

9 luglio 2025

Diritto d'autore ed AI generativa tra riserve di protezione e strategie di licenza

1. Introduzione



Autori

Michele Bertani
michele.bertani@belex.com
Tel. +39 02 771131

Pasquale Tammaro
pasquale.tammaro@belex.com
Tel. +39 02 771131

La prima fase di sviluppo dell'AI generativa ha visto prevalere la strategia *move fast and break things*, che ha consentito gli straordinari progressi tecnologici sotto gli occhi di tutti.

Nella corsa all'innovazione alcuni temi sono rimasti sullo sfondo, tra i quali il rapporto con i titolari delle risorse protette da diritti di proprietà intellettuale – quelli d'autore su ogni genere di opera, i diritti *sui generis* per i *database*, i diritti connessi su audiovisivi, musica registrata, *performance* artistiche, fotografie non creative – utilizzate come *input* per il funzionamento dei sistemi di intelligenza artificiale. Numerosi segnali suggeriscono che questi temi stanno per guadagnare il centro della scena, imponendo agli attori in campo l'adozione di strategie più articolate sul piano della loro politica di gestione dei dati, propri o altrui.

È noto infatti che ogni fase dell'operatività dei vari sistemi di AI generativa – a partire dai cd. *foundation models* a vocazione *general purpose*, per arrivare ai modelli *fine tuned*, addestrati per fini specialistici – richiede il trattamento di una quantità assai elevata di dati. Ed è altrettanto indiscusso che la messa a punto di sistemi capaci di interagire con l'uomo usando un linguaggio naturale (cd. LLM o *Large Language Models*, come *ChatGPT*, *Gemini*, *Claude*, *Deepseek*, *LLaMA*, *Mistral*) oppure elaborando immagini (cd. *diffusion models*, tra i quali *DALL-E*, *Stable Diffusion*, *Midjourney*) si fonda sul trattamento di dati tipicamente protetti dai diritti d'autore o connessi, attraverso operazioni (*in primis*, la loro riproduzione, anche parziale o temporanea) che in linea di principio sono riservate ai titolari di questi diritti.

Sin qui queste attività sono state realizzate prevalentemente senza accordi tra *right holders* ed utilizzatori, sulla base di strategie che tendono da un lato a

limitare l'uso di risorse protette, dall'altro lato a sfruttare quanto più possibile gli spazi di libertà d'uso consentiti dai vari ordinamenti, tra i quali la cd. eccezione di *data mining* prevista dalla dir. UE 790/2019 e la *fair use clause* del *Copyright Act* statunitense.

L'ampiezza di questi spazi di libertà è già peraltro oggetto di contenziosi in varie giurisdizioni, nell'ambito di un conflitto mantenutosi per ora a bassa intensità, seppure con visibilità crescente nel dibattito politico generale ⁽¹⁾.

L'impiego massiccio di investimenti segnala d'altro canto il valore crescente – potenzialmente elevatissimo – delle tecnologie dell'AI generativa. E stimola in parallelo le pretese redistributive dei *right holders*, che già in parte – per ora soltanto all'estero – hanno seguito la via giudiziale, ma potenzialmente potrebbero trovare soddisfacimento sul piano negoziale.

Per i *right holders* che vogliano gestire adeguatamente l'una e l'altra strada è opportuno procedere all'analisi delle risorse coinvolte, l'adozione di misure adeguate per riservarne la protezione, l'elaborazione e l'attuazione di strategie di *licensing*, il monitoraggio delle violazioni e l'*enforcement* dei diritti violati.

Per gli sviluppatori e gli utilizzatori dei sistemi di AI generativa il percorso è simmetrico, in quanto richiede l'esame del fabbisogno e dell'uso dei dati altrui per soddisfare gli obblighi *ex lege* di trasparenza, la predisposizione di strumenti di rilevazione ed esame delle dichiarazioni di riserva, la negoziazione di licenze, la difesa contro l'eventuale *enforcement*.

2. *Training* dei modelli di AI e dichiarazioni di riserva

L'interazione automatizzata in linguaggio naturale richiede che il *Large Language Model* sia sottoposto ad una massiccia attività di *machine learning*, per fargli acquisire l'abilità di comprendere il significato delle richieste dell'utilizzatore e di soddisfarle con un livello appropriato di coerenza semantica e profondità informativa. Un affinamento ulteriore di queste abilità viene poi conseguito dai sistemi di AI sottoposti a *fine tuning* specialistico, ovvero a *reinforcement learning*.

I vari tipi di *machine learning* condotti da operatori automatizzati implicano il trattamento di enormi quantità di dati espressi in linguaggio naturale, ridotti in unità elementari (i cd. *token*) ed analizzati sul piano delle loro relazioni statistiche. Ancor prima, le fasi di raccolta dei dati, preparazione (tramite *cleaning*, *filtering*, *labelling*, etc.) per le attività di *training*, analisi e generazione dell'*output* richiedono forme di trattamento dei testi scritti in linguaggio naturale che rientrano in astratto nel contenuto dei diritti esclusivi d'autore o connessi.

⁽¹⁾ V. recentemente M. WOLF, *Creatives and the AI Industry are Clashing over Copyright*, nel *Financial Times* del 23.6.2025.

Esigenze analoghe sorgono per mettere a punto i *Diffusion Models*, che permettono l'elaborazione artificiale delle immagini, anche partendo da interrogazioni formulate in linguaggio naturale.

Per agevolare lo sviluppo dei sistemi di AI il legislatore europeo ha stabilito un'eccezione specifica a tali diritti, riguardante l'uso delle risorse protette per finalità di estrazione di testo e di dati (*text and data mining* o TDM). Questa eccezione è inderogabile qualora il TDM sia compiuto a scopo di ricerca scientifica. Quando invece il sistema di AI abbia finalità (di altro genere, tra le quali anzitutto quelle) commerciali, il TDM è lecito soltanto a condizione che il *right holder* abbia omesso di effettuare una **dichiarazione di riserva “in modo appropriato”**, che nel caso dei contenuti resi pubblicamente disponibili *online* ne implica la possibilità di “lettura automatizzata” [v. gli artt. 3-4 dir. UE 790/2019 ⁽²⁾].

Il sistema di *opt-out* così delineato impone precisi **oneri di auto-organizzazione al *right holder*** che voglia impedire l'uso indiscriminato di opere e contenuti protetti. Diventa pertanto cruciale adottare misure di riserva adeguate allo scopo, che siano coerenti con le esigenze strategiche del *right holder* sui piani del *marketing*, della distribuzione commerciale, della politica culturale. Al riguardo non si può infatti dimenticare che alcuni tra gli operatori automatici (cd. *web crawlers*) operanti *online* svolgono attività sia di *scraping* (dunque, di raccolta massiva di dati a fini di AI *training*) che di indicizzazione, con la conseguenza che una riserva generalizzata di protezione può ostacolare la classificazione del sito *web* e ridurre la visibilità dell'offerta commerciale del *right holder*.

Come evidenzia la *survey* dell'EUIPO-European Union Intellectual Property Office pubblicata nel maggio 2025, **nel campo delle misure di riserva non c'è spazio per** soluzioni universali, del tipo “*one-size-fits-all*” ⁽³⁾. Piuttosto l'approccio deve essere consapevole, flessibile ed articolato, con l'obiettivo di trovare il giusto *mix* tra le misure di carattere giuridico e quelle di natura tecnica, scegliendo dall'ampia gamma a disposizione le soluzioni più adatte per ogni genere di risorsa protetta, per i diversi contesti (*online* oppure *offline*), le varie destinazione d'uso ⁽⁴⁾.

Resta da dire che la naturale vocazione *crossboarder* dell'operatività dei sistemi di AI generativa – che dispensano *output* ovunque, decentrando in molti luoghi la capacità di calcolo e le attività di memorizzazione dei dati – esalta la rilevanza della rigida regolazione euro-unitaria anche per gli operatori d'oltreoceano. E reciprocamente pone agli utilizzatori europei il problema dell'uso dei patrimoni informativi accessibili nei siti di origine statunitense, assoggettati al regime

⁽²⁾ Per l'attuazione di queste previsioni nell'ordinamento italiano v. gli artt. 70ter-70quater l. 633/1941 sul diritto d'autore.

⁽³⁾ V. EUIPO, *The Development of Generative Artificial Intelligence from a Copyright Perspective*, Maggio 2025.

⁽⁴⁾ Per una prima valutazione circa l'idoneità delle misure di riserva v. nell'esperienza tedesca la sentenza del *Landgericht Hamburg* del 27.9.2024 nel caso *LAION v. Robert Kneschke*.

apparentemente più flessibile della *fair use doctrine* ⁽⁵⁾, l'applicazione della quale al TDM è oggetto di ampia discussione ⁽⁶⁾.

Focus: la decisione USA *Bartz v. Anthropic*, 23 giugno 2025

Un monito e un'opportunità

Negli Stati Uniti il dibattito sull'uso dei contenuti protetti per l'addestramento dei sistemi di AI generativa ha dato vita sin qui ad un contenzioso abbastanza ampio. Una recentissima ordinanza della *Norther District Court of California* ⁽⁷⁾ nel caso *Barz v. Anthropic* (la società che sviluppa l'AI *Claude*) offre qualche importante spunto strategico.

Secondo la corte statunitense l'uso non autorizzato di opere letterarie per l'addestramento dei LLM è coperto da *fair use* quando lo sviluppatore del sistema di AI vi abbia avuto accesso in modo legittimo, mentre è illecito quando si tratti di copie ottenute attraverso *download* da "archivi pirata".

La decisione evidenzia la differenza tra il regime statunitense e la disciplina europea, fondata sul sistema delle riserve di protezione, poiché in base a quest'ultima il *right holder* che eserciti correttamente l'*opt-out* può rendere illecito l'uso dei contenuti per il *training* dei sistemi di AI generativa effettuato senza il suo consenso.

3. Strategie di *licensing* dei titolari dei diritti

Adottare un sistema adeguato di riserva è una precondizione per accedere ad una tutela più efficace delle risorse protette contro il loro uso "trasformativo" da parte dei sistemi di AI generativa. Rappresenta d'altro canto la **premessa necessaria** per l'adozione di **un'adeguata strategia di *licensing***, che il *right holder* deve elaborare in via autonoma, sulla base delle peculiarità della propria posizione, valutando poi se attuarla a livello individuale oppure avvalendosi dell'operato degli organismi di gestione collettiva dei diritti d'autore e connessi.

4. Monitoraggio dell'uso delle risorse protette

L'attuazione di strategie di *licensing* può essere utilmente anticipata dal monitoraggio dell'uso delle risorse del *right holder*, al fine di acquisire conoscenza delle caratteristiche e del modo operativo degli operatori e dei settori del mercato che abbiano interesse ad utilizzarle per finalità di TDM o altrimenti connesse all'operatività dei sistemi di AI generativa.

⁽⁵⁾ V. la § 107 del *Title 17 dello US Code (Copyright Law)*.

⁽⁶⁾ V. al riguardo il documento dello *United States Copyright Office* intitolato "*Copyright and Artificial Intelligence Part 3: Generative AI Training*", pubblicato nel maggio 2025.

⁽⁷⁾ Cfr. US District Court for the Northern District of California del 23.6.2025, *Bartz et al. vs. Anthropic*; sull'operatività della *fair use defence* a beneficio degli sviluppatori di LLM v. tuttavia US District Court for the Northern District of California del 25.6.2025, *Kadrey et al. vs. Meta*.

A questo fine potrà risultare opportuno analizzare periodicamente i *report* di trasparenza che gli sviluppatori dei sistemi cd. *general purpose* sono tenuti a pubblicare in base all'art. 53 r. UE 1689/2024 (cd. *AI Act*). Ugualmente potrà essere opportuno svolgere interrogazioni mirate dei sistemi di AI maggiormente usati, per ricavare elementi circa le modalità del loro funzionamento, tra i quali *in primis* l'impiego non autorizzato delle opere del *right holder* per fini di *training* ovvero *AI inference*.

L'esperienza pratica evidenzia infatti che talvolta, quando interagiscono con gli utenti, i sistemi di AI generativa restituiscono **output sovrapponibili** – in tutto o in parte – alle risorse protette dai diritti d'autore o connessi. Ciò può costituire un **indizio rilevante dell'uso non autorizzato** di tali opere – in forma di memorizzazione massiva dei dati da esse estratti, con successiva *regurgitation* di loro porzioni – nell'attività di *training* del modello di AI, ovvero nella fase di *output generation*, che per migliorare la qualità della risposta automatizzata viene condotta talvolta mediante l'analisi di dati tratti da fonti esterne (es.: risorse pubblicamente disponibili nel *web*, banche dati specialistiche accessibili a pagamento, etc.), mediante un processo di *real time inference*.

L'esito del monitoraggio potrà poi guidare la scelta tra le soluzioni praticabili, quali l'offerta e la negoziazione di una licenza oppure l'*enforcement* dei diritti violati.

5. ***Enforcement dei diritti d'autore e connessi***

Quando il titolare dei diritti sulle risorse utilizzate senza consenso non voglia autorizzarne l'uso o non riesca a negoziare licenze resta la strada dell'*enforcement* per via giudiziale, che nel caso dei diritti d'autore e connessi contempla la possibilità di ottenere provvedimenti come gli **ordini inibitori** assistiti da **penalità di mora**, il **risarcimento del danno**, e la **pubblicazione della sentenza** a spese dell'autore della violazione.

Ai fini della formazione della prova è infine possibile richiedere – *inaudita altera parte* – l'efficace **misura cautelare della descrizione**, oltre che esercitare il cd. **diritto d'informazione**, con **accesso ai dati sull'intera filiera** dei soggetti coinvolti nell'illecito.

Practice di Proprietà Intellettuale, Privacy e Information Technology

Michele Bertani

Pasquale Tammaro

Giovanni Guglielmetti

Tommaso Faelli

Ivan Rotunno



Focus Team Innovazione e Trasformazione digitale

Il *Focus Team* è una costellazione di competenze in diversi ambiti di attività con focus su innovazione e trasformazione digitale.

Tommaso Faelli

Proprietà intellettuale

Alessandro De Nicola

Legal Risk, Compliance & Investigations

Marco Adda

Fiscale

Maurizio Pappalardo

Antitrust

Giulia Bianchi Frangipane

Societario

Luca Perfetti

Amministrativo

Enrico Chieppa

Litigation & Dispute Resolution

Vittorio Pomarici

Lavoro

Gianpaolo Ciervo

Bancario

Matteo Viani

Fiscale

Fabrizio Colonna

Financial Regulatory