Intelligenza artificiale, modelli Llm e la sentenza sul caso Gema

Artificial intelligence, Llm models and the ruling on the Gema case

Cecilia Trevisi, Avvocato

Keywords
Intelligenza artificiale, diritto d'autore, modelli Llm

Jel codes
O33, C63, O32

In tema di Intelligenza artificiale, una recente sentenza tedesca sul caso Gema porta un importante contributo alla tutela del diritto d'autore, dell'industria creativa e dei contenuti. La decisione chiarisce come ci si possa difendere da chi, nell'addestrare un modello Large Language Model, saccheggi senza autorizzazione contenuti e dati prelevati dai siti web attraverso lo scraping svolto con specifici software. Si chiarisce che è vietato addestrare i modelli utilizzando contenuti coperti da tutela autorale senza autorizzazione e senza compenso, ridimensionando il rilievo della riserva d'uso. Una sfida tutt'altro che risolta ma circoscritta all'ambito del Text and Data Mining.

A recent German ruling on the Gema case makes an important contribution to the protection of copyright and the creative and content industry. Clarifying how it can be protected from those who, in training a Large Language Model, plunder content and data taken from websites without authorization through scraping carried out with specific software. Making it clear that it is forbidden to train models using content covered by protection without authorization and without compensation, by reducing the importance of the reserve of use. A challenge that is far from solved but limited to the field of Text and Data Mining.

1. Introduzione

Il recente caso Gema portato davanti al tribunale specializzato bavarese, definito con la sentenza dell'11 novembre 2025[1], cattura l'attenzione perché chiarisce il funzionamento dei trasformatori generativi pre-addestrati (Gpt) e, più in generale, le modalità di addestramento di un modello Llm, offrendo allo stesso tempo spunti interessanti per bypassare le difficoltà tecniche dovute allo scraping, ossia all'estrazione dei dati dai siti web effettuata attraverso specifici software.

Si tratta di una pronuncia favorevole agli autori e, più in generale, all'industria creativa che, tuttavia, lascia aperta la questione, tutt'altro che secondaria, di come formulare correttamente la riserva d'uso, così da impedire che contenuti protetti dal diritto d'autore, diffusi in rete, vengano impiegati per l'addestramento di modelli di Ai senza riconoscere ai titolari il compenso loro spettante.

La riserva d'uso assolve proprio a questa funzione: segnalare che un determinato contenuto non è liberamente utilizzabile. Il nodo centrale è quello di comprendere in che modo tale volontà possa essere espressa e resa intelligibile anche ai modelli di Intelligenza artificiale.

Nella sentenza si chiarisce un punto decisivo: l'impiego di opere protette per l'addestramento dei modelli non necessariamente rientra nel Tdm. Ne consegue che solo per opporsi alle attività di Tdm è necessario adottare strumenti idonei a rendere conoscibile alla macchina la volontà del titolare di esercitare la riserva d'uso; negli altri casi tornano pienamente applicabili i principi generali del diritto d'autore, comprese le relative eccezioni e limitazioni.

1 Landgericht München I, Az.: 42 O 14139/24, 11 novembre 2025.

2. Text and Data Mining - Tdm

La direttiva (Ue) 2019/790 (nota come direttiva copyright) consente il libero uso di contenuti autorali nell'ipotesi di estrazione e analisi dei dati (Text and Data Mining, Tdm) in due soli casi: per la ricerca scientifica e per tutti gli altri scopi (quindi anche per finalità commerciali) purché l'accesso ai contenuti avvenga in modo lecito (quindi, ad esempio, i dati non siano estratti da siti illegali, come BitTorrent) e i titolari dei diritti non abbiamo esercitato la riserva d'uso (opt-out) che deve essere esplicitata in modo che la macchina sia in grado di leggerla.

Questa richiesta di mezzi idonei a esplicitare l'opt-out sta mettendo in crisi tecnici, giuristi e, ovviamente imprenditori. Rispondere alla domanda su quale sia la modalità più efficace per esplicitare la riserva d'uso è quasi impossibile anche se il protocollo robots.txt costituisce, per ora, l'unica modalità tecnologica ritenuta adeguata a tale scopo.

Il problema si è creato proprio in seguito all'indicazione nel testo della normativa europea dell'espressione «mezzi appropriati» per esplicitare la riserva d'uso senza alcuna ulteriore specificazione. Questo significa che se i mezzi non sono effettivamente idonei, i contenuti autorali possono essere impiegati dalle imprese private per l'addestramento di Ai.

La prima reazione è consistita nell'inserimento immediato, su numerosi siti Internet, di disclaimer volti a vietare espressamente lo scraping e l'addestramento dei modelli di Intelligenza artificiale. Il problema, tuttavia, è che la macchina non è in grado di leggerli e, pertanto, non può individuarli, comprenderli o conformarsi a tale manifestazione di volontà.

Viste le difficoltà generate dal meccanismo della riserva, la Commissione europea ha promosso uno studio di fattibilità per costituire un registro centrale degli opt-out. Una soluzione, in realtà, molto criticata perché vorrebbe dire che ogni giorno (ogni minuto, secondo, forse) bisognerebbe aggiornare il registro con i contenuti per cui si intende esercitare la riserva d'uso e catturarne anche la correlazione temporale nel breve, medio e lungo periodo.

3. Lo scenario tecnico

Il caso affrontato dinnanzi al tribunale di Monaco riguarda l'utilizzo di una chatbot da parte di utenti che dispongano di un accesso a Internet e di un dispositivo abilitato a Internet (telefono, pc, tablet, ecc.). Di fatto quello che si verifica utilizzando ChatGpt nelle versioni a pagamento.

L'utente pone una domanda al chatbot inserendo un cosiddetto prompt. L'output che il chatbot crea in risposta alla specifica richiesta, il cosiddetto output, viene visualizzato dall'utente nella finestra di dialogo del sito web o dell'app del chatbot. Inoltre, i prompt e gli output per gli utenti registrati vengono memorizzati come cosiddette chat nella cronologia e possono essere recuperati dall'utente in un secondo momento.

Per comprendere le ragioni del verificarsi di un «rigurgito» (knowledge distillation) – ovvero della produzione di output che riproducono esplicitamente determinati input di addestramento – si è reso necessario analizzare il modo in cui vengono memorizzati tali contenuti all'interno del modello in fase di addestramento. Esame che viene fatto dai giudici tedeschi dovendosi occupare dei rigurgiti di testi di canzoni negli output, quindi materiale protetto dal diritto d'autore, a causa dall'addestramento del modello linguistico (Llm) con le opere corrispondenti.

4. La materia del contendere

L'attore è società tedesca di gestione collettiva dei diritti d'autore[2] che amministra i diritti di sfruttamento economico delle opere musicali[3]. Il convenuto (Open Ai) è una società americana che sviluppa, gestisce e concede in licenza tecnologie di Intelligenza artificiale generativa e, in particolare, sistemi di dialogo testuali per la comunicazione in linguaggio naturale, i cosiddetti chatbot di Intelligenza artificiale, basati su modelli linguistici transformer i cui server sono situati in varie località tra cui la Germania.

L'attore avanza richieste di provvedimenti ingiuntivi, informazioni e danni nei confronti della convenuta in ragione della violazione del diritto d'autore e dei diritti personali ri-

2 Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte.

3 Nelle sue condizioni tariffarie a partire dal 2023 per l'utilizzo di musica su Internet, la società di gestione collettiva obbliga i suoi licenziatari a dichiarare una riserva d'uso per il Tdm e, sul suo sito web, dichiara esplicitamente una riserva di utilizzo per le opere contenute nel suo repertorio.

spetto a nove diversi testi di canzoni[4] in quanto riprodotti, in una versione leggermente modificata, come output in risposta ai prompt degli utenti. La memorizzazione dei contenuti all'interno dei modelli porta al rigurgito, ossia output che riproducono esplicitamente determinati input di addestramento. La convenuta viene considerata come il produttore di tali riproduzioni essendo responsabile dell'addestramento dei modelli con i testi delle canzoni.

Il tribunale bavarese affronta il predetto caso, analizzando prima di tutto le modalità con cui si addestra un modello Llm. Uno sforzo tecnico fondamentale, forse così dettagliato è il primo in ambito europeo, per comprendere come (e se) il modello memorizza i dati di addestramento per poi rigurgitarli.

5. L'addestramento

I modelli su cui si basa il chatbot sono i cosiddetti trasformatori generativi pre-addestrati (Gpt) basati su un'architettura di rete neurale che consente il deep learning. Questi modelli, capaci di generare un linguaggio naturale, sono addestrati per mezzo di grandi quantità di dati, inclusi testi provenienti da pagine web di Internet raccolte (crawlate) a partire da testi accessibili pubblicamente[5].

Le reti neurali sono composte da più livelli di neuroni artificiali combinati in una tipologia complessa. I neuroni ricevono l'input di altri neuroni, eseguono un calcolo (somma di prodotti tra ingressi e pesi filtrati attraverso una funzione non lineare) e producono un output che, a sua volta, può essere utilizzato come input per un altro neurone. Nei modelli transformer, un meccanismo di attenzione multi-testa consente di cogliere dipendenze e correlazioni tra i singoli elementi della sequenza di input e di concentrarsi in modo flessibile su di essi durante la generazione dell'output.

I modelli vengono sottoposti a training. Nella fase di pre-addestramento, i dati di training vengono convertiti in testo leggibile dalle macchine. Il testo viene suddiviso in cosiddetti token, che corrispondono a parole o parti di parole. Successivamente, a ciascun token viene assegnato un indice intero univoco, e i token identici presentano lo stesso indice. In questo modo il testo viene trasformato in un formato numerico che consente l'elaborazione computazionale. Durante lo sviluppo di un modello Transformer, nel corso dell'addestramento viene acquisito il significato semantico delle singole parole, rispettivamente dei token, e la loro vicinanza reciproca.

In un'ulteriore fase di addestramento, il modello viene addestrato con prompt selezionati e con gli output ideali corrispondenti, al fine di adeguare i parametri della rete neurale, così da poter produrre un output il più possibile appropriato. Inoltre, al modello vengono insegnate le preferenze umane: diversi output relativi a prompt tipici vengono confrontati da valutatori umani e giudicati qualitativamente, ad esempio in termini di comprensibilità, rilevanza e cortesia.

Dopo aver completato il processo di addestramento, il modello può rispondere alle domande degli esseri umani tramite il chatbot. Quando l'utente inserisce un prompt, l'input viene suddiviso in token e a ciascun token viene associato un vettore che descrive il significato semantico del token e la sua relazione con gli altri token, relazione che è stata ottimizzata durante l'addestramento. Inoltre, ogni token riceve un vettore posizionale, che ne determina la posizione nell'input e riproduce così strutture sintattiche come la costruzione della frase.

Il vettore posizionale viene sommato al vettore che descrive il significato semantico del token e da questa somma viene generato un vettore comune. Per l'input si crea una matrice di vettori, dalla quale per ogni token vengono derivati altri tre vettori (query, key e value). Successivamente, la collocazione di un token nella struttura linguistica viene analizzata e compresa da una rete neurale.

I passaggi vengono ripetuti più volte. Non è ancora del tutto chiaro cosa faccia effettivamente ciascun livello di una rete neurale. Si ipotizza che i primi livelli riconoscano caratteristiche e schemi semplici, come ad esempio grammatica o sintassi, mentre i livelli più profondi colgano astrazioni e relazioni, come ad esempio relazioni semantiche complesse o significati dipendenti dal contesto.

---

4 I testi in questione sono i seguenti: *Atemlos* di Kristina Bach, *36 Grad* di Thomas Eckart, Inga Humpe, Peter Plate e Ulf Leo Sommer, *Bochum* e *Männer* di Herbert Grönemeyer, *Über den Wolken* di Reinhard Mey, *Junge* di Jan Vetter nonché *Es schneit*, *In der Weihnachtsbäckerei* e *Wie schön, dass du geboren bist* di Rolf Zuckowski.

5 Nel file robots.txt, leggibile dalle macchine, è possibile, tramite i cosiddetti tag per i crawler, impedire l'accesso alle pagine web o consentire l'indicizzazione.

Per la selezione dei token di output in risposta a un prompt entrano in considerazione i cosiddetti logit. Quale di questi venga scelto come output viene determinato con l'aiuto della funzione softmax e del parametro temperature (temperatura). La funzione softmax consente di ottenere una distribuzione di probabilità concreta, in cui la somma dei valori di tutti i logit è pari al 100%; senza questo calcolo matematico aggiuntivo, il valore dei logit può essere un numero arbitrario. Un'ulteriore possibilità per selezionare il token di output sono le strategie di campionamento, in questo modo si evita che token molto improbabili possano essere scelti.

Almeno in singoli casi, viene emesso un output con il prompt corrispondente, il cui contenuto è almeno in parte identico a un contenuto presente nel set di dati di addestramento. Questo si verifica perché la sequenza di token che viene generata dal modello è quella che appare statisticamente plausibile perché, ad esempio, contenuta nell'addestramento in una forma particolarmente stabile o ricorrente, e questo si può verificare, ad esempio, perché tale sequenza di token compare in una moltitudine di diverse pagine web accessibili pubblicamente e quindi inclusa nei dati di addestramento più di una volta.

Tuttavia, l'output del modello non è ancora l'output che il chatbot mostra all'utente come output di un prompt. L'output viene creato sulla base dell'output del modello dopo che è stato elaborato dalla cosiddetta struttura di decoding. Durante il decoding, mediante meccanismi di randomizzazione impiegati in modo mirato (meccanismi che introducono un elemento di casualità nella generazione degli output), si producono casualità e variazioni. Il decoding non è, quindi, parte del modello, bensì un passaggio successivo a esso. La generazione di output non deterministici da parte del chatbot (e non del modello) si basa dunque su scelte progettuali.

Nonostante la randomizzazione artificiale nel decoding, i chatbot, nel caso di specie, riproducevano i contenuti memorizzati in modo coerente e con una varianza minima, perché quando si generavano contenuti memorizzati, le probabilità che i token componessero il contenuto memorizzato raggiungevano valori molto alti, spesso vicini al 100%. La riproduzione del contenuto memorizzato diverrebbe così inevitabile e avverrebbe in modo deterministico. A questo proposito, i chatbot funzionerebbero come un database.

Si giunge in questo modo a una prima conclusione decisamente rilevante sul piano giuridico, ossia è possibile stabilire se un output contiene contenuti memorizzati mediante un confronto tra l'output e i dati di addestramento. Se l'output si ritrova nei dati di addestramento, si presume una memorizzazione[6]. È vero che esiste la possibilità teorica di un'identità casuale tra output e dato di addestramento. Tuttavia, almeno a partire da una certa lunghezza del testo, la probabilità di una simile coincidenza sarebbe così estremamente bassa da diventare praticamente trascurabile.

6. L'accertata violazione dei diritti d'autore dei testi delle canzoni

Chiarito che, anche senza conoscere i dati di addestramento e quindi senza possibilità di confronto, la ricerca informatica può stabilire, sulla base della combinazione di diversi metodi se un output includa o meno contenuti memorizzati, nel caso Gema il tribunale è giunto alla conclusione che i testi delle canzoni in questione sono stati memorizzati nel modello nella loro interezza e senza alterazioni. Laddove, nella riproduzione dei contenuti memorizzati, si verificassero modifiche del testo, ciò non dipenderebbe dal modello né dalle informazioni in esso incorporate, bensì dai meccanismi di decoding, che, introducendo una variabilità artificiale, indurrebbero allucinazioni. La riproduzione dei testi controversi non costituirebbe un caso isolato provocato *ad hoc*, ma si verificherebbe in modo costante e ricorrente.

7. L'esercizio corretto della riserva d'uso (opt-out)

Aspetto rilevante per l'accertamento della contraffazione dei contenuti protetti dal diritto d'autore si è rilevata essere la libera disponibilità di tali contenuti in rete e, quindi, l'ef-

6 Al contrario di quanto deciso nel Regno Unito dalla High Court of Justice nella controversia Getty Images v. Stability in cui si è affermato che non vi sia nessuna prova tra la memorizzazione ovvero derivazione diretta tra i materiali di Getty Images (fotografie) e gli output di Stability, poiché si tratterebbe di «prodotti di pattern appresi» e non di utilizzo di copie dei materiali.

fettivo e corretto esercizio della riserva d'uso necessaria per escludere l'attività di Tdm.

L'attrice sostiene che i testi delle canzoni oggetto della controversia non fossero liberamente disponibili su Internet, essendo stata esplicitata la riserva d'uso[7] rispetto a ciascuno dei predetti contenuti, né sarebbero state concesse licenze per l'utilizzo online dei testi delle canzoni. Oltretutto Gema nel concedere licenze in altri ambiti ritiene di obbligare sistematicamente i propri licenziatari al rispetto della riserva d'uso proprio per il Tdm. In altre parole: i contenuti non sarebbero liberamente accessibili, quindi il loro uso finalizzato all'addestramento dei modelli andrebbe autorizzato dal titolare dei diritti e pagato il corrispettivo per tale utilizzo.

In base alla ricostruzione tecnica delle modalità di addestramento, secondo l'attrice le violazioni dei diritti d'autore si sarebbero verificate sia nel modello sia al momento dell'output. L'inserimento dei testi delle canzoni come dati di addestramento nei modelli e la successiva restituzione dei testi in risposta alle richieste effettuate dagli utenti, comporta già di per sé una presunzione del fatto che i testi siano stati memorizzati e riprodotti nei modelli.

Senza dimenticare che i titolari dei diritti non vengono a conoscenza dell'impiego delle loro opere per l'addestramento di Ai né prima né dopo l'uso, anzi i dati di training utilizzati dalle imprese interessate in fase di addestramento vengono considerati come segreti commerciali, evitando in questo modo qualsiasi tipo disclosure.

Punto nodale resta quello di determinare se la riserva d'uso così come esercitata dall'attrice sia (o meno) efficace.

I testi delle canzoni oggetto della controversia sarebbero stati tutti messi su Internet con il consenso dei titolari dei diritti e senza misure di protezione. Ognuno dei nove testi controversi sarebbe stato acquisito tramite crawling da una pagina web priva di file robots.txt, oppure i file robots.txt presenti non avrebbero vietato il crawling da parte del CcBot. In altre parole, i file robtots.txt impostati non vietavano lo scraping.

Tuttavia il protocollo robots.txt (leggibile dalle macchine), ancorché sia riconosciuto come uno standard per il corretto esercizio del opt-out, non è del tutto adatto per la dichiarazione di riserve d'uso nel settore del Tdm. Con robots.txt non sarebbe possibile formulare riserve d'uso riferite a singole opere, poiché tale protocollo consente solo di escludere in modo generale directory, sottodirectory o url con determinate stringhe di caratteri, senza poter identificare con precisione singoli contenuti o opere specifiche. Inoltre, in robots.txt non potrebbero essere disciplinati neppure tipi di utilizzo o condizioni di licenza.

La convenuta ha dunque ritenuto che, nel caso di specie, non vi sarebbero state riserve d'uso efficaci.

Il file robots.txt costituirebbe pertanto un metodo idoneo e formalmente valido (a quanto pare l'unico)[8] per dichiarare la riserva d'uso mentre la riserva d'uso dichiarata dall'attrice inserita solo nel 2024 non sarebbe stata formulata in una forma leggibile dalle macchine e, quindi, è stata ritenuta inadeguata.

Il tribunale di Monaco è stato lapidario, ritenendo che l'eccezione della riserva d'uso non sia invocabile perché la riproduzione integrale dei testi delle canzoni all'interno dei modelli non costituirebbe Tdm. I testi, in quanto dati di addestramento, non sarebbero stati soltanto «valutati» o analizzati, bensì recepiti integralmente nei parametri del modello, con conseguente interferenza con gli interessi di sfruttamento economico degli autori. Non si è quindi trattato di una mera estrazione di testo e dati.

In base ai considerando della direttiva Dsm emerge che, con l'introduzione delle eccezioni per il Tdm, non si intende soltanto promuovere l'innovazione e le nuove tecnologie, ma anche tutelare gli autori. Il considerando 8 della direttiva Dsm indica che, qualora non possano essere invocate eccezioni o limitazioni, per tali atti è necessaria l'autorizzazione del titolare dei diritti («Se non sussistono eccezioni né limitazioni è richiesta un'apposita autorizzazione ai titolari dei diritti»). Anche il successivo considerando 18 sottolinea che i titolari dei diritti potranno continuare a concedere licenze per l'utilizzo delle loro opere o di altri materiali protetti che non rientrano né nell'eccezione obbligatoria prevista dalla direttiva per il Tdm a fini di ricerca scientifica, né nelle eccezioni e limitazioni applicabili ai sensi della direttiva 2001/29/Ce per gli atti di riproduzione temporanea[9].

7 L'art. 4 della direttiva (Ue) 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale che consente l'attività di Tdm riguarda o casi in cui i titolari dei diritti d'autore non abbiano esercitato la riserva d'uso (opt-out) con mezzi appropriati; quindi se i titolari di tali diritti non si sono attivati per adottare nel web e con mezzi appropriati le misure volte a escludere che i propri contenuti on-line siano caricati nei data set utilizzati per l'addestramento di Ai significa che tali contenuti (anche autorali), qualora l'accesso sia legale, possono essere legittimamente impiegati dalle imprese private per l'addestramento.

8 Viste le difficoltà generate dal meccanismo della riserva, la Commissione europea ha promosso uno studio di fattibilità per un registro centrale degli opt-out dall'eccezione di Tdm.

9 L'eccezione di copia temporanea di fatto non è applicabile in quanto alcune delle fasi del processo di Tdm comporta una riproduzione permanente dei dati trattati.

Secondo il giudice tedesco un'interpretazione – presumibilmente favorevole alla tecnica e all'innovazione – che volesse ritenere coperte dall'eccezione anche le riproduzioni nel modello, sarebbe esclusa alla luce del chiaro tenore letterale della direttiva Dsm.

8. Conclusioni

La disciplina delle eccezioni consente, nel Tdm, atti preparatori di riproduzione in un contesto in cui gli interessi di sfruttamento degli autori non sono compromessi, poiché vengono estratte mere informazioni e l'opera, in quanto tale, non viene riprodotta[10]. Nel caso delle riproduzioni nel modello, invece, lo sfruttamento dell'opera verrebbe compromesso in modo duraturo e, per tale via, verrebbero lesi gli interessi legittimi dei titolari dei diritti. Un'applicazione analogica della disposizione di eccezione – che non prevede alcuna remunerazione per lo sfruttamento – lascerebbe dunque autori e titolari dei diritti privi di tutela. Ciò contrasterebbe chiaramente con il considerando 17 della direttiva Dsm.
Sotto il profilo della responsabilità sugli output, il tribunale di Monaco sembra non avere dubbi e individua quale unico responsabile OpenAi in quanto responsabile dell'architettura dei modelli, escludendo che la responsabilità ricada sugli gli utenti che con i loro prompt mettono in moto il processo di generazione automatizzato. Su tale ultimo aspetto conservo delle perplessità perché se gli utenti fossero a conoscenza della non originalità dei risultati prodotti dagli output, una responsabilità concorrente, almeno in alcuni casi, potrebbe essere ravvisata.

10 Nel testo in inglese della direttiva si distingue tra «extraction» (estrazione, come i dati da una banca dati) e «mining» (analisi automatica dei dati per ottenere informazioni). Il mining a fini di addestramento verrebbe attuato attraverso atti di semplice estrazione di dati e informazioni, transitori o comunque temporanei, di per sé irrilevanti o tuttalpiù ricompresi nell'ambito dell'art. 5(1) della Direttiva Infosoc, in quanto non hanno a oggetto i contenuti protetti dal copyright.