

Vettori d'autore: tutela autoriale di vettori e matrici frutto di apprendimento automatico

1. Introduzione

Il dibattito giuridico contemporaneo sull'intelligenza artificiale generativa appare, a un'analisi rigorosa, viziato da una fallacia prospettica: l'ossessiva concentrazione sull'*output esterno* (il testo, l'immagine o il suono generati a valle dell'interazione con l'utente) a discapito dell'*output interno* del processo di addestramento. Sebbene la dottrina si sia ampiamente interrogata sulla tutelabilità delle opere create "dall'IA" o "con l'IA", una zona d'ombra normativa avvolge tuttora l'asset di maggior valore economico e strategico dell'intera filiera algoritmica: i pesi del modello (*model weights*).

Oggetto della presente indagine non è dunque la singola istanza generativa, bensì la natura giuridica di quelle matrici multidimensionali di valori numerici (parametri) che, cristallizzando le correlazioni statistiche apprese durante il *training*, costituiscono l'essenza operativa di un *Large Language Model* (LLM) o di un sistema di diffusione per la creazione di immagini.

La questione non è meramente speculativa. I pesi di un LLM rappresentano investimenti nell'ordine delle centinaia di milioni di dollari; costituiscono il nucleo del valore economico delle imprese che li sviluppano; formano oggetto di accordi di licenza e di controversie giudiziarie in molteplici giurisdizioni. Ciononostante, la loro qualificazione giuridica rimane incerta. Il presente contributo intende dimostrare che le categorie tradizionali del diritto d'autore italiano ed europeo offrono strumenti adeguati ad inquadrare questa nuova fatispecie, a condizione di procedere con una concreta ed esatta analisi del fenomeno. Per questo motivo si proverà almeno ad individuare la terminologia informatica pertinente, rimandando, per motivi di spazio, alla letteratura di settore per un quadro più esplicativo sulle specificità del processo algoritmico.

Una premessa metodologica è d'obbligo, laddove la dottrina dovesse ancora insinuare dei dubbi: l'Intelligenza Artificiale (leggasi «l'attuale insieme di tecniche di apprendimento automatico computazionale»), ai fini di questo contributo, non sarà mai considerata in alcun modo come un soggetto, bensì unicamente come un processo, ovvero uno specifico tipo di attività ordinata e predeterminata. Un processo complesso, che coinvolge molteplici attori rigorosamente umani (autori dei dati, curatori dei *dataset*, progettisti dell'architettura, tecnici del *training*), ciascuno dei quali compie scelte potenzialmente rilevanti ai fini della tutela autorale; così come altrettanti diversi oggetti, con caratteristiche e tutele differenti.

In nessun caso sarà considerato giuridicamente paragonabile ad un agente.

Da tale premessa discende un corollario: come una fotografia non può essere «fatta» da una fotocamera, così un modello di IA (e ancora meno il suo *output*) non può essere «creato» dalla macchina, bensì dai soggetti umani che hanno concorso a determinarne le condizioni di produzione.

La tesi centrale parte quindi dal seguente assunto: i pesi di un modello di Intelligenza Artificiale sono proteggibili ai sensi del diritto d'autore nella misura in cui risultano dall'addestramento su opere dell'ingegno di carattere creativo. Un modello addestrato su testi letterari, composizioni musicali, opere d'arte visiva produce pesi che incorporano, in forma trasformata, l'apporto creativo degli autori originari. Per converso, un modello addestrato esclusivamente su dati empirici (rilevazioni fisiche,

serie statistiche) non incorpora alcun apporto creativo, potrà semmai formare oggetto di tutela brevettuale e sarà certamente potenziale oggetto di protezione quale segreto commerciale.

L'indagine si concentrerà sul quadro normativo italiano ed europeo, con particolare riferimento alla Legge n. 633 del 1941 (Legge sul Diritto d'Autore, LDA), alla Direttiva 96/9/CE sulle banche dati, alla Direttiva 2001/29/CE (Direttiva InfoSoc) e alla Direttiva UE 2019/790 (Direttiva Copyright).

2. Anatomia tecnica del processo di *training*

2.1 Architettura del modello

Un modello di apprendimento automatico, nella sua declinazione più complessa di rete neurale artificiale, è una struttura matematica composta da unità di calcolo (neuroni) organizzate in strati (*layer*) successivi. Ciascun neurone riceve in ingresso valori numerici, li moltiplica per coefficienti detti pesi (*weights*), applica una funzione di attivazione non lineare e produce un valore in uscita che diviene *input* per lo strato successivo.

L'architettura di un *Large Language Model* contemporaneo si fonda sul meccanismo dell'*attention*, introdotto nel 2017 con il modello *Transformer*¹. Tale architettura consente al modello di «pesare» differentemente le diverse parti dell'input, stabilendo relazioni contestuali tra elementi anche distanti nella sequenza. I parametri che governano questo meccanismo (matrici *Query*, *Key*, *Value*, pesi degli strati *feed-forward*, *embedding* dei *token*) costituiscono, nel loro insieme, i «pesi» del modello.

Un modello come GPT-4 conta centinaia di miliardi di parametri: numeri in virgola mobile, tipicamente rappresentati con 16 o 32 bit, organizzati in matrici multidimensionali. Prima dell'addestramento, questi valori sono inizializzati casualmente; al termine del *training*, codificano le correlazioni statistiche apprese dal corpus di dati.

2.2 Il *dataset* di *training*

Il *dataset* di addestramento è l'insieme dei dati su cui il modello «apprende». Per un modello linguistico, si tratta di un corpus testuale che può comprendere miliardi di documenti: pagine web, libri, articoli, codice sorgente.²

La composizione del *dataset* non è casuale. La selezione delle fonti, il bilanciamento tra domini, la rimozione di contenuti inappropriati, la deduplicazione dei testi: tutte queste operazioni richiedono scelte discrezionali che incidono sostanzialmente sul comportamento finale del modello. Il curatore del *dataset* compie un'attività analoga a quella del compilatore di un'antologia.

La preparazione dei dati include la *tokenizzazione*: la segmentazione del testo in unità discrete (*token*). Un *tokenizer* può operare a livello di parole, sotto-parole o caratteri, secondo algoritmi come *Byte-Pair Encoding* (BPE) o in altri e nuovi modi ancora da scoprire. La scelta dell'algoritmo influenza la capacità del modello di gestire lingue diverse e termini tecnici.

2.3 Il processo di *training*

L'addestramento di una rete neurale procede per iterazioni. A ogni passo, un *batch* di dati viene presentato al modello (*forward pass*); il modello produce una predizione (nel caso dei LLM, la

¹ A. VASWANI et al., *Attention Is All You Need*, in Advances in Neural Information Processing Systems, 2017.

² Per un'analisi della composizione dei *dataset* di addestramento dei principali LLM, v. J. DODGE et al., *Documenting Large Webtext Corpora*, in Proceedings of EMNLP, 2021.

distribuzione di probabilità sul *token* successivo); la predizione viene confrontata con il dato reale mediante una funzione di perdita (*loss function*); l'errore viene comunicato ai neuroni precedenti attraverso la rete (*backpropagation*), calcolando il gradiente della perdita (*loss*) rispetto a ciascun peso; infine, i pesi vengono aggiornati nella direzione che minimizza l'errore (*gradient descent*), secondo un tasso di apprendimento (*learning rate*) e un algoritmo di ottimizzazione.

Questo ciclo si ripete miliardi di volte. Il *training* di un modello di frontiera richiede settimane di calcolo su migliaia di GPU, con un consumo energetico nell'ordine dei gigawattora.³

Le scelte compiute in questa fase (numero di epoche, *learning rate*, dimensione del *batch*, tecniche di regolarizzazione, selezione dei *checkpoint*) non sono meramente tecniche: determinano quali correlazioni il modello apprenderà, quali «dimenticherà», come bilancerà fedeltà ai dati e capacità di generalizzazione. Il *data scientist* che sovrintende a questo processo compie scelte creative non dissimili da quelle di un regista che seleziona le inquadrature da montare. Quanto più i dati di partenza saranno artistici e sfuggiranno alle generalizzazioni, tanto più il suo contributo sarà di carattere «intuitivo» e quindi «creativo», diventando difficile concepire metriche oggettive di ottimizzazione.

2.4 I pesi come *output*

Al termine del *training*, i pesi costituiscono una rappresentazione compressa e trasformata del *dataset*. Non si tratta di una copia: le opere originarie non sono «contenute» nei pesi come file in un archivio. Si tratta piuttosto di una *codifica statistica* delle regolarità presenti nei dati.

Un traduttore umano che abbia letto migliaia di romanzi non «contiene» quei romanzi nella propria mente, ma ne ha estratto competenze linguistiche e sensibilità stilistiche. Analogamente, un LLM addestrato su un corpus letterario non «memorizza» le opere: apprende distribuzioni di probabilità su sequenze di *token* che riflettono i *pattern* statistici del corpus.

Questa analogia non deve indurre in tentazioni umanizzanti: questo è l'esatto valore aggiunto concesso da ormai un secolo di lavoro sulla codificazione della conoscenza astratta; la somiglianza della struttura artificiale con quella naturale non può portare a confusioni ontologiche giuridicamente rilevanti, a meno di voler riconoscere soggettività giuridica anche ai tavoli perché si sorreggono su delle «gambe».

Questa distinzione è cruciale sul piano giuridico. La capacità di un modello di riprodurre, in certi casi, frammenti di opere del *training set* (fenomeno noto come *memorization*) è un effetto collaterale, non una funzione intenzionale. La memorizzazione è un *bug*, non una *feature*.⁴

2.5 Soggetti coinvolti e rispettivi apporti

La filiera produttiva di un modello di apprendimento automatico coinvolge quindi molteplici soggetti, ciascuno con un apporto specifico e una corrispondente posizione giuridica.

Gli autori delle opere di *training* creano i contenuti (testi, immagini, musiche) che costituiscono la materia prima del *dataset*; il loro apporto è creativo a titolo originario. Il *curatore del dataset* seleziona, organizza e bilancia i dati, svolgendo un'attività che può integrare i requisiti della banca dati ai sensi dell'art. 102-bis LDA, ovvero dell'opera collettiva ai sensi dell'art. 3 LDA, a seconda che

³ E. STRUBELL et al., *Energy and Policy Considerations for Deep Learning in NLP*, in Proceedings of ACL, 2019.

⁴ Si veda il successivo 5.4 in questo senso.

prevalga l'investimento o la creatività nella selezione. L'*architetto del modello* progetta la struttura della rete neurale: l'implementazione specifica è tutelabile come *software* ai sensi dell'art. 2, n. 8, LDA, mentre i principi matematici sottostanti rimangono liberamente appropriabili. Il *data scientist* conduce il *training*, compiendo scelte discrezionali su iperparametri, regolarizzazione, *checkpoint*: il suo contributo creativo si fonde inscindibilmente nel risultato finale. L'*AI Company*, infine, organizza e finanzia la produzione, assumendo il rischio economico dell'impresa: la sua posizione è funzionalmente analoga a quella del produttore cinematografico, come proveremo ad argomentare successivamente.

3. Pesi al vaglio delle categorie esistenti

3.1 Pesi come software?

La prima ipotesi da vagliare è che i pesi costituiscano «programmi per elaboratore» ai sensi dell'art. 2, n. 8, LDA. Tale norma protegge i programmi «in qualsiasi forma espressi purché originali quale risultato di creazione intellettuale dell'autore», escludendo espressamente «le idee e i principi che stanno alla base di qualsiasi elemento di un programma, compresi quelli alla base delle sue interfacce».

L'esclusione di questa ipotesi appare motivata da ragioni sistematiche. Un programma per elaboratore, nella definizione consolidata, è un insieme di istruzioni che, eseguite da un elaboratore, producono un risultato determinato. I pesi non sono istruzioni: sono *dati* letti e utilizzati da un programma (il codice che implementa l'architettura del modello), ma non ne fanno parte. La distinzione è netta sul piano ontologico: il codice di *PyTorch* o *TensorFlow* è *software*; i *files* «*.safetensors*» che contengono i pesi sono dati.

Si potrebbe obiettare che i pesi, combinati con il codice, formano un sistema inscindibile, e che la loro separazione sarebbe artificiosa. Ma questo argomento prova troppo: anche un database è «inscindibile» dal software che lo interroga, senza per questo divenire esso stesso *software*. Il legislatore ha infatti distinto con cura la tutela del programma (Direttiva 2009/24/CE, recepita agli artt. 64-bis ss. LDA) dalla tutela della banca dati (Direttiva 96/9/CE). La tutela del programma per elaboratore copre l'implementazione dell'architettura neurale, non i parametri appresi durante l'addestramento.

3.2 Pesi come banca dati?

La seconda ipotesi è che i pesi costituiscano una «banca dati» ai sensi della Direttiva 96/9/CE, recepita agli artt. 64-*quinquies* e 102-*bis* ss. LDA. La direttiva offre due livelli di tutela: il diritto d'autore sulla struttura (art. 3) e il diritto *sui generis* sull'investimento (art. 7).

L'art. 1, par. 2, definisce la banca dati come «una raccolta di opere, dati o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili». Il requisito degli «elementi indipendenti» è dirimente. La giurisprudenza della Corte di Giustizia ha chiarito che tale requisito implica che ciascun elemento possieda un valore informativo autonomo: nel caso *Fixtures Marketing* (C-444/02), la Corte ha precisato che gli elementi devono essere «separabili gli uni dagli altri senza che il valore del loro contenuto [...] ne risulti pregiudicato».

I singoli valori numerici di una matrice di pesi non soddisfano tale requisito. Ciascun peso acquista significato esclusivamente in relazione a tutti gli altri; non è possibile «accedere individualmente» a un peso isolato per estrarne informazione utile. La struttura è intrinsecamente olistica, non modulare.

Il diritto *sui generis* di cui all'art. 7, poi, tutela l'investimento «rilevante» nel «conseguimento, verifica o presentazione» del contenuto. Anche qualora si volesse forzare la qualificazione, l'operazione di training non è di sistematizzazione o disposizione, ma di vera e propria de-strutturazione e rielaborazione.

3.3 Pesi come opera autonoma?

La terza ipotesi è che i pesi costituiscano puramente un'«opera dell'ingegno di carattere creativo» ai sensi dell'art. 1 LDA, proteggibile a titolo originario.

Questa qualificazione incontra un ostacolo strutturale nell'art. 6 LDA, a norma del quale «il titolo originario dell'acquisto del diritto di autore è costituito dalla creazione dell'opera, quale particolare espressione del lavoro intellettuale». L'opera deve essere espressione *di un autore umano*, come confermato dalla costante giurisprudenza nazionale ed europea.

È significativo, in proposito, il recente orientamento della Cassazione in tema di intelligenza artificiale. Nell'ordinanza n. 1107 del 16 gennaio 2023, la Suprema Corte ha ribadito che la tutela autorale presuppone «l'apporto creativo di un soggetto umano», escludendo che un elaborato generato autonomamente da un sistema informatico possa essere qualificato come opera dell'ingegno.⁵

I pesi, considerati in sé, non sono pensati e tantomeno «creati» materialmente da nessun individuo specifico in senso giuridicamente rilevante: emergono da un processo di ottimizzazione matematica. L'obiezione non è che manchino di creatività (la creatività, come abbiamo visto, c'è, ma è stratificata nei contributi umani a monte), bensì che non esiste un singolo autore cui imputare la «creazione» dei pesi come atto unitario. Da qui nasce la bizzarra necessità per una Corte Suprema di ricordare che ad un sistema informatico, ovvero un singolare ammasso di minerali elettrificati, non può essere attribuito ingegno o creatività.

Questo *impasse* può essere superato riqualificando i pesi non come opera originaria, ma come opera derivata e composta, nel senso che si chiarirà al successivo capitolo 4. In tale prospettiva, la creatività non è assente, ma distribuita tra molteplici soggetti umani i cui apporti confluiscono in un risultato finale.

3.4 Pesi come segreto industriale

L'ultima ipotesi è che i pesi siano tutelabili come «informazioni aziendali riservate» ai sensi degli artt. 98-99 del D.lgs. 10 febbraio 2005, n. 30 (Codice della Proprietà Industriale, CPI). Molte imprese, di fatto, non pubblicano i pesi dei propri modelli, mantenendoli segreti.

L'art. 98 CPI richiede che le informazioni: (a) siano segrete, nel senso che non siano «nel loro insieme o nella precisa configurazione e combinazione dei loro elementi generalmente note o facilmente accessibili agli esperti»; (b) abbiano «valore economico in quanto segrete»; (c) siano «sottoposte, da parte delle persone al cui legittimo controllo sono soggette, a misure da ritenersi ragionevolmente adeguate a mantenerle segrete».

La tutela del segreto industriale, tuttavia, presenta limiti strutturali che la rendono inadeguata a risolvere il problema della qualificazione dei pesi.

In primo luogo, si tratta di una tutela *fattuale*, non di un diritto di proprietà intellettuale in senso proprio: protegge contro l'acquisizione, l'utilizzo e la divulgazione illecita, ma non conferisce una

⁵ Cass. civ., ord. 16 gennaio 2023, n. 1107, in *Dir. aut.*, 2023, p. 215 ss.

privativa opponibile *erga omnes*. Chi sviluppi indipendentemente pesi funzionalmente analoghi, senza appropriarsi illecitamente di quelli altrui, non commette alcun illecito.

In secondo luogo, la tutela viene meno con la divulgazione, anche se involontaria. I modelli distribuiti con licenze *open source* (LLaMA, Mistral, Falcon) non beneficiano di alcuna protezione una volta che i pesi sono stati resi pubblici.

In terzo luogo, e soprattutto, il segreto industriale non riconosce i diritti degli autori delle opere di training. Se i pesi incorporano, in forma trasformata, l'apporto creativo di migliaia di autori, la tutela come segreto industriale giova esclusivamente all'impresa che detiene il modello, lasciando gli autori originari privi di tutela e di remunerazione.

La protezione come segreto industriale può dunque affiancarsi, ma non sostituirsi, a una corretta qualificazione sul piano del diritto d'autore.

4. Pesi come opera composta e derivata

4.1 La categoria dell'opera composta nella L. 633/1941

Nessuna delle categorie esaminate singolarmente appare soddisfacente. La soluzione va cercata in una diversa direzione: i pesi come opera composta risultante dalla combinazione inscindibile di apporti creativi distinti.

Le «opere composte» non sono definite esplicitamente dalla legge sul diritto d'autore, ma la dottrina e la giurisprudenza le identificano in quelle opere unitarie che risultano dall'assemblaggio di contributi eterogenei, ciascuno dotato di autonoma tutelabilità ma funzionalmente integrato nel tutto. L'esempio paradigmatico è l'opera cinematografica, disciplinata agli artt. 44-50 LDA.

L'art. 44 LDA individua i coautori del film negli autori del soggetto, della sceneggiatura, della musica e nel direttore artistico. Ciascuno di questi contributi è autonomamente tutelabile (il soggetto come opera letteraria, la musica come composizione etc...); ma è la loro combinazione, sotto la direzione creativa del regista e l'organizzazione economica del produttore, a generare l'opera cinematografica come entità distinta.

Le opere drammatico-musicali quali le opere liriche, le operette, le composizioni musicali con parole, i balletti, possono fornire ulteriori esempi significativi, in quanto combinazione di elementi che, seppur fruibili separatamente (la musica, il testo, la coreografia...), combinate offrono una unità di funzioni che altrimenti non sarebbe percepibile. La caratteristica distintiva dell'opera composta è dunque la sinergia funzionale: i singoli contributi, pur conservando autonoma tutelabilità, acquistano un valore aggiunto dalla loro combinazione.

Un precedente significativo e più vicino alla dimensione digitale è offerto dalla giurisprudenza in materia di videogiochi. Con sentenza del 17 ottobre 1983, il Tribunale di Torino ha stabilito: «I videogiochi non sono considerati programmi per elaboratore, bensì opere composte [...] assimilati all'opera cinematografica, in quanto il risultato creativo è composto dalla combinazione di una storia e di un *software*, fra loro tuttavia nettamente distinti e separatamente tutelabili».⁶

L'analogia con i modelli di IA è stringente. Anche un LLM risulta dalla combinazione di elementi eterogenei: le opere del *dataset*, la struttura del *dataset*, l'architettura del modello, le scelte del

⁶ Trib. Torino, 17 ottobre 1983, in *Dir. aut.*, 1984, p. 57 ss.

training. Elementi «nettamente distinti e separatamente tutelabili», ma che solo combinati producono il modello funzionante.

Approfittando dell'analogia, potremmo identificare i «coautori» del modello di IA negli autori delle opere di *training*, nel curatore del *dataset*, nell'architetto del modello e nel *data scientist* che supervisiona l'addestramento, tenendo conto che essi possono talvolta sovrapporsi.

4.2 Struttura soggettiva: la creatività stratificata

La nozione di «creatività stratificata» consente di superare l'obiezione per cui i pesi non avrebbero un autore umano. La creatività c'è, ma è distribuita tra molteplici soggetti i cui apporti si sedimentano nel risultato finale.

Tale stratificazione può essere così schematizzata:

Soggetto	Apporto	Tutela
Autori opere di <i>training</i>	Creazione originaria	Diritto d'autore pieno (artt. 1-5 LDA)
Curatore del <i>dataset</i>	Selezione/organizzazione	Banca dati (art. 102-bis LDA)
Architetto del modello	Implementazione	<i>Software</i> (art. 2, n. 8, LDA)
<i>Data scientist</i>	Scelte di <i>training</i>	Contributo umano all'opera composta
<i>AI Company</i>	Organizzazione produzione	Diritti connessi del produttore (art. 45 LDA)

Nessuno di questi soggetti, singolarmente, è «autore» dei pesi. Ma tutti insieme concorrono a determinarne la configurazione finale, esattamente come nessun singolo soggetto è «autore» di un film, pur essendo il film opera dell'ingegno.

4.3 La categoria dell'opera derivata ex art. 4 L. 633/1941

La qualificazione come opera composta spiega *chi* contribuisce a *cosa*. Ma come si giustifica la propagazione dei diritti degli autori delle opere di *training* nei pesi? Qui interviene la categoria dell'opera derivata.

L'art. 4 LDA dispone: «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».

L'incipit della norma («senza pregiudizio dei diritti esistenti sull'opera originaria») costituisce il cardine della nostra ricostruzione. Esso sancisce la sopravvivenza dei diritti degli autori originari anche dopo la trasformazione dell'opera in una nuova forma. La connessione tra opera originaria (o *database* di opere originali) e opera derivata è diretta ed innegabile, dunque lo sfruttamento commerciale di quest'ultima non potrà essere effettuato senza il necessario consenso del titolare dei diritti sulla prima.

Il *training* opera una «trasformazione» delle opere originarie in una «altra forma»: da testo (o immagine, o suono) a vettore numerico. Si tratta di un mutamento radicale di *genus*, ma la norma non richiede che la forma derivata appartenga allo stesso genere dell'originale. La traduzione è già trasformazione da una lingua a un'altra; il passaggio da romanzo a film è trasformazione da un

medium a un altro; l'adattamento teatrale di un'opera letteraria è trasformazione da un sistema di segni a un altro. Il passaggio da opera a vettore è, analogamente, trasformazione da un sistema semiotico (linguaggio umano) a un altro (rappresentazione matematica).

4.4 Trasformazione di *genus* e sopravvivenza dei diritti

L'obiezione prevedibile è che nei pesi l'opera originaria non sia più «riconoscibile», e che dunque non si possa parlare di derivazione.

Tuttavia, essa coglie solo parzialmente il fenomeno. I pesi non «risalgono all'idea» scartando l'espressione: *trasformano* l'espressione in un nuovo formato. La dipendenza funzionale permane: senza quelle specifiche opere nel *training set*, i pesi sarebbero diversi. La trasformazione non è dissolvimento dell'espressione, ma sua *codifica* in un diverso sistema di rappresentazione: l'opera originaria non è leggibile nei vettori, ma senza di essa quei vettori non esisterebbero.

Peraltro, l'art. 18 LDA conferma l'ampiezza del diritto di elaborazione: «Il diritto esclusivo di elaborare comprende tutte le forme di modificaione, di elaborazione e di trasformazione dell'opera previste nell'art. 4». L'espressione «tutte le forme di trasformazione» è volutamente onnicomprensiva e non suggerisce un *numerus clausus*.

4.5 Diritti morali e loro sopravvivenza nell'opera derivata

Un profilo essenziale concerne la sopravvivenza dei diritti morali degli autori delle opere di *training*.

L'art. 20 LDA stabilisce: «Indipendentemente dai diritti esclusivi di utilizzazione economica dell'opera [...] ed anche dopo la cessione dei diritti stessi, l'autore conserva il diritto di rivendicare la paternità dell'opera e di opporsi a qualsiasi deformazione, mutilazione od altra modificaione [...] che possano essere di pregiudizio al suo onore o alla sua reputazione».

I diritti morali sono intrasferibili, irrinunciabili e imprescrittibili; possono essere fatti valere senza limiti di tempo e *erga omnes*.

Nel contesto dei modelli di IA, ciò comporta che se un modello produce *output* che deformano o mutilano le opere originarie in modo pregiudizievole per l'onore degli autori, questi ultimi conservano il diritto di opporsi. Il diritto morale all'integrità sopravvive dunque alla trasformazione in vettori numerici, alla stregua di qualsiasi altro diritto della personalità che dovesse emergere nel caso concreto⁷.

Parimenti, il diritto alla paternità trova fondamento negli obblighi di trasparenza dell'AI Act (Reg. UE 2024/1689), che impone ai fornitori di rendere disponibile «una sintesi sufficientemente dettagliata» dei contenuti di *training* (art. 53, par. 1, lett. d).

4.7 Il ruolo del produttore: l'*AI Company* e i diritti connessi

L'art. 45 LDA stabilisce: «L'esercizio dei diritti di utilizzazione economica dell'opera cinematografica spetta a chi ha organizzato la produzione dell'opera stessa». Il produttore non è autore, ma imprenditore: assume il rischio economico, coordina i fattori produttivi, acquisisce i diritti di sfruttamento.

⁷ Si pensi ai diritti d'immagine nel caso di un modello generativo di video progettato per produrre artefatti coerenti di uno o più individui specifici.

La posizione dell'*AI Company* è strutturalmente analoga. OpenAI, Anthropic, Google non sono «autori» dei modelli che sviluppano, ma organizzatori della produzione. Come il produttore cinematografico, l'*AI Company* finanzia l'acquisizione dei dati e l'infrastruttura di calcolo, coordina i contributi dei diversi soggetti (curatori, architetti, *data scientist*), assume il rischio economico dell'impresa e immette il prodotto sul mercato.

Applicando analogicamente l'art. 45 LDA, all'*AI Company* spetterebbe l'esercizio dei diritti di utilizzazione economica dei pesi, fermi restando i diritti degli autori originari. Il modello cinematografico prevede infatti meccanismi di tutela per gli autori: l'art. 46 LDA stabilisce che «gli autori dell'opera cinematografica hanno diritto a un equo compenso a carico degli esercenti dei locali cinematografici per ciascuna proiezione in pubblico delle opere stesse»; l'art. 46-bis prevede un equo compenso per la comunicazione al pubblico. Queste sarebbero disposizioni rivolte ai fornitori e ai *deployer* di sistemi di IA ai sensi dell'art. 3 paragrafo 1, numeri 3 e 4 dell'AI Act.

Tale ricostruzione ha il pregio di allocare coerentemente diritti e responsabilità: l'*AI Company* può sfruttare commercialmente il modello, ma deve remunerare gli autori delle opere di *training*; gli autori conservano pretese economiche e morali, ma non possono paralizzare lo sfruttamento dell'opera composta.

5. L'eccezione TDM e i suoi limiti

5.1 Il text and data mining nella Direttiva 2019/790

La Direttiva (UE) 2019/790 ha introdotto agli artt. 3 e 4 eccezioni obbligatorie per il *text and data mining* (TDM), recepite in Italia agli artt. 70-ter e 70-quater LDA

L'art. 2, n. 2, definisce il TDM come «qualsiasi tecnica di analisi automatizzata volta ad analizzare testi e dati in formato digitale avente lo scopo di generare informazioni inclusi, a titolo non esaustivo, modelli, tendenze e correlazioni». La definizione comprende espressamente la generazione di «modelli», rendendo applicabile l'eccezione al *training* di sistemi di apprendimento automatico.

L'art. 3 prevede un'eccezione per la ricerca scientifica, riservata a organismi di ricerca e istituti di tutela del patrimonio culturale. L'art. 4, di portata più generale, estende l'eccezione a qualsiasi soggetto: «Gli Stati membri dispongono un'eccezione o una limitazione [...] per le riproduzioni e le estrazioni effettuate da opere o altri materiali cui si abbia legalmente accesso ai fini dell'estrazione di testo e di dati».

Il punto cruciale della nostra analisi è il seguente: l'eccezione TDM legittima l'atto di estrazione e riproduzione ai fini dell'analisi, ma non incide sulla natura giuridica del risultato. Se il risultato del TDM è un'opera derivata, tale resta indipendentemente dalla liceità del processo che l'ha generata.

Un'analogia chiarisce il punto. L'eccezione per la citazione di cui all'art. 70 LDA legittima la riproduzione di brani di un'opera a fini di critica, discussione o insegnamento. Ma se un autore compone un'opera interamente fatta di citazioni (si pensi ai *found poetry* o ai *cut-up* letterari), tale opera rimane derivata dalle opere citate, e il suo sfruttamento commerciale richiede il consenso degli autori originari. La liceità della citazione non cancella la derivazione.

Analogamente, l'eccezione TDM può legittimare il *training* in quanto atto di riproduzione ed estrazione, ma non trasforma i pesi (il risultato) in un'opera libera da vincoli. Se i pesi sono

qualificabili come opera derivata da opere protette, i diritti degli autori originari sopravvivono «senza pregiudizio» ai sensi dell'art. 4 LDA.

Il Considerando 17 della Direttiva 790/2019 conferma questa lettura: «Il pregiudizio potenzialmente arrecato ai titolari dei diritti da detta eccezione dovrebbe essere minimo». Tale affermazione presuppone che l'eccezione non annulli i diritti, ma ne limiti l'esercizio in un contesto specifico.

5.2 Regime di *opt-out* e implicazioni

L'art. 4, par. 3, della Direttiva subordina l'eccezione generale all'assenza di «riserva» da parte dei titolari: «L'eccezione [...] si applica a condizione che l'utilizzo delle opere [...] non sia stato espressamente riservato dai titolari dei diritti in modo appropriato, ad esempio attraverso strumenti che consentano lettura automatizzata».

Il meccanismo dell'*opt-out* consente agli autori di opporsi preventivamente all'uso per il *training*, ma fa ricadere su di essi l'onere di attivarsi, ribaltando il principio per cui ogni sfruttamento richiede consenso preventivo.

La giurisprudenza non ha ancora chiarito le conseguenze dell'*opt-out* tardivo: se un autore esprime riserva dopo il *training*, può pretendere la rimozione dai pesi? Sul piano tecnico, l'operazione (*machine unlearning*) è difficilmente operabile e di sicuro antieconomica. Sul piano giuridico, la questione resta aperta.

5.3 Critica della sentenza GEMA v. OpenAI

La sentenza del Landgericht München I nel caso *GEMA v. OpenAI* (28 novembre 2024, causa n. 42 O 14139/24) costituisce la prima pronuncia europea che affronta *ex professo* la «memorizzazione» di opere protette nei parametri di un modello linguistico.

Il tribunale ha ritenuto che la *Memorisierung* dei testi delle canzoni nei parametri costituisca riproduzione ai sensi dell'art. 2 della Direttiva InfoSoc, e che l'eccezione TDM non si applichi perché la memorizzazione «eccede l'analisi transitoria». L'argomentazione si fonda su un'analogia con la compressione MP3: come un file MP3 contiene l'opera in forma compressa ma ricostruibile, così i pesi conterrebbero le opere «codificate come puzzle» (*wie ein Puzzle gespeichert*).

Tale analogia è tecnicamente errata. La compressione MP3 è un algoritmo *reversibile per progetto*: ideato per ricostruire un'approssimazione dell'onda sonora originale. I pesi di un LLM non «contengono» le opere in alcun senso analogo. Non esiste un algoritmo che, dati i pesi, restituisca le opere di *training*: queste sono *trasformate* in correlazioni statistiche che non preservano l'informazione originale.

Il fenomeno della *memorization* è un effetto collaterale, non una funzione. Dipende da fattori contingenti (frequenza nel *dataset*, dimensione del modello, temperatura del *sampling*) ed è oggetto di tecniche di mitigazione proprio perché considerato un comportamento indesiderato. Fondare la qualificazione giuridica su un *bug* anziché sul funzionamento ordinario del sistema è metodologicamente scorretto.

La sentenza confonde inoltre due piani: la riproduzione *durante* il *training* (quando le opere sono caricate in RAM per il calcolo) e la presunta riproduzione *nei* pesi. La prima può essere coperta dall'eccezione TDM; la seconda non si configura tecnicamente come «copia».

6. Output e apporto dell'utente

6.1 Il prompt come «ultimo parametro»

Quando un utente interroga un modello generativo, fornisce un *prompt* che condiziona l'*output*. Sul piano tecnico, il *prompt* viene tokenizzato, trasformato in vettori di *embedding*, e utilizzato come input per il processo di generazione autoregressiva. La risposta del modello è funzione congiunta dei pesi (fissi al momento dell'inferenza) e del *prompt* (variabile e deciso dall'utente).

In questo senso, il *prompt* può essere concepito come l'«ultimo parametro» del sistema: un input che, combinato con la struttura sedimentata nei pesi, determina l'*output* specifico. L'analogia con la fotografia è ancora una volta illuminante: il fotografo sceglie inquadratura, esposizione, momento dello scatto; la fotocamera (con le sue caratteristiche tecniche) trasforma quelle scelte in immagine. Né il fotografo né la fotocamera, da soli, «creano» la foto: è la loro interazione a produrre il risultato.

6.2 Riformulazione del problema: apporto utente vs. apporto sistema

Il dibattito corrente tende a formulare il problema nei termini: «qual è l'apporto creativo dell'utente rispetto all'apporto della macchina?». Tale formulazione è viziata dall'antropomorfismo più volte rievocato. La macchina non ha «apporto» in senso giuridicamente rilevante: è un *medium* attraverso cui si esprimono gli apporti umani sedimentati nei pesi.

La domanda corretta è dunque: «qual è l'apporto creativo dell'utente rispetto all'apporto degli altri soggetti umani che hanno contribuito a creare il sistema?».

Così riformulato, il problema ammette soluzioni coerenti con le categorie esistenti. Se l'apporto dell'utente è minimo, l'*output* dipende massimamente dai pesi, che a loro volta dipendono dal *dataset*: i diritti risalgono la catena fino agli autori delle opere di *training*. Se invece l'apporto dell'utente è significativo, l'*output* è il risultato di una combinazione tra l'apporto dell'utente e quello del sistema, sicché l'utente può essere considerato coautore, non autore esclusivo.

Si consideri anzitutto l'ipotesi di un *prompt* banale, quale «Scrivi una storia d'amore». In tal caso, l'*output* è determinato quasi interamente dai *pattern* appresi nel *training*: l'utente non è autore, e i diritti, semmai, spettano agli autori delle opere di *training* e agli altri soggetti della filiera.

Diversa è l'ipotesi del *prompt elaborato*, quale «Scrivi una storia d'amore ambientata in una Venezia distopica del 2147, con protagonisti un'archivista di memorie e un contrabbandiere di libri, nello stile di Calvino ma con dialoghi alla Tarantino». Qui l'apporto dell'utente è sostanziale: egli ha fornito ambientazione, personaggi, istruzioni stilistiche. L'*output* è opera derivata cui l'utente contribuisce in misura significativa, potendo rivendicare una quota di coautorialità.

Vi è infine l'ipotesi del *fine-tuning* (o *LoRA*), in cui l'utente non si limita a formulare *prompt*, ma modifica i pesi del modello mediante addestramento aggiuntivo su un proprio *dataset*. In questo caso l'*output* del modello dipende in maniera molto più sostanziale dall'apporto dell'utente. Quest'ultimo potrà vantare una quota maggiore di diritti, che saranno comunque vincolati ai diritti sul modello base, che a sua volta incorpora i diritti sulle opere di *training* originarie. Si configura dunque una catena di derivazione a più livelli, che attualmente è gestibile solo con altrettanti livelli di licenze.

7. Conclusioni

7.1 Sintesi ricostruttiva

L'analisi condotta consente di trarre le seguenti conclusioni.

In primo luogo, i pesi di un modello di apprendimento automatico non sono riconducibili né alla categoria del *software* (difettando del requisito delle istruzioni eseguibili), né a quella delle banche dati (difettando del requisito degli elementi indipendenti individualmente accessibili), né a quella delle opere dell'ingegno originarie (difettando di un singolo autore umano identificabile), né a quella dei segreti industriali (che possono al massimo essere una tutela residuale, che non riconosce i diritti degli autori a monte).

In secondo luogo, i pesi possono essere qualificati come *opera composta* (risultante dalla combinazione inscindibile di apporti creativi eterogenei, secondo il modello dell'opera cinematografica) e *derivata* (in quanto trasformazione delle opere del *dataset* di *training* in un nuovo formato di rappresentazione). Tale duplice qualificazione, lungi dall'essere contraddittoria, coglie due dimensioni complementari del fenomeno: la dimensione sincronica (chi contribuisce cosa) e la dimensione diacronica (come i diritti si propagano attraverso le trasformazioni).

Tale qualificazione, tuttavia, è possibile solo quando il *training* verte su opere dell'ingegno di carattere creativo. Modelli addestrati esclusivamente su dati empirici (rilevazioni fisiche, serie statistiche, dati comportamentali) esulano dal perimetro del diritto d'autore, potendo semmai beneficiare di tutela brevettuale o di protezione come segreto industriale.

In quarto luogo, l'eccezione TDM introdotta dalla Direttiva 2019/790 legittima l'atto di *training* in quanto riproduzione ed estrazione ai fini dell'analisi, ma non incide sulla natura giuridica del risultato. I diritti degli autori delle opere di *training* sopravvivono nei pesi ai sensi dell'art. 4 LDA.

Infine, l'*AI Company* assume una posizione analoga al produttore cinematografico ai sensi dell'art. 45 LDA: le compete l'esercizio dei diritti di utilizzazione economica del modello, fermi restando i diritti (patrimoniali e morali) degli autori a monte.

La ricostruzione proposta ha conseguenze operative immediate.

Gli autori delle opere alla base del *training* possono vantare diritti sullo sfruttamento commerciale dei pesi che sopravvivono alla trasformazione delle opere originali in vettori. Tali diritti possono essere esercitati tramite *collecting societies*, sul modello GEMA/SIAE.

L'*AI Company* deve acquisire i diritti necessari tramite licenze e *opt-in* o verificare l'applicabilità dell'eccezione TDM (assenza di *opt-out*). È inoltre soggetta agli obblighi di trasparenza AI Act.

L'utente finale, infine, a seconda dell'entità del proprio apporto, può essere coautore dell'*output* (nel caso di *prompt* elaborati o *fine-tuning*) ovvero mero fruitore. Nel primo caso, i suoi diritti coesistono con quelli degli altri soggetti della filiera.

7.2 Intervento legislativo: necessità o scelta?

Si è sostenuto che le categorie esistenti del diritto d'autore sono sufficienti a inquadrare il fenomeno. Ciò non significa che un intervento legislativo sia superfluo, ma che la sua funzione sarebbe diversa da quella di «colmare lacune».

Il legislatore potrebbe intervenire per operare scelte politiche redistributive: definire quote minime di compenso, stabilire meccanismi di gestione collettiva obbligatoria, imporre obblighi di trasparenza più stringenti. Esattamente come avvenne per il cinema (artt. 44-50 LDA) e per la fonografia (artt. 72 ss. LDA), dove il legislatore non «inventò» nuove categorie, ma allocò diritti e compensi tra i soggetti di filiere emergenti.

In assenza di intervento legislativo, la giurisprudenza dovrà sviluppare i principi qui delineati. Le corti europee saranno chiamate a precisare i confini della tutela: quali atti costituiscono TDM lecito, quando l'*opt-out* è efficace, come si calcolano i compensi. La sentenza *GEMA v. OpenAI*, pur con i limiti evidenziati, raggiunge la giusta conclusione.

Il diritto d'autore ha sempre saputo adattarsi alle trasformazioni tecnologiche: dalla stampa alla fotografia, dal fonografo al cinema, dalla radio a Internet. L'Intelligenza Artificiale generativa rappresenta solo l'ennesima sfida. Le categorie forgiate in tre secoli di evoluzione normativa offrono strumenti concettuali robusti. Sta agli interpreti applicarli con rigore tecnico e sensibilità sistematica, resistendo alla tentazione di invocare vuoti normativi là dove le norme esistono e attendono solo di essere correttamente applicate.

Bibliografia

Normativa: L. 22 aprile 1941, n. 633; D.Lgs. 10 febbraio 2005, n. 30; D.Lgs. 8 novembre 2021, n. 177; Dir. 96/9/CE; Dir. 2001/29/CE; Dir. 2009/24/CE; Dir. (UE) 2019/790; Reg. (UE) 2024/1689.

Dottrina: E. Piola Caselli, *Trattato del diritto d'autore*, Torino, 1927; A. Vaswani et al., *Attention Is All You Need*, NeurIPS, 2017.

Giurisprudenza: Cass. civ., ord. 16 gennaio 2023, n. 1107; Trib. Torino, 17 ottobre 1983, in *Dir. aut.*, 1984, p. 57 ss.; CGUE, C-444/02, *Fixtures Marketing*; LG München I, 28 novembre 2024, 42 O 14139/24 (*GEMA v. OpenAI*).