paid The Times fair 'compensation. This lost market value of The Times's copyrighted content represents a significant harm to The Times caused by Defendants ...”*.

¹⁰ Per ulteriori informazioni, si veda il seguente [LINK](#).

utilizzare materiali protetti dal diritto d'autore¹¹.

Da ultimo, la Società avrebbe affermato che *“la legge sul copyright non vieta l'addestramento”*¹², mettendo a fuoco il vero dilemma legale che emerge dalla controversia: l'addestramento degli algoritmi di IA tramite opere coperte dal diritto d'autore, costituisce di *per sé* una violazione di tale diritto?

Conclusioni

In conclusione, si rileva che l'azione intentata del NYT potrebbe avere effetti dirompenti sullo sviluppo (o non sviluppo) dell'intelligenza artificiale. A riguardo, non sfuggono gli opposti interessi delle parti in causa. La nota casa editrice statunitense si fa portavoce dell'esigenza di porre delle limitazioni all'intelligenza artificiale, e al relativo riconoscimento della facoltà di utilizzare ai propri fini opere protette dal diritto d'autore, per non vedere frustrato alle basi il suo *business model*. D'altra parte, sono

altresi apprezzabili le ragioni di *Open AI* che non vorrebbe essere “frenata” nello sviluppo della sua tecnologia e si dichiara dunque legittimata ad allenare i suoi algoritmi utilizzando tutti i materiali accessibili sul *web*. Lo scontro tra questi due colossi si pone in uno scenario globale in cui i Paesi sono chiamati a prendere in tempi brevi una posizione, spesso – come testimoniato da una recente sentenza in Cina¹³ – non uniforme.

Si tratta chiaramente di un importante crocevia etico e giuridico, tale per cui occorrerà, nel futuro prossimo, rimettere in discussione una serie di concetti storicamente legati alla capacità umana di creare, come l'originalità, poiché le creazioni dell'IA sono spesso il risultato di analisi e di elaborazioni di grandi quantità di dati che seguono algoritmi e tecniche di apprendimento automatico; e come la creatività, sino ad ora intesa come forma di espressione personale ed intenzionale.

¹¹ La posizione assunta da *Open AI* sarebbe la seguente: *“... Because copyright today covers virtually every sort of human expression - including blog posts, photographs, forum posts, scraps of software code, and government documents - it would be impossible to train today's leading AI models without using copyrighted materials ...”*.

¹² La Stampa sopra riportata afferma in particolare che *“... OpenAI said it complies with all copyright laws when training its models ...”*, e che *“... we believe that legally copyright law does not forbid training ...”*.


¹³ Il documento in analisi si inserisce in un contesto particolarmente sensibile sulla scena internazionale: il 27 novembre 2023, in Cina, il Tribunale di Internet di Pechino (*Beijing Internet Court*) ha ufficialmente ricosciuto il diritto d'autore di un'immagine generata da intelligenza artificiale (per consultare il documento ufficiale, si veda il seguente [LINK](#)). La disputa legale vedeva come attore il Sig. Li, il quale aveva utilizzato il software *Stable Diffusion* per creare le immagini su cui rivendicava la sua protezione. Il convenuto, un *blogger*, aveva impiegato l'immagine generata dall'IA di Li per illustrare un articolo online, scatenando così la controversia. A riguardo, il Tribunale di Pechino ha notato che Li ha compiuto *“un certo grado di investimento intellettuale”* nella scelta dei testi di richiesta, noti come *prompt*, così come nella configurazione dei parametri e nella progettazione della presentazione. Secondo il Tribunale, la “originalità” dell'immagine risiede quindi nelle selezioni ripetute con cui Li ha regolato i parametri per ottenere un'immagine che riflettesse la sua *“scelta estetica e il suo giudizio personalizzato”*, riconoscendo dunque che, per essere tutelata dal diritto d'autore, un'opera d'arte generata dall'IA deve avere una quantità sostanziale di lavoro intellettuale fornito dall'autore.




Jacopo Piemonte

ASSOCIATE

 j.piemonte@dejalex.com

 +39 02 72554.1

 Via San Paolo 7
20121 – Milano


 +32 (0)26455670


 Chaussée de La Hulpe 187
1170 – Bruxelles



Giulia Beneduci

ASSOCIATE

 g.beneduci@dejalex.com

 +39 02 72554.1


 Via San Paolo 7
20121 – Milano



Federico Aluigi

ASSOCIATE

 f.aluigi@dejalex.com

 +32 (0)26455670

 Chaussée de La Hulpe 187
1170 – Bruxelles

MILANO

Via San Paolo, 7 · 20121 Milano, Italia
T. +39 02 72554.1 · F. +39 02 72554.400
milan@dejalex.com

ROMA

Via Vincenzo Bellini, 24 · 00198 Roma, Italia
T. +39 06 809154.1 · F. +39 06 809154.44
rome@dejalex.com

BRUXELLES

Chaussée de La Hulpe 187 · 1170 Bruxelles, Belgique
T. +32 (0)26455670 · F. +32 (0)27420138
brussels@dejalex.com

MOSCOW

Potapovsky Lane, 5, build. 2, 4th floor, office 401/12/9 · 101000, Moscow, Russia
T. +7 495 792 54 92 · F. +7 495 792 54 93
moscow@dejalex.com