

Diritti d'autore e intelligenza artificiale

Tra politica, gestione diritti e opportunità

Piero Attanasio

**TORROSSA
RETREAT**

Fiesole – 18 aprile 2024

Sommario

L'AI pre-Chat-GPT: norme sul controllo dei dati

AI e diritto d'autore

I diritti sui contenuti usati per addestrare l'AI

Il *TDM Reservation Protocol* del W3C

Aprire le scatole nere dell'IA

I contratti con per l'uso dell'AI

I diritti sugli output dell'AI

Le «opere derivate» grazie all'AI

Qualche conclusione di ordine generale

L'AI prima di Chat-GPT

Premessa

Fenomeni sempre più sovra/internazionali richiedono **regole almeno europee**

L'Europa su queste tematiche sta legiferando tramite **Regolamenti**

Immediatamente in vigore in tutti gli Stati membri

Non richiedono il «recepimento» e quindi non consentono adeguamenti alle realtà locali

Anche gli atti esecutivi di secondo livello sono a livello europeo

Su AI e © sono rilevanti due regolamenti e una Direttiva (con relativo decreto di recepimento)

Il DMA = Digital Market Act (2022),

l'AI-Act = Artificial Act (2024)

La Direttiva CDSM = Copyright in the Digital Single Market (2019)

E il Decreto legislativo di recepimento (2021)



Il problema:

“Who owns the robots rules the world” (Freeman, 2015)

“The quintessential task of many digital platforms is making predictions of various sorts (...) Data is the oil that powers these predictions” (Calvano and Polo, 2020).

“Platforms can use this information asymmetry to facilitate interaction and increase welfare for users. These data externalities attract users to the platform” (Martens, 2020)

“The collection and use of big user data enables [platforms] to continuously improve the quality of their offerings” (Fast et al., 2021), which “may result in monopolistic market power of platforms which they can use for their own benefit, at the expense of users” (Martens, 2020)

La ricerca di soluzioni:

Necessità di norme diverse ma coordinate: diritto del lavoro, fiscale, privacy, concorrenza (Tirole, 2019). E sostegno all'innovazione delle PMI, e tutela dei diritti d'autore...

Il **DMA** (Digital Market Act) è un tentativo di affrontare con una politica della concorrenza ex ante (Khan, 2017)

Obbliga i gate keepers a condividere i dati con i partner commerciali che li generano



A che punto siamo

1. La Commissione europea individua i «gate-keepers» (soggetti con posizioni dominanti in mercati digitali europei).
Amazon è uno di quelli individuati
2. Ogni «gate-keeper» deve presentare i propri «impegni» per rispettare gli obblighi, tra cui quello sulla condivisione dei dati
Amazon sostiene che non deve condividere i dati con gli editori
3. Gli editori, tramite FEP, stanno contestando questa presa di posizione
4. La Commissione europea dovrà decidere

***L'AI generativa
(Large Language Model)
(General purpose AI)***

L'AI generativa e il diritto d'autore. Le tre domande fondamentali

1. Se ho l'accesso a un contenuto devo chiedere l'autorizzazione per darlo in pasto a un algoritmo di intelligenza artificiale?
2. L'output di un procedimento che utilizza l'intelligenza artificiale generativa è un'opera protetta dal diritto d'autore e, se lo è, chi ne è il titolare?
3. L'output di un procedimento che utilizza l'intelligenza artificiale generativa è un'opera derivata delle opere che hanno contribuito a crearla?

***Prima domanda:
È legittimo l'uso di opere
protette in applicazioni AI?***

Posso utilizzare i contenuti a cui ho accesso per addestrare le mie AI?



Esiste il diritto di *text and data mining (TDM)*?

Qualcuno sostiene che «*the right to read includes the right to mine*»

La Direttiva CDSM del 2019 sembra risolvere il problema:

Stabilisce due eccezioni: assoluta per gli enti di ricerca per applicazioni no-profit, con la possibilità di riservare il diritto in caso applicazioni profit ⁽¹⁾.

Eccezione e riserva confermano l'esistenza del diritto

Ciò significa che, *per le applicazioni commerciali*, l'uso di opere protette per il *training* dipende anche dal tempo:

Fino a giugno 2021 era illegittimo,

Dal giugno 2021 si può fare in regime di eccezione, fino a riserva dei diritti

N.B.: In entrambi i casi, il TDM è possibile solo se chi lo fa ha un «*accesso legittimo*» ai testi

(1) Rispettivamente artt. 3 e 4 della Direttiva CDSM e 70-ter e 70-quater della Legge italiana sul diritto d'autore (LdA)

Come si può riservare i diritti



Se i contenuti sono concessi in licenza, in primis la riserva può essere espressa contrattualmente

Per i contenuti a libero accesso: **Rights Reservation Protocol W3C** e l'importanza di adottarlo

Nato per essere incorporato nelle pagine web è oggi anche per gli EPUB

Allo studio la soluzione per i PDF

Altre soluzioni:

Riserve da parte di società di gestione collettiva (prime esperienze in Francia e Germania)

La SIAE in Italia ha modificato il proprio regolamento (art. 3, ultimo comma) per specificare che non autorizza i propri clienti al TDM sulle opere gli utilizzatori cui accedono

L'espressione «tutti i diritti riservati», comprende i diritti TDM?

Una licenza cc-nc contiene una riserva per il TDM commerciale?

Ma come attuare le nuove norme se non si sa nulla di come sono addestrati gli algoritmi?

Rilevanza nel settore accademico

N.B.: riservare i diritti ≠ vietare

Significa riservarsi i diritti di autorizzare caso per caso, a titolo gratuito o oneroso

Perché fare l'opt-out se si vende solo alle biblioteche universitarie?

Le università, per ragioni di ricerca, possono usare i contenuti senza autorizzazione, anche in caso di opt-out (Art. 3 Direttiva, Art. 70-ter LdA)

Tuttavia

In caso di progetti di ricerca università / imprese, l'opt-out è valido

Può essere fatto per via contrattuale

Le novità contenute nell'AI-Act

L'AI-Act non ha alcuna norma nuova sul diritto d'autore!

Richiama gli artt. 3 e 4 della Direttiva CDSM

Chiede alle piattaforme AI di darsi regole interne per controllare il rispetto dei diritti d'autore

La vera novità è l'articolo sulla **trasparenza**

Ogni piattaforma dovrà fornire una «**sintesi dettagliata**» delle fonti utilizzate

Il livello di dettaglio è definito in modo funzionale: quanto serve ai titolari dei diritti per capire se sono state utilizzate proprie opere

Pur non essendo una nuova norma sul diritto d'autore, è essenziale per l'*enforcement*

A che punto siamo

L'AI-Act ha istituito una nuova Agenzia europea (**AI-Agency**) con diversi compiti per l'attuazione del Regolamento

In tema di trasparenza dovrà fornire un template e linee guida per la piattaforme AI sul «*detailed summary*»



Per cosa sono usati i contenuti degli editori

1. Per il *training* iniziale

Quando le macchine apprendono a parlare

2. Per il *fine tuning*

Quando le macchine imparano a non dire sciocchezze (o a dirne di meno)

3. Per la RAG - *Retrieval-Augmented Generation*

Quando le macchine si aggiornano

4. In futuro per altro che oggi non immaginiamo

Spunti interessanti per capire le diverse fasi in Cristianini, 2024

I contenuti editoriali sono più preziosi (oggi) nelle fasi 2 e 3

Le richieste degli editori sul *detailed summary*

Dovrà contenere informazioni sui dataset utilizzati e sui modi per accedervi così da controllare effettivamente cosa è stato usato.

Altre esigenze

Origine della richiesta	Contenuti del <i>detailed summary</i>
In Europa il regime giuridico è diverso prima e dopo il 2021 e prima e dopo l'opt-out	When: deve essere noto quando il TDM è stato effettuato
In Europa l'eccezione vale solo se l'accesso ai contenuti è legale	Where: dove è stato effettuato il TDM (se da fonti legittime o pirata)
Il valore dipende dalla fase in cui un'opera viene utilizzata	Why: per cosa sono state usate le opere (training, fine tuning, RAG...)
Le opere possono essere state acquisite in modi diversi	How: come sono state acquisite le opere (licenza, sul web ad accesso aperto, ecc.)

I contenuti immessi in applicazioni AI dai titolari dei diritti

Il problema

Quando uso mie opere in applicazioni AI (es. per generare metadati, parole chiave, abstract...), i miei contenuti sono usati anche per training, fine-tuning, RAG...?

In genere, le applicazioni gratuite richiedono:

- Una dichiarazione che attesti che si ha l'autorizzazione di usare le opere (e una manleva)

- La cessione del diritto di uso da parte della piattaforma AI

Le applicazioni corporate a pagamento hanno spesso condizioni più eque per gli utilizzatori

AIE sta organizzando una serie di seminari sul tema (www.aie.it/Cosafacciamo/Corsieformazione.aspx)

Prossimo corso 11-13 giugno

***Seconda domanda:
Quali diritti sull'output
dell'AI***

Criteria per capire cosa è protetto e cosa non lo è

Il principio generale del diritto d'autore

«**Sono protette** ai sensi di questa legge **le opere dell'ingegno di carattere creativo**»

Di conseguenza

Un'opera ottenuta usando strumenti di AI è protetta se vi è una creatività umana

Qual è il confine?

La legge lo lascia indefinito, ma come era indefinito in passato

Per decenni abbiamo risposto a domande del tipo:

Quando una data fotografia si può definire «opera fotografica»?

(la distinzione è ripresa dalla Direttiva CDSM per le fotografie di opere in pubblico dominio)

Quando un post su un social merita la protezione?

Ci sono, e ci saranno, grandi discussioni tra i giuristi, ma il principio generale è molto semplice!

Alcuni esempi

Il copyright office USA ha accordato la protezione al testo di un fumetto creato con l'AI ma non ai disegni, in quanto l'autore si era limitato a usare il testo come prompt e l'AI aveva fatto il resto.

Una traduzione o un abstract fatti interamente con l'AI, o con minimi interventi redazionali, non sono opere protette (ma resta il diritto sull'opera tradotta)

Un testo creato con ripetuti prompt, selezione dei risultati, interventi umani significativi, ecc. è certamente un'opera protetta (es. Tanica, 2022)

***Terza domanda:
L'output dell'AI è un'opera
derivata***

Cos'è un'opera derivata?

La risposta è nell'art. 4 LdA:

**Senza pregiudizio dei diritti esistenti sull'opera originaria, sono altresì protette
le elaborazioni di carattere creativo dell'opera stessa,**

quali le **traduzioni** in altra lingua,

le trasformazioni da una in altra forma letteraria od artistica, le modificazioni ed aggiunte che costituiscono un rifacimento sostanziale dell'opera originaria, gli adattamenti, le riduzioni,

i compendi,

le variazioni non costituenti opera originale.

Derivata da cosa?

Poiché la macchina ha imparato a parlare e si è costruita una sua «visione del mondo» (Cristianini, 2024) a partire da altre opere, si può ritenere che i suoi output sono derivati da tutto quanto ha ingerito?

La risposta più comune è no: il procedimento non è dissimile a quello propedeutico a ogni creazione umana

Ma se riporta quanto appreso dalla RAG, è un'opera derivata?

Valgono le regole stabilite per le **citazioni**?

Ma allora di cosa parliamo?

Se un utilizzatore mette nel «contesto» di un'applicazione un'opera e chiede di fare una traduzione (o una sintesi, o una versione divulgativa di un testo scientifico, ecc.)

1. Siamo in presenza di una «**elaborazione**» (primo criterio dell'art. 4)

Se vi è un significativo intervento «creativo» umano che rende l'opera «originale»

2a. L'elaborazione ha «**carattere creativo**» (secondo criterio dell'art. 4)

Se l'opera viene pubblicata così come prodotta dall'IA

2b. L'elaborazione non può essere considerata «opera derivata» ai sensi della legge, e quindi non è protetta. Resta invariata la protezione dell'opera originale

Esempio tristemente noto in ambito universitario

Uno studente fa fare una sintesi di un manuale a Chat-GPT e vende il risultato su una piattaforma di «riassunti»

Chi è responsabile in caso di produzione di elaborazioni non autorizzate?

Nel caso in cui **l'originale è stato usato come «contesto»** nel prompt:

È certamente responsabile l'utilizzatore

Si discute se lo sia anche la piattaforma AI

È la ragione per cui le piattaforme chiedono dichiarazioni e manleve (che hanno il valore delle dichiarazioni che gli editori chiedono agli autori)

Un dubbio: se la macchina è intelligente non dovrebbe capire quando un'operazione è illegittima?

Per meglio dire: non dovrebbe essere addestrato a capirlo?

Chi è responsabile in caso di produzione di elaborazioni non autorizzate?

Nel caso in cui l'originale era tra le fonti del LLM ed «emerge» sulla base di un prompt: Es.

Prompt: «Scrivi il testo di una canzone in cui una 16enne degli anni '50 dice all'amato di essere troppo giovane per un storia d'amore»

Risposta: il testo di *Non ho l'età* di Gigliola Cinquetti con poche varianti

Esiste una causa su queste basi in USA, con risultati reali analoghi

Evidenze di questo genere, hanno due conseguenze:

Certamente dimostrano che quella canzone è stata utilizzata in fase di «apprendimento»
(violazione in fase di input)

Cause in USA ritengono che c'è una violazione anche in fase di output, perché – sia pure forzata a farlo – la macchina «riproduce» un'opera protetta



*Qualche riflessione teorica
finale di politica*

Il ruolo del diritto d'autore nell'economia dei dati

La tutela del diritto d'autore è sempre più uno strumento per garantire condizioni equi sui mercati culturali

Un apparente paradosso: stabilendo un monopolio su ogni singola opera, tempera lo strapotere di monopoli basati sul controllo delle piattaforme distributive e dei dati

Evidente nella Direttiva CDSM nell'articolo 17, pensato per riequilibrare il rapporto tra produttori musicali e piattaforme di distribuzione (cd *value gap*);

e nell'articolo 15, nato per rafforzare il potere di mercato degli editori di giornali nei confronti di GoogleNews et similia

Oggi si ripete in ambito di intelligenza artificiale: difendere i diritti d'autore è (può essere) uno strumento per distribuire più equamente il potere di mercato

Riflessi politici

Se si cerca un fenomeno inedito nella storia dei mercati culturali e non è **gratis vs. a pagamento**, che è un tema che si è posto già per la TV generalista (cfr. tra molti Picard, 1989)

C'è una storia più complessa

Tradizionalmente, nei mercati culturali il potere di mercato era basato sul controllo dei contenuti

Solidarietà di filiera: dall'autore al libraio sono tutti alleati contro la pirateria libraria; idem dallo sceneggiatore al *broadcaster*, ecc.

Che si vendano accessi o pubblicità, l'interesse è tutelare i contenuti «appetibili» per i diversi pubblici

I social media cambiano il quadro creando un conflitto tra:

chi controlla i contenuti (autori, editori, distributore...) e

chi controlla le relazioni «sociali» in cui si scambiano contenuti

La solidarietà di filiera si è spezzata. Gli interessi di big tech e operatori culturali divergono

Conclusioni

Nella discussione sulla Direttiva CDSM del 2019 era frequente la rappresentazione dello scontro tra «popolo del web» e «industrie culturali»

Hanno prevalso le seconde perché siamo riusciti a far capire quanto fuorviante fosse quella rappresentazione

Nella discussione sull'AI prevale la rappresentazione di una battaglia tra i pochi giganti dell'IA e i molti operatori culturali

Lo spiegava già bene, con ironia, Jaron Lanier (2019) nel commentare la Direttiva CDSM

Consigli di lettura

Articoli e libri citati

- Calvano E., Polo M. (2020), Market Power, Competition and Innovation in Digital Markets: A Survey, *Information Economics and Policy*, Vol. 54, 100853.
- Cristianini N. (2024), *Machina Sapiens*, Bologna, Il Mulino.
- Fast V., Schnurr D. and Wohlfarth M. (2021), Regulation of Data-driven Market Power in the Digital Economy: Business Value Creation and Competitive Advantages from Big Data (January 31, 2021).
- Freeman R.B. (2014), Who Owns the Robots Rules the World, *IZA World of Labor*, May 2014.
- Kahn L.M. (2017), Amazon's Antitrust Paradox, *The Yale Law Journal*, 126.
- Martens B. (2020), An Economic Perspective on Data and Platform Market Power, *JRC Digital Economy Working Paper*, 2020-09.
- Lanier J. (2019), A Love Letter to the Pretend Pirates of Europe from the Real Pirates of Silicon Valley, *Billboard Pro*. Marz 2019
- Picard R.G. (1989), *Media Economics: Concepts and Issues*, Newbury Park, Calif., Sage Publication
- Picard R.G. (2005), *Economia e finanza dei media*, Milano, Guerini e Associati.
- Tanica R. (2022), *Non siamo mai stati sulla Terra*, Milano, Il Saggiatore.
- Tirole J. (2019), "Regulating the Disrupters", *Livemint*, Jan 2019.

TORROSSA
RETREAT

Grazie dell'attenzione

Piero Attanasio
piero.attanasio@aie.it

AIE Associazione
Italiana
Editori 