



Il diritto alla ricerca

Nuove frontiere e profili evolutivi del diritto d'autore

19 dicembre 2023, Biblioteca Vallicelliana - Roma

Le eccezioni di TDM per finalità di ricerca scientifica e per uso commerciale nei sistemi di machine learning generativa

Avv. Deborah De Angelis

Estrazione di testo e dati (TDM)

Il TDM è ampiamente utilizzato in vari settori, tra cui economia, finanza, scienza, statistica, marketing, medicina, industrie creative e ricerca scientifica. Consente l'estrazione automatizzata di informazioni preziose dai database, consentendo attività come le previsioni del mercato azionario, l'analisi demografica, il clustering dei clienti e altro ancora.

Il Text and Data Mining (TDM) è un processo computazionale che prevede l'analisi e la rielaborazione di grandi quantità di dati grezzi e strutturati. L'obiettivo è quello di estrarre informazioni, identificare modelli e scoprire nuove tendenze e approfondimenti da diverse fonti di informazioni e dati.

Ogni giorno condividiamo, alle volte anche inconsapevolmente, una grande quantità di informazioni e dati che generiamo attraverso l'utilizzo di varie piattaforme, app, smartphone e altri device. Ciò consente un'analisi predittiva delle preferenze e del comportamento degli utenti.

Il dibattito internazionale - Organizzazione mondiale per la P.I. (SCCR)



American University Washington College of Law

Digital Commons @ American University Washington College of Law

Joint PIJIP/TLS Research Paper Series

10-2022

Research Exceptions in Comparative Copyright

Sean Flynn

American University Washington College of Law, sflynn@wcl.american.edu

Luca Schirru

American University Washington College of Law, 5211.luca@gmail.com

Michael Palmedo

American University Washington College of Law, mpalmedo@american.edu

Andrés Izquierdo

American University Washington College of Law, andres@wcl.american.edu

Le attività di TDM rappresentano una componente della rivoluzione digitale nel campo della ricerca scientifica e per fini diversi.

Nel mondo, ciascuna legge sul diritto d'autore prevede almeno un'eccezione per promuovere gli usi a fini di ricerca delle opere protette dal diritto d'autore, ma tali eccezioni **variano notevolmente** da un Paese all'altro.

I moderni modelli di ricerca scientifica (IA), che utilizzano metodi computazionali per analizzare le informazioni nei libri, articoli, database ed altre opere e materiali protetti necessitano di **maggiore flessibilità** negli utilizzi consentiti dalle varie forme di eccezione già previste nelle legislazioni in materia di diritto d'autore.

Eccezioni per TDM per scopi di ricerca scientifica

Figure 3.

TDM Exceptions in Comparative Copyright PIJIP

Country	Commercial	Lawful Access	Protection from contracts	Uses	Users	Works	Typology
Japan	Y	N		Use	Open	Open	Green
Singapore	Y	Y	Y	Reproduction, communication	Open	Open	Green
Germany	N			Reproduction, communication, storage	Open	Open	Green
Estonia	N	N		“processing”	Open	Open	Green
UK	N	Y	Y	Reproduction	Open	Open	Blue
Switzerland				Reproduction	Open	Open	Blue
EU DSM Art 3	*	N	Y	Reproduction, storage	Cultural institutions	Open	Purple
France	N	Y		Reproduction, communication (decree)	Open	Scientific writings	Orange
Ecuador	*			TDM “acts carried out by their users”	Libraries, Archives safe harbor	Open	Red



Gestione/protezione dei dati nella fase di **input** per addestramento dell'IA

Quali tipologie di dati?

Informazioni, dati, banche dati, opere e materiali protetti dal diritto d'autore e diritti connessi, contenuti in pubblico dominio o rilasciati con licenza Creative Commons.

Violazione di quali diritti nella fase di training?

Riproduzione, comunicazione al pubblico, modificazione, elaborazione.

Due sistemi a confronto: civil law e common law.

Il mondo guarda all'UE. Definizione

Art. 2, 2 CDSM

Qualsiasi tecnica di analisi automatizzata di testi e dati in formato digitale avente lo scopo di generare informazioni inclusi, a titolo non esaustivo, modelli, tendenze e correlazioni.

Art. 70 ter, comma 2, LdA

Qualsiasi tecnica automatizzata volta ad analizzare **grandi quantità** di testi, **suoni**, **immagini**, dati o **metadati*** in formato digitale per generare informazioni, compresi schemi, tendenze e correlazioni (art. 70 ter, comma 2, LdA).

**con cui il legislatore italiano ha ampliato l'ambito applicativo dell'eccezione anche alle informazioni sui dati organizzate in modo strutturato o semistrutturato.*



Artt. 3 e 4 direttiva 2019/790/UE (DCSM)

FINALITA'

ART. 3 → ricerca scientifica

ART. 4 → per scopi diversi dalla ricerca scientifica



Artt. 3 e 4 direttiva 2019/790/UE (DCSM)

BENEFICIARI

ART. 3 → organismi di ricerca e istituti culturali

ART. 4 → chiunque



Artt. 3 e 4 direttiva 2019/790/UE (DCSM)

OGGETTO

ART. 3 → le opere e i materiali protetti

ART. 4 → le opere e i materiali protetti per i quali i titolari dei diritti non abbiano espressamente riservato, in modo appropriato (machine-readable standard) l'utilizzo per TDM.

Recepimento italiano (D. Lgs. 8 novembre 2021, n. 177)

TDM per scopi di ricerca scientifica

Art. 70 ter LdA

consente le **riproduzioni** effettuate da enti di ricerca e istituti culturali a fini di ricerca scientifica, per **l'estrazione** di testi e dati da opere o altri materiali **disponibili su reti o banche dati** a cui **hanno lecitamente accesso**, nonché la **comunicazione al pubblico** dei risultati della ricerca se espressi in **nuove opere originali**.



Protezione contrattuale

Come stabilito dall'art. 7 (1) della Direttiva CDMS, le **clausole contrattuali in conflitto** con l'eccezione TDM per scopi scientifici sono **nulle**.



Eccezione per l'estrazione di testi e dati per scopi diversi dalla ricerca scientifica (art. 70 quater LdA)

Sono consentite **le riproduzioni e le estrazioni** da opere o altri materiali contenuti in reti o in banche dati **a cui si ha accesso legittimo** ai fini dell'estrazione di testi e di dati.

L'eccezione è subordinata alla possibilità di esercitare la **facoltà di opt-out** (riserva delle loro opere o altri materiali protetti dai diritti connessi o banche dati).

Opt out - Come?

Al momento non vi è un protocollo o uno standard generalmente riconosciuto per il *machine readable expression* della facoltà di opt-out. Esistono un certo numero di approcci: protocolli sviluppati dagli editori, servizi sviluppati da artisti e alcune specifiche proposte dalle aziende di IA.

L'UE dovrebbe intervenire per fornire una lista di standard da adottare.

Il modello europeo di opt-out potrebbe rappresentare lo standard anche in USA, dove si escluderebbe la violazione e quindi si rientrerebbe nel *fair-use* in presenza dei requisiti stabiliti dalla direttiva EU, ossia l'accesso legale e l'opt-out.

Convergenza tra il sistema di civil law e common law.

Posso conservare e memorizzare i contenuti oggetti di TDM?



Articolo 3 Dir./Articolo 70 ter LdA

Memorizzazione e conservazione

Le copie delle opere o di altri materiali sono **memorizzate** con un **adeguato livello di sicurezza** e possono essere **conservate** e **utilizzate** solo per scopi di ricerca scientifica, compresa la verifica dei risultati della ricerca (art. 70 ter, comma 6).

Articolo 4 Dir./Articolo 70 quater LdA

Conservazione e archiviazione

Le riproduzioni e le estrazioni possono essere conservate **solo per il tempo necessario** all'attività di text and data mining (art. 4 della Direttiva CDSM, art. 70 quater, comma 2, LdA).



Misure di protezione tecnologica

I titolari dei diritti sono autorizzati ad applicare misure per garantire la sicurezza e l'integrità delle reti e delle banche dati in cui sono ospitate le opere o altri materiali (art. 70 ter, comma 7, LdA, art. 70 quater, comma 3).

Tali misure non devono andare oltre quanto necessario e possono essere definite anche sulla base di accordi tra associazioni di titolari di diritti, istituti di patrimonio culturale e organizzazioni di ricerca (art. 70 ter, comma 8, LdA).

Finora tali negoziati non hanno avuto luogo.

Esercizio dell'eccezione Vs. Misure di protezione



Il legislatore italiano **non ha previsto espressamente alcuna procedura rapida e specifica** per i casi in cui le misure di protezione tecnologica impediscano al beneficiario di compiere gli atti consentiti da entrambe le eccezioni TDM.

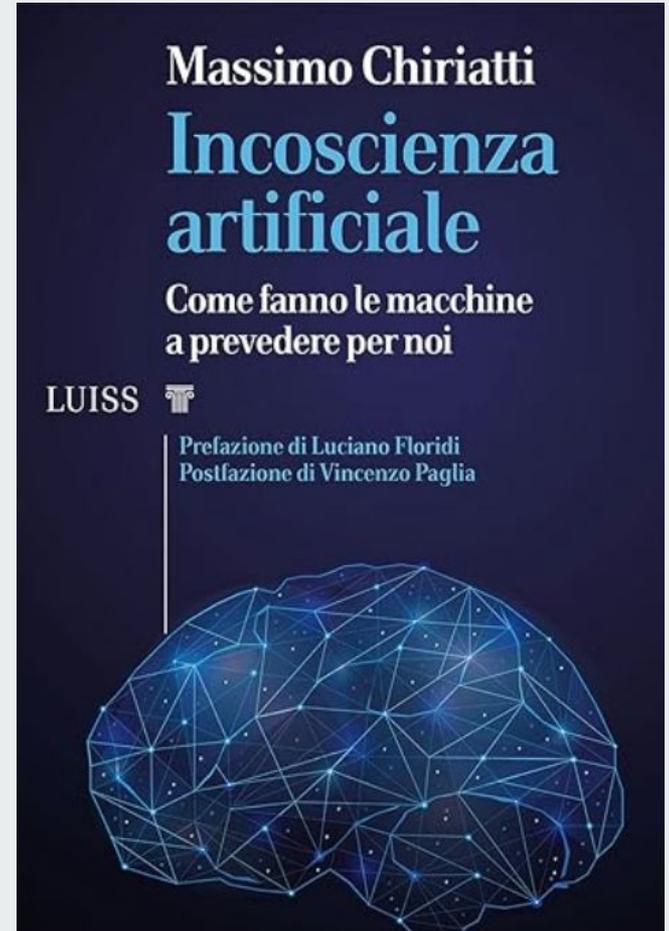
Ma l'art. 70 sexies LdA stabilisce che le gli **istituti di istruzione** (art. 70-bis, comma 1), gli **istituti di tutela del patrimonio culturale** (art. 70-ter, comma 3) e **gli organismi di ricerca** (art. 70-ter, comma 4), che hanno acquisito il **possesso legittimo di esemplari dell'opera** o del materiale protetto, oppure **vi hanno avuto accesso legittimo**, hanno diritto di estrarne una copia, **con i limiti e le finalità previste dalla legge**, purché l'estrazione di tale copia **non sia in contrasto con lo sfruttamento normale** dell'opera o degli altri materiali, e a condizione che non arrechi un **ingiustificato pregiudizio** ai titolari dei diritti (secondo il principio del *three-step-test*, richiamato dall'art. 7 (2) della Direttiva Copyright per la sua applicazione alle eccezioni incluse nel Titolo II).



Sistemi di machine learning generativa

Il ML è una branca dell'intelligenza artificiale che si occupa di creare sistemi che apprendono o migliorano le performance in base ai dati che utilizzano.

ML generativo è un tipo di intelligenza artificiale in grado di creare nuovi contenuti e idee, tra cui conversazioni, storie, immagini, video e musica. IA generativa risponde al comando (prompt) impartito dall'utente umano.



TDM & IA generativa



La definizione di TDM è sufficientemente ampia da comprendere molti processi di apprendimento automatico, come riferito dalla Commissione europea, nella propria risposta ad un'interrogazione parlamentare.

La Commissione ritiene che la creazione di opere d'arte da parte dell'IA non meriti un intervento legislativo specifico, rientrando tale disciplina nel quadro della direttiva Copyright, [Interrogazione parlamentare E-000479/2023(ASW)].

https://www.europarl.europa.eu/doceo/document/E-9-2023-000479-ASW_EN.html

AI ACT Notizie delle ultime ore...!

una serie di norme per i sistemi e i modelli di IA per scopi generali (GPAI), per affrontare le questioni sollevate dai modelli di IA generativa come GPT4 e Midjourney

Obblighi

di

trasparenza

1. preparare e rendere disponibile al pubblico un riassunto sufficientemente dettagliato dei contenuti utilizzati per l'addestramento del modello o del sistema e informazioni sulla politica interna del fornitore per la gestione degli aspetti legati al copyright;
2. dovranno dimostrare di aver adottato misure adeguate per garantire che la formazione del modello o del sistema avvenga nel rispetto del diritto dell'Unione sul diritto d'autore e sui diritti connessi, in particolare dell'art. 4.3 del DSMD (rafforzamento dell'opt-out per esprimere la riserva dei diritti di riproduzione e estrazione per le finalità di TDM che per i contenuti online deve essere fatto attraverso un sistema machine readable).

Interazione tra l'AI Act e le pratiche di sviluppo del software libero, gratuito e open source

Notevole aumento di modelli fondamentali di intelligenza artificiale rilasciati con licenze open source, come BLOOM di Big Sciences e Stability.ai.

Il regolamento non si applica ai sistemi di IA forniti con licenze libere e open source, a meno che non si tratti di sistemi di IA ad alto rischio.

I modelli di IA open source sono definiti come "modelli di IA resi accessibili al pubblico con una licenza libera e open source i cui parametri, compresi i pesi, le informazioni sull'architettura del modello e le informazioni sull'utilizzo del modello, sono resi disponibili al pubblico".

DE IURE CONDENDO: LINEE GUIDA EU PER OPT OUT



Nel complesso, le disposizioni sul diritto d'autore contenute nell'AI ACT rappresentano un passo nella giusta direzione, consolidando l'attuale approccio legislativo equilibrato adottato dall'UE nella Direttiva CDSM del 2019.

Inoltre, l'obbligo di pubblicare sintesi sufficientemente dettagliata dei contenuti utilizzati per l'addestramento di cui all'articolo 1(d) renderà più facile per i terzi comprendere le fonti dei dati di addestramento, compreso il soddisfacimento del criterio di accesso legittimo.

Necessità linee guida EU per OPT-OUT.

OPT-OUT ex art. 4 e strumenti Creative Commons

Le licenze CC, basandosi sulle legislazioni del diritto d'autore, non si applicano in presenza di limitazioni e eccezioni, o del fair use.

I titolari dei diritti possono scegliere di esercitare l'opt out "in modo appropriato, ad esempio, con mezzi leggibili dalla macchina nel caso di contenuti resi pubblicamente disponibili online".

I termini delle licenze CC e, in particolare, NoDerivatives (ND) e NonCommercial (NC), non costituiscono una riserva di diritti "in modo appropriato" per il TDM, ai sensi dell'articolo 4 e del considerando 18 (e di qualsiasi testo nazionale di recepimento).

La riserva ai sensi dell'articolo 4 può essere considerata compatibile con le licenze CC solo se opera all'interno dell'articolo 4 (rendendo l'eccezione non disponibile) ed è rilasciata con atto separato e distinto dalla licenza CC.

Attribuzione (diritti morali) e sistemi di machine learning



Altra questione aperta è la **tutelabilità del diritto morale** d'autore nei sistemi di ML generativa, considerato che tutte le licenze CC prevedono l'attribuzione della paternità (BY).

Nelle cause giudiziarie statunitensi, gli attori hanno presentato una richiesta di risarcimento per contestare **la mancata attribuzione ai titolari dei diritti d'autore delle opere coperte da copyright presenti nei dati di addestramento dell'IA generativa.**

In generale, il Copyright Act statunitense non tutela in modo significativo i diritti morali, tuttavia, oltre alla previsione contrattuale delle licenze CC, la sezione 1202 del Digital Millennium Copyright Act (DMCA) protegge i titolari del copyright contro la rimozione/modifica delle informazioni di gestione del copyright in relazione alle loro opere.



Grazie per la vostra attenzione!

info@ddastudiolegale.it

Questa presentazione è rilasciata con licenza [CC BY SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).