



# *Text and Data Mining (TDM) e violazioni del *copyright*: l’addestramento della GenAI e il diritto umano alla scienza alla luce di alcuni elementi di prassi recente*

**Angela Maria Gallo\***

SOMMARIO: 1. Introduzione. – 2. Le eccezioni di TDM nella direttiva CDSM e il loro recepimento nell’ordinamento giuridico tedesco. – 3. La presunta violazione della normativa tedesca sul *copyright* da parte della Large-scale Artificial Intelligence Open Network (LAION). – 4. Il fondamento giuridico delle eccezioni di TDM in alcune norme di diritto internazionale e dell’Unione europea. – 5. Alcuni ulteriori elementi di prassi in materia di TDM. – 6. Conclusioni.

## **1. Introduzione**

Le eccezioni di *Text and Data Mining* (TDM) previste dagli articoli 3 e 4 della direttiva 2019/790 sul diritto d’autore e sui diritti connessi nel mercato unico digitale<sup>1</sup> (Copyright in the Digital Single Market –

---

\* Dottoranda di ricerca in Diritto pubblico, comparato ed internazionale, curriculum Ordine internazionale e diritti umani presso Sapienza – Università di Roma.

<sup>1</sup> Per un approfondimento sugli obiettivi raggiunti e sui fallimenti della direttiva, C SGANGA, *Dall’armonizzazione alla frammentazione: obiettivi e fallimenti della Direttiva Copyright (2019/790/UE) in materia di ricerca, educazione e accesso al patrimonio culturale*, in *Rivista italiana di informatica e diritto*, 2023, pp. 47-66.

CDSM), rappresentano la risposta del legislatore europeo all'esigenza di garantire, in particolare ai soggetti che svolgono attività di ricerca, la possibilità di accedere ad un'ampia quantità di informazioni senza incorrere nei limiti imposti dalla normativa sulla proprietà intellettuale. Si tratta di una tecnologia integrata di elaborazione del linguaggio naturale (NLP), di classificazione di *pattern* e di apprendimento automatico (ML), che consiste in un processo computazionale di analisi e modellazione di contenuti testuali non strutturati in linguaggio naturale<sup>2</sup>.

Orbene, l'art. 3 della direttiva CDSM consente agli enti di ricerca e agli istituti di tutela del patrimonio culturale di effettuare attività di riproduzione ed estrazione di opere coperte da *copyright*; mentre, l'art. 4 consente attività analoghe a fini commerciali solo se i titolari dei diritti d'autore non hanno espresso una riserva in merito (diritto di *opt-out*). Ambedue le eccezioni si applicano esclusivamente nei casi in cui i soggetti che eseguono il TDM dispongono di un accesso legale alle opere.

Tuttavia, dal momento che l'addestramento dell'intelligenza artificiale si fonda, come il TDM, su una analisi automatizzata dei dati «avente lo scopo di generare informazioni inclusi, a titolo non esaustivo, modelli, tendenze e correlazioni<sup>3</sup>», alcuni fornitori di modelli di IA potrebbero invocare l'applicazione, a seconda dei casi, di una delle due eccezioni previste dalla direttiva 2019/790, per giustificare l'utilizzo, senza licenza o autorizzazione, di opere protette da *copyright* ai fini dell'addestramento<sup>4</sup>.

---

<sup>2</sup> Per una definizione esaustiva in materia, si veda, fra gli altri, C. ZONG, R. XIA, J. ZHANG, *Text and data Mining*, Singapore, 2021, p. 2 ss.

<sup>3</sup> Direttiva (UE) 2019/790 del Parlamento europeo e del Consiglio, del 17 aprile 2019, sul diritto d'autore e sui diritti connessi nel mercato unico digitale e che modifica le direttive 96/9/CE e 2001/29/CE, art. 2.

<sup>4</sup> Per un approfondimento sul tema dell'addestramento della GenAI attraverso il TDM: C. GEIGER, G. FROSIO, O. BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU (October 17, 2019)*, in C. SAIZ GARCÍA, R. EVANGELIO LLORCA (dirs.), *Propiedad intelectual y mercado único digital europeo*, Valencia, 2019, pp. 27-71; C. GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the Failed Text and Data Mining Limitations in the CSDM Directive*, in *Research Paper Series*, n. 66, 2021, pp. 383-394.

Una fattispecie di questo tipo è stata di recente decisa dal Tribunale di Amburgo nella causa *R. Kneschke c. Large-scale Artificial Intelligence Open Network (LAION)*, con la sentenza n. LG Hamburg 310 O 227/23 del 27 settembre 2024<sup>5</sup>.

Nello specifico, i giudici tedeschi erano chiamati a stabilire se le attività di estrazione ed analisi dei dati, poste in essere a fini di ricerca, potessero integrare una violazione del diritto d'autore qualora tra i dati utilizzati per addestrare l'IA fosse compresa un'opera coperta da *copyright*.

Sebbene la pronuncia non abbia risolto tutte le questioni emerse nel corso del procedimento, ha chiarito alcuni profili relativi all'applicabilità delle eccezioni di TDM all'addestramento dell'IA, offrendo al contempo spunti di riflessione sulla nozione stessa di attività di ricerca.

Di seguito, dopo una breve ricognizione del quadro normativo sovranazionale e tedesco in materia, si analizzeranno la sentenza e altri elementi di prassi rilevanti, tenendo conto delle linee di tendenza attualmente seguite dagli Stati e delle norme di diritto internazionale e dell'Unione europea in proposito. L'obiettivo è verificare se le eccezioni di *text and data mining* – in quanto strumenti funzionali alla ricerca scientifica e, al contempo, tecnologia alla base dell'addestramento dell'IA, oggi impiegata come mezzo di informazione – possano trovare fondamento nel diritto alla scienza, inteso quale diritto alla diffusione della conoscenza.

A tal fine, si rifletterà anche sulla possibilità che tale diritto costituisca un limite ai diritti di proprietà intellettuale riconosciuti e tutelati a livello internazionale.

## ***2. Le eccezioni di TDM nella direttiva CDSM e il loro recepimento nell'ordinamento giuridico tedesco***

Come noto, prima della direttiva UE 2019/790 sul diritto d'autore, il quadro giuridico europeo in materia di riproduzioni di opere coperte da *copyright* era piuttosto rigido, in quanto la tutela autoriale era estesa

---

<sup>5</sup> Tribunale di Amburgo 27 settembre 2024, n. LG Hamburg 310 O 227/23, disponibile sul sito [www.openfuture.eu](http://www.openfuture.eu).

anche a banche dati e software, imponendo divieti di riproduzione, temporanei o permanenti, che di fatto escludevano le attività di *text and data mining*<sup>6</sup>.

Successivamente, a seguito dell'evoluzione della tecnologia e della crescente possibilità di utilizzare il materiale pubblicato *online*, il legislatore europeo ha riconosciuto la necessità di introdurre eccezioni specifiche per il TDM, chiarendo i casi in cui tale attività sia lecita e quando, invece, possa integrare una violazione del diritto d'autore<sup>7</sup>.

Orbene, come abbiamo visto, gli articoli 3 e 4 della direttiva CDSM disciplinano due distinte eccezioni: la prima riguarda l'uso di dati estratti a fini di ricerca scientifica; la seconda, invece, consente il TDM per fini generali, inclusi quelli commerciali, in deroga alla normativa sul diritto d'autore. Tuttavia, di recente, l'Unione europea nel fornire un'interpretazione sistematica delle due eccezioni<sup>8</sup>, ha sottolineato che quelle previste dall'art. 3 sono incondizionate: i titolari dei diritti non possono opporsi al loro utilizzo, neppure esercitando il diritto di *opt-out*. Va detto, però, che esse sono riservate agli organismi di ricerca e agli istituti di tutela del patrimonio culturale<sup>9</sup> che operano per finalità di interesse pubblico, anche nell'ambito di partenariati con soggetti privati, purché questi ultimi non godano di un accesso preferenziale ai risultati.

---

<sup>6</sup> M. ZANCAN, *La nuova direttiva sul diritto d'autore e sui diritti connessi nel mercato unico digitale*, in *MediaLaws*, disponibile sul sito [www.medialaws.eu](http://www.medialaws.eu); V. BOCCHETTI, *La direttiva dell'Unione europea sul diritto d'autore*, in *DUE*, 2019; R. DUCATO, A. STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, disponibile sul sito [www.uclouvain.be](http://www.uclouvain.be). Per un quadro giuridico UE sul *text and data mining*, M. BOTTIS, M. PAPADOPOULOS, C. ZAMPAKOLAS, P. GANATSIU, *Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit*, in *Erasmus Law Review*, 2019, p. 190 ss.

<sup>7</sup> Per un'analisi sulla previsione delle eccezioni di TDM nella direttiva *Copyright*, C. GEIGER, G. FROSIO, O. BULAYENKO, *The Exception for Text And Data Mining (TDM) in the Proposed Directive on Copyright in The Digital Single Market – Legal Aspects*, disponibile sul sito [www.europarl.europa.eu](http://www.europarl.europa.eu).

<sup>8</sup> UE, Commissione per la Giustizia, le Libertà civili e gli Affari interni, *Generative AI and Copyright. Training, Creation, Regulation*, p. 35, disponibile sul sito [www.europarl.europa.eu](http://www.europarl.europa.eu)

<sup>9</sup> Ai sensi dell'art. 2, par. 1 e 3 della direttiva CDSM un istituto di tutela del patrimonio culturale può essere una biblioteca o un museo accessibili al pubblico, un archivio o un istituto di tutela del patrimonio cinematografico o sonoro, mentre un organismo di ricerca può essere un ente senza scopo di lucro o un ente incaricato da uno Stato membro di svolgere una missione di ricerca di servizio pubblico.

Ancora, l'art. 3 consente l'archiviazione e la conservazione di copie delle opere e di altri materiali estratti, a condizione che sia garantito un livello di sicurezza adeguato in fase di memorizzazione, in modo da rendere possibile, in un secondo momento, la verifica dei risultati della ricerca. Ciò è importante, poiché la ricerca scientifica empirica richiede generalmente che i dati di ricerca restino disponibili a fini di corroborazione<sup>10</sup>. Inoltre, i titolari dei diritti, ossia quelli individuati dagli artt. 5, lettera a), e 7, par. 1, della direttiva 96/9/CE, dall'art. 2 della direttiva 2001/29/CE, e dall'art. 15, paragrafo 1, della direttiva 2019/790<sup>11</sup>, possono applicare misure idonee a garantire la sicurezza e l'integrità delle reti e delle banche dati contenenti le opere o altri materiali oggetto di analisi ed estrazione dei dati. A tal proposito, come specificato nel considerando n. 16 della direttiva CDSM, tali misure intendono evitare che richieste massive di accesso e di scaricamento delle opere compromettano la sicurezza e l'integrità dei sistemi. Tra gli strumenti di controllo che i titolari dei diritti d'autore possono adottare rientrano la verifica dell'indirizzo IP e le procedure di autenticazione degli utenti.

Le eccezioni previste dall'art. 4, invece, consentono attività di TDM a fini generali, inclusi quelli commerciali, e possono essere esercitate da qualsiasi soggetto, purché i titolari dei diritti d'autore non si siano avvalsi di una riserva di *opt-out*, condizione inserita su proposta della delegazione olandese in fase di discussione in seno al Gruppo di Lavoro del Consiglio.

Tale riserva deve essere espressa in modo appropriato, ossia mediante l'uso di strumenti leggibili da una macchina, compresi metadati e termini e condizioni di un sito web o di un servizio. Tuttavia, i servizi possono essere opportunamente riservati in altro modo, attraverso, ad esempio, accordi contrattuali o dichiarazioni unilaterali o, ancora, aggiungendo metadati di tipo robot.txt ai loro contenuti *online*<sup>12</sup>.

---

<sup>10</sup> P. BERNT HUGENHOLTZ, *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*, in *Kluwer Copyright Blog*, 24 July 2019, disponibile sul sito [www.legalblogs.wolterskluwer.com](http://www.legalblogs.wolterskluwer.com)

<sup>11</sup> Si tratta, rispettivamente, del diritto di riproduzione delle banche dati, del diritto di estrazione e/o reimpiego del costituente della banca dati, del diritto di riproduzione della direttiva *InfoSoc* e del diritto sulle pubblicazioni di carattere giornalistico.

<sup>12</sup> Considerando n. 18, direttiva (UE) 2019/790, cit.

In ogni caso, ambedue le eccezioni presuppongono un accesso legittimo alle opere protette da *copyright*, come chiarito dal considerando n. 14 della direttiva, secondo cui è legittimo l'accesso derivante da accordi commerciali o dalla libera pubblicazione *online* dei contenuti.

Orbene, le eccezioni in parola sono contemplate dagli artt. 60d e 44b del *Gesetz über Urheberrecht und verwandte Schutzrechte*<sup>13</sup> (UrhG, la legge tedesca sul *copyright*). Non si tratta propriamente di un recepimento “totale” della direttiva CDSM, almeno per quanto concerne le eccezioni di TDM a fini di ricerca scientifica, in quanto esse erano già state previste dall'ordinamento giuridico tedesco nel 2017, a seguito della decisione del legislatore di non attendere l'emanazione della direttiva 2019/790, al tempo incerta<sup>14</sup>; tuttavia l'art. 60d è stato in parte modificato per adattarlo ai nuovi requisiti previsti dalla direttiva. Al contrario, le eccezioni di TDM a fini commerciali rappresentano una previsione normativa di nuovo conio per l'ordinamento giuridico tedesco, in quanto frutto del recepimento della direttiva CDSM, avvenuta nel 2021<sup>15</sup>.

Ora, l'articolo 44b (1) dell'UrhG contiene la definizione giuridica di *text and data mining*, a cui si fa riferimento anche nell'articolo 60d della predetta legge, e adotta, in forma lievemente modificata, la definizione dell'articolo 2 (2) della direttiva CDSM<sup>16</sup>. In questo contesto, il *text and data mining* è inteso come «l'analisi automatizzata di singole o multiple opere digitali o digitalizzate al fine di estrarre informazioni, in particolare su modelli, tendenze e correlazioni<sup>17</sup>».

---

<sup>13</sup> Il testo del codice è interamente disponibile sul sito [www.gesetze-im-internet.de](http://www.gesetze-im-internet.de).

<sup>14</sup> G. SPINDLER, *Text- und Datamining im neuen Urheberrecht und in der europäischen Diskussion*, p. 273, disponibile sul sito [www.mohrsiebeck.com](http://www.mohrsiebeck.com); B. RAUE, *Free Flow of Data? The Friction Between the Commission's European Data Economy Initiative and the Proposed Directive on Copyright in the Digital Single Market*, in *IIC - International Review of Intellectual Property and Competition Law*, 2017, p. 656.

<sup>15</sup> Universität Bremen, *Sonderfall: Text und Data Mining*, disponibile sul sito [www.uni-bremen.de](http://www.uni-bremen.de).

<sup>16</sup> B. RAUE, *Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§ 44b, 60d UrhG)*, in *Journal of Copyright and Media Law*, 2021, p. 797.

<sup>17</sup> Art. 44b (1) del Gesetz über Urheberrecht und verwandte Schutzrechte.

Sia l'art. 44b che l'art. 60d dell'UrHG ammettono la digitalizzazione di informazioni analogiche a fini di text e data mining, tanto per i ricercatori quanto per i privati, ma escludono da queste eccezioni le attività poste in essere esclusivamente per la creazione di un archivio digitale.

Orbene, l'art. 44b consente le riproduzioni di opere ai fini di *text e data mining* eseguito a scopi commerciali, purché esse siano legalmente accessibili e l'autore non abbia espresso una riserva in proposito, che deve essere effettuata in un linguaggio leggibile dalla macchina. Inoltre, le predette attività sono consentite solo a condizione che siano richieste riproduzioni meramente temporanee nella memoria del computer.

La permanenza delle copie oltre il tempo necessario richiede la stipula di accordi di licenza con i titolari dei diritti, offrendo loro così una forma di remunerazione per l'uso<sup>18</sup>.

Infine, l'art. 44b autorizza le attività di *text and data mining* anche in relazione ai programmi per computer<sup>19</sup>.

L'art. 60d UrHG, invece, consente le attività di TDM di opere coperte da *copyright* per la ricerca scientifica, purché l'ente che effettua la ricerca *a)* persegua scopi non commerciali, *b)* reinvesta tutti i profitti nella ricerca scientifica; *c)* operi nell'ambito di un contratto riconosciuto dallo Stato.

L'analisi della normativa tedesca rivela alcuni profili interessanti.

Anzitutto, gli organismi di ricerca che collaborano con imprese private, le quali, però devono esercitare un'influenza determinante sull'ente di ricerca o possedere un accesso preferenziale ai risultati della ricerca scientifica, sono esclusi dalla lista dei beneficiari delle eccezioni di TDM, in antitesi con quanto disposto dal considerando n. 11 della direttiva CDSM, che prevede che gli organismi di ricerca e gli istituti di tutela del patrimonio beneficiano delle eccezioni di TDM anche quando le loro attività di ricerca siano svolte nel quadro di partenariati pubblico-privato. In questo modo, si riconosce la necessità avvertita dai predetti organismi ed istituti, spesso dettata dalla mancanza di fondi pubblici per finanziare l'uso di nuove tecnologie, di effettuare l'estrazione di testo e di dati utilizzando strumenti tecnologici

---

<sup>18</sup> B. RAUE, *Die Freistellung*, cit., p. 798.

<sup>19</sup> Art. 69a (4) del Gesetz über Urheberrecht und verwandte Schutzrechte.

all'avanguardia, messi a loro disposizione dai privati<sup>20</sup>. Ancora, molto interessante risulta la scelta del legislatore tedesco di interrompere l'accesso al pubblico ai risultati della ricerca una volta completata la fase di revisione della qualità della stessa, nonché di cancellare le copie quando non sono più necessarie per l'estrazione di testo e dati<sup>21</sup>. Ciò si discosta dal considerando n. 15 della direttiva CDSM, che ammette la possibilità di conservare le copie in un ambiente sicuro per consentire la corroborazione dei risultati anche al termine delle attività di TDM<sup>22</sup>.

### 3. *La presunta violazione della normativa tedesca sul copyright da parte della Large-scale Artificial Intelligence Open Network (LAION)*

Orbene, le norme fin qui illustrate trovano applicazione concreta nella sentenza *Kneschke c. LAION*, di cui di seguito riportiamo brevemente i fatti occorsi.

Un fotografo tedesco, Robert Kneschke, il 27 aprile 2023 ha depositato un ricorso presso il Tribunale di Amburgo in quanto sosteneva che la *Large-scale Artificial Intelligence Open Network* (LAION), un'organizzazione no profit dedita alla creazione di *dataset* utilizzati per l'addestramento dell'intelligenza artificiale, avesse violato le norme tedesche in materia di diritto d'autore, poiché aveva incluso nel dataset "LAION 5B" una fotografia da lui scattata, senza il suo consenso<sup>23</sup>. La fotografia, infatti, era stata affidata ad un'agenzia di foto stock, che la commercializzava per conto del fotografo sul suo sito *online*, nel quale era presente un *disclaimer* che riportava le condizioni d'uso delle fotografie ivi pubblicate. Dunque, un'anteprima della predetta fotografia, seppur in bassa risoluzione e con filigrana, era accessibile al pubblico sul sito web dell'agenzia.

In aggiunta, l'attore sosteneva che il predetto set di dati fosse stato successivamente utilizzato per l'addestramento di un'IA a fini

---

<sup>20</sup> Considerando n. 11 della direttiva (UE) 2019/790, cit. Per approfondimenti sull'argomento, E. ROSATI, *An EU Text And Data Mining Exception for the Few: Would It Make Sense?*, in *Journal of Intellectual Property Law & Practice*, 2018, p. 429 ss.

<sup>21</sup> Art. 44b del Gesetz über Urheberrecht und verwandte Schutzrechte.

<sup>22</sup> Considerando n. 15 della direttiva (UE) 2019/790, cit.

<sup>23</sup> *R. Kneschke c. Large-scale Artificial Intelligence Open Network (LAION)*, sopra citata.



commerciali (in particolar modo per l’addestramento di IA di proprietà di Stability AI<sup>24</sup>).

La convenuta, invece, asseriva che la fotografia fosse già presente in un *dataset* formato da oltre 5 miliardi di coppie immagine-testo, disponibile online sin dal 2008<sup>25</sup>, dal quale aveva estratto i dati per la creazione di LAION 5B, nell’ambito di una ricerca scientifica il cui oggetto di studio era la correlazione tra il contenuto dell’immagine e la descrizione di testo corrispondente. Infatti, il *dataset* da essa creato non includeva copie di opere, ma consisteva in una tabella contenente collegamenti ipertestuali a immagini o file immagine, pubblicamente disponibili su Internet, nonché un “testo alternativo” concernente informazioni sul contenuto dell’immagine<sup>26</sup>.

Dunque, l’attore, invocando la normativa tedesca sul diritto d’autore, chiedeva la cancellazione della fotografia dal *dataset* LAION 5B, nonché un risarcimento per il danno patito.

La pronuncia del Tribunale di Amburgo assume rilievo poiché chiarisce diversi aspetti giuridici centrali relativi alla relazione tra TDM e l’addestramento dell’IA. In primo luogo, i giudici hanno stabilito che *a)* le attività di TDM condotte ai fini dell’addestramento dell’intelligenza artificiale – quando utilizzano i dati come strumenti di apprendimento e non come forme di espressione artistica – non violano il diritto d’autore.

Infatti, l’estrazione e l’analisi dei dati mirano alla creazione di dataset destinati alla fase successiva di *training* dei modelli, senza determinare la produzione di opere concorrenti. Poiché l’*output* dell’IA non è prevedibile *ex ante*, non è possibile stabilire se e in quale misura possa riprodurre opere protette, né se arrechi un danno economico agli autori originari, dunque non è detto che sia idoneo a concorrere sul mercato con le opere utilizzate in fase di addestramento.

---

<sup>24</sup> Stability AI è un’impresa che si occupa della creazione di modelli di intelligenza artificiale, nota soprattutto per aver sviluppato *Stable Diffusion*, uno dei modelli *open source* più popolari per la generazione di immagini a partire da descrizioni testuali. Oltre alla generazione di immagini, l’azienda sviluppa altri modelli per generare diversi tipi di contenuti, tra cui video, musica e oggetti 3D.

<sup>25</sup> Si tratta del database creato dalla *Common Crawl Foundation*, disponibile sul sito [www.commoncrawl.org/overview](http://www.commoncrawl.org/overview)

<sup>26</sup> R. Kneschke c. *Large-scale Artificial Intelligence Open Network (LAION)*, sopra citata.

Ne consegue che l'attività di TDM, limitandosi alla fase di analisi e raccolta dei dati, non configura una violazione del *copyright*, in quanto non utilizza i dati raccolti come “opere espressive”<sup>27</sup>.

Il Tribunale ha poi affrontato b) il tema delle misure tecnologiche di protezione (*Technological Protection Measures* – TPM), previste dall'art. 44b UrhG, le quali consentono ai titolari di diritti d'autore di limitare l'uso delle proprie opere a fini commerciali, purché espresse in un linguaggio intellegibile per la macchina. Tradizionalmente, tale linguaggio è quello dei metadati o di specifici codici informatici (Java, SQL, XML<sup>28</sup>). Tuttavia, i giudici hanno osservato che l'evoluzione tecnologica dell'IA consente oggi la lettura di linguaggi naturali, che devono quindi ritenersi inclusi tra i “formati leggibili dalla macchina”<sup>29</sup>.

A sostegno di questa interpretazione, il Tribunale richiama l'art. 53, par. 1, lett. c), del regolamento (UE) 1689/2024 (AI Act), che impone ai fornitori di modelli per finalità generali di adottare una politica conforme alla normativa UE in materia di diritto d'autore, e di individuare e rispettare, anche attraverso tecnologie all'avanguardia, una riserva di diritti espressa a norma dell'articolo 4, paragrafo 3, della direttiva (UE) 2019/790. Ebbene, secondo i giuridici, la locuzione “tecnologia all'avanguardia” include inequivocabilmente le applicazioni di IA, che sono in grado di catturare il contenuto di un testo scritto in linguaggio naturale.

A tal proposito, riteniamo di poter concordare con il Tribunale, anche in ragione del fatto che l'IA esegue comandi che gli esseri umani scrivono in un linguaggio naturale. Inoltre, riprendendo *mutatis mutandis* quanto affermato altrove sul principio di neutralità tecnologica<sup>30</sup>, la norma parla di “formato leggibile alla macchina”,

---

<sup>27</sup> Ivi. Traduzione a nostra cura. In questo senso, T. W. DORNIS, S. STÖBER, *Copyright Law and Generative AI Training - Technological and Legal Foundations (Urheberrecht und Training generativer KI-Modelle - Technologische und juristische Grundlagen)*, Baden-Baden, 2024, p. 67 ss.

<sup>28</sup> M. D. PAPADOPOULOU, K. KOLOTOUROU, M. BOTTIS, *The Exception of Text and Data Mining from the Academic Libraries Standpoint*, in *Open Journal of Social Sciences*, vol. 9, n. 5, 2021, p. 514.

<sup>29</sup> *R. Kneschke c. Large-scale Artificial Intelligence Open Network (LAION)*, sopra citata, punto 4. Traduzione a nostra cura.

<sup>30</sup> G. M. RUOTOLO, *Scritti di diritto internazionale ed europeo dei dati*, Bari, 2021, p. 66 ss. Sul principio di neutralità tecnologica si veda anche V. SHADIKHODJAEV,

tuttavia, essendo mutata “la macchina” che deve leggere il linguaggio, la quale attualmente è dotata di una tecnologia più avanzata, la predetta locuzione, di cui all’art. 44b UrhG, ben può includere il linguaggio naturale tra quelli intelligibili dai sistemi di IA.

Un altro punto focale della sentenza è quello relativo alla correlazione tra l’attività di TDM e la ricerca scientifica. A tal proposito, i giudici hanno chiarito che c) una ricerca scientifica aspira ad acquisire nuove conoscenze e, per farlo, la fase di lavoro antecedente all’acquisizione assume grande importanza. Infatti, per la formulazione di conclusioni empiriche, è necessario partire da una raccolta di dati ed informazioni ben strutturata; analogamente, le attività di TDM sono prodromiche all’addestramento dell’IA per fini di ricerca e rientrano, quindi, nell’eccezione di cui all’art. 3 della direttiva CDSM.

In questo senso, i giudici chiariscono che anche la creazione di un set di dati come quello oggetto della fattispecie può essere considerata ricerca scientifica, in quanto, sebbene l’acquisizione di nuove conoscenze non sia ancora tangibile, essa rappresenta una fase di lavoro prodromica all’acquisizione successiva di nuove nozioni. Peraltro, il fatto che la LAION stesse conducendo una ricerca scientifica è pure confermato dalla pubblicazione gratuita del set di dati, reso disponibile ai ricercatori. Inoltre, va detto, che per i giudici tedeschi l’utilizzo del set di dati da parte di aziende commerciali per l’addestramento o l’ulteriore sviluppo dei loro sistemi di IA è irrilevante, perché anche la ricerca delle aziende commerciali, seppur privata e condotta per altri scopi, è ancora ricerca, pur non rientrando nelle eccezioni di cui all’art. 60d UrhG<sup>31</sup>. Infine, riprendendo quanto esplicitamente previsto nel considerando 42 della direttiva *InfoSoc* e il considerando n. 11 della direttiva CDSM, i giudici hanno chiarito che, per verificare «se la ricerca sia non commerciale, è rilevante solo il tipo concreto di attività scientifica, mentre l’organizzazione e il finanziamento dell’istituzione in cui si svolge la ricerca sono irrilevanti»<sup>32</sup>.

---

*Technological Neutrality and Regulation of Digital Trade: How Far Can We Go?*, in *EJIL*, vol. 32, n. 4, 2021, p. 1221 ss.

<sup>31</sup> R. Kneschke c. *Large-scale Artificial Intelligence Open Network (LAION)*, sopra citata., punto 3. Traduzione a nostra cura.

<sup>32</sup> *Ibidem*.

Orbene, riteniamo che una siffatta interpretazione della normativa UE in materia di *copyright* sia effettivamente conforme alla *ratio legis* sottesa alle norme sul *text and data mining*, che da ultimo è stata esplicitata anche in uno studio condotto dalla Commissione per la Giustizia, le Libertà civili e gli Affari interni, su richiesta della Commissione JURI del Parlamento europeo, nel quale si sottolinea che il *text and data mining* è una tecnica di analisi automatizzata necessaria per estrarre informazioni da grandi volumi di testo e dati, spesso a supporto della ricerca scientifica o dell'analisi empirica<sup>33</sup>.

Non si tratta, quindi, di una rielaborazione dei dati ma solo di osservazione e confronto degli stessi; dunque, fino a quando le attività poste in essere tanto da ricercatori quanto da privati si fermano all'analisi dei dati e alla loro riproduzione per identificare delle corrispondenze, non siamo in presenza di una violazione delle norme sul *copyright*, in quanto le predette attività non risultano in contrasto con lo sfruttamento normale dell'opera e non arrecano un ingiustificato pregiudizio agli interessi legittimi del titolare.

#### 4. *Il fondamento giuridico delle eccezioni di TDM in alcune norme di diritto internazionale e dell'Unione europea*

Ora, a noi pare che le eccezioni di TDM trovino fondamento giuridico in alcune norme programmatiche di diritto internazionale che tutelano il diritto alla scienza, e in particolare, l'art. 15 del Patto Internazionale sui Diritti Economici, Sociali e Culturali (ICESCR), nonché in altri strumenti di *soft law*.

Esso è inteso come il diritto a godere dei benefici del progresso scientifico e delle sue applicazioni, come chiarito dal Comitato sui Diritti Economici, Sociali e Culturali nel *General Comment* n. 21 del 2009<sup>34</sup>. In particolare, il predetto Comitato sostiene che il diritto alla scienza sia indissolubilmente legato a quello all'istruzione, in quanto ambedue sono prodromici alla partecipazione degli individui alla vita

---

<sup>33</sup> *Generative AI and Copyright. Training, Creation, Regulation*, cit., p. 43.

<sup>34</sup> United Nations, Committee on Economic, Social and Cultural Rights, General comment No. 21, *Right of everyone to take part in cultural life (art. 15, para. 1 (a), of the International Covenant on Economic, Social and Cultural Rights)*, p. 1, disponibile sul sito [www.refworld.org](http://www.refworld.org).

culturale. Inoltre, poiché nel predetto *General Comment*, il Comitato afferma che tutti i diritti relativi alla sfera culturale e intellettuale dell'individuo sono indissolubilmente intrecciati, riteniamo che la piena realizzazione di uno di questi diritti possa avvenire solo in un contesto in cui anche gli altri vengono parimenti tutelati. Dunque, gli Stati non possono esimersi dal tutelare il diritto alla scienza se tutelano quello all'istruzione, e, di conseguenza, non possono vietare l'uso di nuove tecnologie che faciliterebbero la piena realizzazione dei predetti diritti. Ancora, nel successivo *General Comment* n. 25 su "Science and Economic, Social and Cultural Rights"<sup>35</sup>, rilasciato dal Comitato sui Diritti Economici, Sociali e Culturali nel 2020, relativo all'art. 15 del Patto Internazionale sui Diritti Economici, Sociali e Culturali, si precisa che i benefici della scienza cui fa riferimento il predetto articolo attengono sia ai risultati materiali delle ricerche scientifiche e delle loro applicazioni, sia alla diffusione delle conoscenze scientifiche. Tra questi benefici rientra anche il ruolo della scienza nell'educare gli individui a sviluppare le loro capacità critiche, al fine di renderli pienamente partecipi di una società democratica<sup>36</sup>.

Anche nel diritto dell'Unione europea è possibile individuare il fondamento giuridico delle eccezioni di TDM, a prescindere dalle disposizioni di diritto derivato fin qui analizzate. Infatti, l'art. 13 CDFUE sancisce la libertà delle arti e della scienza, compresa quella accademica, che derivano dalla libertà di pensiero ed espressione, parimenti tutelate dalla Carta<sup>37</sup>. Si tratta di un *unicum* nel contesto normativo di riferimento, in quanto il diritto alla scienza non è contemplato dalla Convenzione europea dei diritti dell'uomo<sup>38</sup> (CEDU), a cui, come noto, il catalogo di diritti della CDFUE si allinea, così come si evince anche dalle Dichiarazioni relative alla Carta dei diritti fondamentali dell'Unione europea, contenute all'interno del

<sup>35</sup> United Nations, Committee on Economic, Social and Cultural Rights, General comment No. 25 (2020) on science and economic, social and cultural rights (article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights), p. 2, disponibile sul sito [www.digitallibrary.un.org](http://www.digitallibrary.un.org).

<sup>36</sup> V. ZAMBRANO, *Il diritto a godere dei benefici del progresso scientifico: profili problematici di un diritto ancora in cerca di effettività*, in *CI*, n. 3, 2020, p. 220.

<sup>37</sup> Commentary of the Charter of Fundamental Rights of the European Union, 2006, p. 132, disponibile sul sito [www.sites.uclouvain.be](http://www.sites.uclouvain.be).

<sup>38</sup> C. P. R. ROMANO, A. BOGGIO, *The human right to science*, New York, 2004, p. 385.

Trattato di Lisbona del 2009<sup>39</sup>. Tuttavia, sebbene solo indirettamente, poiché non esplicitamente contemplato dalla norma, riteniamo che, nella tutela prevista dall'art. 13 CDFUE sia ricompreso anche il diritto di diffondere la conoscenza. Ciò è deducibile dal fatto che, in alcuni strumenti giuridici adottati dall'UE in materia di ricerca scientifica, si fa spesso riferimento ai concetti di scienza aperta e di libertà della ricerca scientifica. In particolar modo, nel reg. 2021/695, che istituisce il programma quadro di ricerca e innovazione Horizon Europe, oltre ad un esplicito riferimento al rispetto dei diritti e delle libertà riconosciute dalla CDFUE, (art 13 compreso), si chiarisce che «l'accesso aperto alle pubblicazioni scientifiche e ai dati della ricerca, nonché la diffusione e lo sfruttamento ottimali delle conoscenze, offrono la possibilità di migliorare la qualità, l'impatto e i benefici della scienza. Sono anche in grado di accelerare il progresso delle conoscenze rendendole più attendibili, efficienti e accurate, più facilmente comprese dalla società e idonee a rispondere alle sfide sociali<sup>40</sup>».

Ancora, nella risoluzione del Parlamento europeo del 17 gennaio 2024, recante raccomandazioni alla Commissione concernenti la promozione della libertà della ricerca scientifica nell'UE<sup>41</sup>, al punto 5, si specifica che «la libertà della ricerca scientifica implica il diritto dei singoli ricercatori [...] di raccogliere materiale empirico e di utilizzare metodi di ricerca scientifica validi, [...] di proporre nuove idee e teorie e diffonderle liberamente». Peraltro, al punto 7, lettera d), si sottolinea che, tra i diritti dei ricercatori è incluso «il diritto di pubblicare, condividere, diffondere e comunicare apertamente, a livello sia interno che esterno, i risultati e i dati delle loro ricerche». Dunque, da quanto appena affermato deduciamo che anche nel quadro giuridico UE il

---

<sup>39</sup> Per ulteriori approfondimenti sui rapporti tra CEDU e CDFUE, E. CANNIZZARO, *La cornice europea*, in AA.VV., *Diritti della persona e nuove sfide del processo penale: atti del 32° Convegno nazionale*, Salerno, 25-27 ottobre 2018, Milano, 2019, p. 38; G. BRONZINI, *La CEDU e la Carta dei diritti fondamentali: overlap e distinzioni*, in *Questioni di giustizia*, 2019.

<sup>40</sup> Regolamento (UE) 2021/695 del Parlamento europeo e del Consiglio, del 28 aprile 2021, che istituisce il programma quadro di ricerca e innovazione Orizzonte Europa e ne stabilisce le norme di partecipazione e diffusione, e che abroga i regolamenti (UE) n. 1290/2013 e (UE) n. 1291/2013, considerando n. 8.

<sup>41</sup> Risoluzione del Parlamento europeo, del 17 gennaio 2024, recante raccomandazioni alla Commissione concernenti la promozione della libertà della ricerca scientifica nell'UE, disponibile sul sito [www.eur-lex.europa.eu](http://www.eur-lex.europa.eu).

diritto alla scienza implica la diffusione della conoscenza, poiché essa contribuisce «allo sviluppo e al miglioramento delle condizioni di vita delle persone»<sup>42</sup>.

Tuttavia, come noto, la CFUE contempla tra i diritti fondamentali anche quello alla proprietà, dal quale deriva la protezione della proprietà intellettuale<sup>43</sup>. Dunque, una previsione di siffatta portata suggerisce che il diritto alla scienza e quello alla proprietà intellettuale vadano necessariamente bilanciati. A tal proposito, la Corte di Giustizia dell'UE ha più volte sottolineato come, sebbene il diritto di proprietà intellettuale rientri nel catalogo di diritti sanciti dalla CEDU, in particolar modo dall'art. 1 del relativo Protocollo addizionale, cui la CDFUE si allinea, è pure vero che i diritti di proprietà intellettuale non vanno concepiti come assoluti e che, quindi, vanno bilanciati tenendo conto dei diritti altrui, che, nella fattispecie, potrebbero subire lesioni<sup>44</sup>. In particolar modo, la Corte, dinanzi ad un conflitto fra due diritti fondamentali, nessuno dei quali di natura subordinata al restante, effettua un bilanciamento tenendo conto degli obblighi derivanti dalla tutela degli stessi e dell'interesse sotteso alla necessità di limitare l'uno o l'altro diritto<sup>45</sup>. Quanto appena affermato non è rinvenibile in una sola pronuncia della Corte di giustizia, bensì nel suo orientamento giurisprudenziale, ricostruibile attraverso l'analisi di molteplici

<sup>42</sup> *Ivi*, considerando F.

<sup>43</sup> CDFUE, art. 17: «Ogni persona ha il diritto di godere della proprietà dei beni che ha acquisito legalmente, di usarli, di disporne e di lasciarli in eredità [...] 2. La proprietà intellettuale è protetta». Per approfondimenti, C. GEIGER, *Intellectual Property Shall Be Protected!? Article 17 (2) of the Charter of Fundamental Rights of the European Union: a Mysterious Provision with an Unclear Scope*, in *European Intellectual Property Review (EIPR)*, 2009, p. 113; M. HUSOVEC, *The Essence of Intellectual Property Rights Under Article 17(2) of the EU Charter*, in *GLJ*, 2019, pp. 840-863; J. GRIFFITHS, L. McDONAGH, *Fundamental Rights and European IP Law – the case of Art 17(2) of the EU Charter*, in C. GEIGER (ed.), *Constructing European Intellectual Property Achievements and New Perspectives*, Cheltenham, 2013, pp. 75-93.

<sup>44</sup> Corte giust. 12 settembre 2006, C-479/04, *Laserdisken/Kulturministeriet*, punto 65; 29 gennaio 2008, C-275/06, *Promusicae/Telefónica*, punti 62-68. Per un approfondimento sul bilanciamento tra diritti di proprietà intellettuale e diritti fondamentali, A. SCENNA, *Il bilanciamento tra i diritti d'autore e i diritti fondamentali alla luce della direttiva UE 2019/790. Note alla sentenza della Corte di giustizia europea*, in *DPCE online*, 2022.

<sup>45</sup> Corte giust. 27 marzo 2014, C-314/12, *Telekabel/Constantin e Wega*, punti 46 e 63; 16 febbraio 2012, C-360/10, *SABAM/Netlog*, punti 42-43; 19 aprile 2012, C-461/10, *Bonnier Audio/Perfect Communication Sweden*, punto 56.

sentenze in cui la Corte ha dovuto affrontare il tema del bilanciamento tra il diritto alla proprietà intellettuale e gli altri diritti fondamentali<sup>46</sup>.

Ora, appurato il fondamento giuridico internazionale ed europeo del diritto alla scienza, e volendo aderire a un'impostazione secondo cui il diritto alla scienza persegue il fine ultimo della diffusione della conoscenza, intesa come bene pubblico<sup>47</sup>, riteniamo che questo approccio comporti inevitabilmente una limitazione del diritto internazionale della proprietà intellettuale<sup>48</sup>.

Inoltre, consapevoli del fatto che le tecnologie dell'informazione mutano repentinamente e che non di rado nuovi strumenti tecnologici diventino il mezzo attraverso il quale la conoscenza è veicolata, riteniamo di poter affermare che anche l'intelligenza artificiale sia attualmente diventata un mezzo veicolare di conoscenza.

Dunque, poiché la ricerca scientifica mira all'acquisizione di nuove conoscenze, e l'intelligenza artificiale può essere un mezzo attraverso il quale acquisirle, le attività di TDM – in quanto collaterali alla creazione di *dataset* per l'apprendimento dell'IA – quando fatte oggetto di eccezioni nella prassi statale, previste a discapito della tutela della proprietà intellettuale, potrebbero esser ritenute misure positive volte ad ottemperare agli obblighi derivanti dall'art. 15 del Patto Internazionale sui Diritti Economici, Sociali e Culturali.

Infine, sempre in ordine alle limitazioni della proprietà intellettuale, giova ricordare, che anche la Convenzione di Berna per la protezione delle opere letterarie e artistiche, all'art. 9, prevede eccezioni ai diritti esclusivi, e quindi riserva agli Stati parte la facoltà di permettere la riproduzione delle opere coperte dal diritto d'autore *a)* in casi speciali; *b)* purché non arrechi alcun tipo di danno allo sfruttamento normale dell'opera; e *c)* non causi un pregiudizio ingiustificato ai legittimi interessi dell'autore.

---

<sup>46</sup> Per un approfondimento in tema di orientamento giurisprudenziale della Corte di giustizia e diritti di proprietà intellettuale, M. RICOLFI, *L'influenza del diritto europeo sul diritto commerciale italiano: valori, principi, interessi*, disponibile sul sito [www.orizzontideldirittocommerciale.it](http://www.orizzontideldirittocommerciale.it).

<sup>47</sup> Per un approfondimento della scienza come bene pubblico globale, International Science Council, *Science as a global public good*, disponibile sul sito [www.it.council.science.com](http://www.it.council.science.com).

<sup>48</sup> G. M. RUOTOLO, *Right to Science and Open Access to Legal Knowledge in International and European Law*, in G. PERUGINELLI, S. FARO, (eds.), *Knowledge of The Law in The Big Data Age*, Amsterdam, 2019, p. 105.



Questo test, noto con il nome di “Berne Three Step Test” è stato ripreso sia da strumenti di diritto internazionale (TRIPs, Trattato sul *Copyright* WIPO e Trattato sulle Esecuzioni e sui Fonogrammi) sia da strumenti regionali<sup>49</sup>; si tratta di un aspetto che non è sfuggito ai giudici tedeschi, i quali hanno effettuato il test, così come previsto dall’art. 7, par. 2 della direttiva CDSM, giungendo alla conclusione che l’utilizzo della fotografia da parte della LAION non pregiudica lo sfruttamento dell’opera da parte dell’autore<sup>50</sup>.

Il Tribunale, peraltro, ha letto le eccezioni di TDM come espressione della volontà del legislatore europeo di diffondere la conoscenza attraverso i dati estraibili dalle opere coperte da *copyright*, facendo presente che «poiché l’uso di conoscenze ottenute mediante l’estrazione di testi e dati per addestrare reti neurali artificiali, che possono poi competere con gli autori, non può mai essere escluso in caso di dubbio sulla base degli attuali sviluppi tecnologici, la tesi opposta richiederebbe in ultima analisi addirittura che l’estrazione di testi e dati ai sensi dell’articolo 44b UrhG sia vietata in toto; tuttavia, un tale completo superamento della disposizione di limitazione sarebbe ovviamente contrario all’intenzione legislativa e non può quindi rappresentare un risultato interpretativo praticabile»<sup>51</sup>.

### 5. Alcuni ulteriori elementi di prassi in materia di TDM

Dunque, le eccezioni di TDM alimentano ulteriormente il dibattito sulla questione della diffusione della conoscenza a scapito della tutela della proprietà intellettuale, un problema che negli ultimi tempi è diventato ancora più evidente a seguito della diffusione dell’IA. Di seguito faremo cenno ad alcuni ulteriori elementi di prassi, mettendo a confronto diversi ordinamenti giuridici in cui è consentita la fruizione

<sup>49</sup> Sull’argomento, fra gli altri, si vedano: C. GEIGER, D. GERVAIS, M. R. F. SENFTLEBEN, *The Three-Step Test Revisited: How to Use the Test’s Flexibility in National Copyright Law*, in *American University International Law Review*, 2014 p. 583 ss.

<sup>50</sup> P. WYMEERSCH, *EU Copyright Exceptions and Limitations and the Three-Step Test: One Step Forward, Two Steps Back*, in *GRUR International*, vol. 72, 2023, pp. 631-642.

<sup>51</sup> *R. Kneschke c. Large-scale Artificial Intelligence Open Network (LAION)*, sopra citata, punto 2, lettera c. Traduzione a nostra cura.

di opere protette attraverso le attività di TDM anche a fini commerciali, come il Giappone, e ordinamenti in cui le predette attività possono essere eseguite solo in presenza di determinate condizioni, come gli Stati Uniti e l'Unione europea, così da verificare se l'ordinamento giuridico UE stia facendo registrare un mutamento di paradigma riguardo ai limiti imposti alla tutela della proprietà intellettuale.

Le attività di TDM, come noto, sono diventate di grande rilevanza nella prassi degli Stati: ne è un esempio il Giappone, che ha introdotto, nella legge di riforma sul *copyright*, entrata in vigore il 1° gennaio 2019, l'art. 30-4, che consente l'uso e lo sfruttamento di un'opera protetta dal diritto d'autore, fatta salva l'ipotesi in cui ciò non comporti un danno ingiusto per l'autore. In particolar modo, la norma consente l'uso di opere coperte da *copyright* nei casi in cui: «(i) l'azione viene effettuata per testare o mettere in pratica una tecnologia connessa alla registrazione di suoni, immagini di un'opera o ad altri tipi di sfruttamento; (ii) se viene utilizzata per l'analisi dei dati (vale a dire l'estrazione, il confronto, la classificazione o altre analisi statistiche della lingua, dei suoni, delle immagini o di altri dati elementari di un gran numero di opere o di un grande volume di altri dati di questo tipo; (iii) se è sfruttata nel corso dell'elaborazione di dati informatici o altrimenti sfruttata in un modo che non implica che ciò che è espresso nell'opera sia percepito dai sensi umani (per le opere di programmazione informatica, tale sfruttamento esclude l'esecuzione dell'opera su un computer), al di là di quanto stabilito nei due punti precedenti»<sup>52</sup>. Dunque, il Giappone consente l'estrazione e l'analisi dei dati di opere coperte dal *copyright* senza il consenso dei titolari del diritto d'autore, purché non siano violati i loro interessi. A tal proposito, l'Agenzia giapponese per gli Affari Culturali (ACA) ha ulteriormente chiarito nel suo *Approach to AI and Copyright Report*, adottato il 23 gennaio 2024, che la norma va interpretata tenendo in considerazione le finalità per le quali è effettuata l'attività di TDM: se essa consiste esclusivamente nella raccolta e nell'analisi dei dati, utili ai fini del training dell'IA, allora sarà considerata legittima ai sensi dell'art 30-4 del *Copyright Law of Japan*; al contrario, se si pone come obiettivo

---

<sup>52</sup> CRIC, *Copyright Law of Japan*, disponibile sul sito [www.cric.or.jp](http://www.cric.or.jp). Traduzione a nostra cura.

l'uso dell'espressione artistica dell'autore per generare, a sua volta, attraverso l'IA, un *output* simile, allora il suo utilizzo non sarà consentito, a meno che il fornitore del modello di IA non abbia ottenuto una licenza a tal fine<sup>53</sup>.

Ciò che risulta interessante dall'analisi di questa prassi è che i titolari di diritti d'autore non possono opporre riserve di *opt-out* per evitare che le proprie opere siano oggetto di attività di TDM, né è richiesto un accesso legittimo alle opere per procedere alle suddette attività.

Questa scelta del governo giapponese di aprire il più possibile alle attività di *text and data mining* anche nel contesto del *training* dell'intelligenza artificiale, è dettato dall'idea di fondo per cui un'opera protetta da rappresenti un mezzo attraverso il quale soddisfare un'esigenza intellettuale o emotiva di un soggetto, che è disposto a corrispondere una somma in denaro per appagare la propria esigenza. Dunque, il motivo per cui esiste una legge sul diritto d'autore è rinvenibile nella necessità di garantire all'autore di un'opera una remunerazione adeguata per averla creata. Di conseguenza, se lo sfruttamento di un'opera non è finalizzato né a goderne né a far sì che un altro soggetto ne goda per appagare un proprio desiderio intellettuale, non è necessario garantire all'autore la ricezione di un compenso per essa; pertanto, il diritto d'autore non deve coprire tale sfruttamento<sup>54</sup>. Questa percezione del diritto d'autore è attualmente unica a livello globale, ragion per cui il Giappone è considerato il paradiso dell'intelligenza artificiale, dove l'IA può essere addestrata senza incorrere in violazioni delle norme sul *copyright*.

A questa visione aperta del diritto d'autore e del conseguente sfruttamento delle opere da esso tutelate anche senza l'autorizzazione del loro creatore, si contrappone quella statunitense, più restrittiva di quella giapponese ma comunque più ampia di quella europea.

Infatti, gli USA, come noto, hanno da tempo consolidato, attraverso un precedente giurisprudenziale basato sulla libertà di espressione,

---

<sup>53</sup> K. TOSAKI, H. TAJIMA, C. KOMIYA, *Report on AI and Copyright Issues by Japanese Government*, April 2024, disponibile sul sito [www.noandt.com](http://www.noandt.com).

<sup>54</sup> T. UENO, *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication*, in *GRUR International*, 2021, p. 150.

sancita dal Primo Emendamento della Costituzione, il *fair use*, una dottrina volta a limitare i diritti d'autore in favore del libero sfruttamento delle opere coperte da *copyright* per attività di ricerca, d'insegnamento, di studio, d'informazione, di commento e di critica<sup>55</sup>. La predetta clausola può essere invocata solo a seguito della verifica della finalità dell'uso, della natura dell'opera, della quantità di materiale coperto da *copyright* utilizzato e dell'impatto dell'opera trasformativa sul mercato dell'originale. Ebbene, è proprio grazie a questa dottrina che, negli ultimi anni, gli Stati Uniti stanno sperimentando la legalizzazione delle attività di TDM.

A tal proposito, i giudici della Court of Appeals for the Second Circuit, nella causa n. 13-4829cv del 16 ottobre 2015, *Authors Guild, Inc. v. Google Inc.*<sup>56</sup>, hanno statuito che le attività di TDM poste in essere da Google per realizzare il progetto *Google Books*, rientrassero nel *fair use*, in quanto trasformative di opere già esistenti, delle quali venivano forniti frammenti, che potevano altresì indurre il lettore ad acquistare l'opera. Questa interpretazione è stata di recente confermata anche dalla Corte distrettuale della California del Nord, che, nell'ordinanza del 23 giugno 2025, emessa a seguito della decisione del caso *Bartz et al. v. Anthropic PBC*, ha statuito che la creazione di set di dati, composti da oltre 7 milioni di opere, necessarie per addestrare l'intelligenza artificiale Claude, rientra nel *fair use* in quanto rappresenta un uso trasformativo di opere coperte da *copyright*, oltre che un utilizzo volto alla diffusione di conoscenza e informazione. Il giudice paragona Claude a uno studente che legge libri per apprendere, ma dai quali non copia, in quanto restituisce *output* in cui non vi è traccia delle opere utilizzate per il suo apprendimento automatico. Infine, la stessa Corte ha recentemente deciso anche il caso *Kadrey et al. v. Meta Platforms* con un'ordinanza del 25 giugno 2025, nella quale ribadisce che l'addestramento del modello LLama di Meta, avvenuto a seguito di attività di *text and data mining*, poste in essere per la creazione di *dataset*, rientra nel *fair use* in quanto altamente trasformativo delle opere originali.

---

<sup>55</sup> Copyright Act of USA, Section 107, *Limitations on exclusive rights: Fair use*, [www.copyright.gov](http://www.copyright.gov).

<sup>56</sup> United States Court of Appeals for the Second Circuit, sentenza del 16 ottobre 2015, n. 13-4829-cv, [www.copyright.gov](http://www.copyright.gov).

Insomma, i giudici statunitensi individuano nel *text and data mining* un'attività prodromica alla creazione di banche dati, che rientrano nel *fair use* in quanto rappresentano un uso trasformativo di opere coperte da *copyright* e agevolano la diffusione dell'informazione e della conoscenza.

Diversamente dagli Stati Uniti e dal Giappone, l'Unione europea mantiene un approccio più prudente e orientato all'equilibrio tra innovazione e diritti d'autore. Infatti, il 23 gennaio 2025 l'Unione europea ha pubblicato un invito a presentare offerte per uno studio di fattibilità su un registro centrale delle deroghe, nell'ambito delle eccezioni di *text and data mining*, attraverso il quale vorrebbe cercare, assieme ai portatori di interessi, quali i titolari dei diritti dei vari settori creativi e gli sviluppatori di IA, una soluzione che concili il rispetto dei diritti dei creatori di contenuti e lo sviluppo tecnologico dell'IA (per il quale, lo ricordiamo, la stessa UE ha lanciato l'iniziativa InvestAI al fine di mobilitare 200 miliardi di euro di investimenti<sup>57</sup>). Un registro simile consentirebbe di monitorare le riserve espresse dagli autori, impedendo l'utilizzo non autorizzato delle loro opere per il *training* dell'IA a fini commerciali. Si tratta, tuttavia, di una misura che introduce un freno alle attività di TDM, limitandone la libertà d'uso a meno che non sia verificata l'assenza di una riserva.

Tuttavia, è ormai riconosciuto tanto dalle organizzazioni internazionali quanto dagli Stati che il TDM costituisce uno strumento essenziale per la ricerca scientifica, poiché il suo esercizio è prodromico e di grande importanza per il progresso scientifico. A tal proposito, l'Organizzazione Mondiale della Proprietà Intellettuale (OMPI), una delle agenzie dell'ONU, che nel 2023 ha lanciato un "Progetto pilota sul Text and Data Mining (TDM) a supporto della ricerca e dell'innovazione nelle università e in altre istituzioni orientate alla ricerca in Africa<sup>58</sup>", con il quale mira ad istruire le università e gli istituti di ricerca africani circa un uso consapevole del TDM, nel

---

<sup>57</sup> Commissione europea, *L'UE lancia l'iniziativa InvestAI per mobilitare 200 miliardi di euro di investimenti nell'intelligenza artificiale*, [www.digital-strategy.ec.europa.eu](http://www.digital-strategy.ec.europa.eu).

<sup>58</sup> WIPO, *Pilot Project on Text and Data Mining (TDM) to Support Research and Innovation in Universities and Other Research-Oriented Institutions in Africa*, disponibile sul sito [www.dacatalogue.wipo.int](http://www.dacatalogue.wipo.int).

rispetto dei trattati sulla proprietà intellettuale, in modo da favorire anche l'uso di strumenti di IA e rafforzare le capacità del personale di università e istituti di ricerca in Africa.

Dunque, la prassi comparata mostra che, mentre l'Unione europea privilegia un equilibrio tra libertà di ricerca e tutela del diritto d'autore, Giappone e Stati Uniti adottano un approccio più aperto, valorizzando la funzione del TDM come motore di diffusione della conoscenza.

Tuttavia, la recente giurisprudenza tedesca *Kneschke c. LAION* mostra una crescente apertura verso un'interpretazione evolutiva del diritto d'autore, che si avvicina ai modelli più flessibili di matrice extraeuropea.

In prospettiva, questo potrebbe condurre a un graduale allineamento dell'Europa con le esperienze degli Stati Uniti e del Giappone, in cui il TDM è legittimo purché non arrechi un pregiudizio concreto all'autore né utilizzi l'opera per riprodurre il contenuto intellettuale o artistico.

Attualmente, però, si assiste a un differente approccio di Europa e Stati Uniti per quanto concerne la regolamentazione delle fattispecie collegate alle nuove tecnologie, con la particolarità che, mentre in genere gli Stati Uniti si assestano su posizioni di minore eteroregolamentazione e maggiore attenzione ai diritti individuali, in questo caso l'interesse alla diffusione della conoscenza, di rilevanza pubblicistica, assume maggiore rilievo rispetto a quanto prevedibile sulla base della prassi.

## 6. Conclusioni

Il presente contributo ha esaminato la sentenza *R. Kneschke c. LAION*, emessa dal Tribunale di Amburgo il 27 aprile 2023. La pronuncia è stata definita “storica” da molte testate giornalistiche, poiché per la prima volta un giudice europeo ha affrontato il tema delle eccezioni di TDM previste dalla direttiva CDSM, valutandone la compatibilità con il diritto d'autore in un contesto di addestramento dell'intelligenza artificiale.

I giudici, analizzando a fondo la questione, hanno concluso che le attività di *text and data mining* poste in essere dalla LAION non violavano il diritto d'autore del fotografo tedesco Kneschke, in quanto svolte ai fini di una ricerca scientifica; dunque, l'attività rientrava

nell'eccezione di TDM prevista dall'art. 3 della direttiva CDSM, ovvero dall' art. 60d dell'UrhG. Tra gli aspetti più innovativi, la sentenza ha proposto un'interpretazione evolutiva della nozione di "linguaggio intellegibile dalla macchina" per l'espressione delle riserve di *opt-out*, affermando che – in considerazione dell'avanzamento tecnologico dell'IA – tale linguaggio può comprendere anche quello naturale. Ciò riflette l'adattamento del diritto ai mutamenti tecnologici, imposto dal principio di neutralità tecnologica, secondo cui l'interpretazione delle norme deve evolversi insieme agli strumenti a cui esse si applicano.

Inoltre, il Tribunale ha chiarito che le attività di *web scraping* e di estrazione di opere protette non assumono automaticamente natura commerciale solo perché la GenAI è in grado di generare contenuti simili o potenzialmente concorrenti. La valutazione deve basarsi sull'uso concreto dei dati e sugli effetti sul mercato delle opere originali, non su presunzioni. Dunque, i giudici tedeschi si sono dimostrati aperti a nuove forme di interpretazione del diritto d'autore, proprio come già accade in Giappone e negli Stati Uniti, dove le attività di TDM sono sempre ritenute legali, purché non arrechino un pregiudizio all'autore dell'opera e non mirino a utilizzare il significato intellettuale della stessa.

Infine, riteniamo che le attività di TDM trovino fondamento giuridico anche nel diritto alla scienza, tutelato dal diritto internazionale, in particolare attraverso l'art. 15 del Patto Internazionale sui Diritti Economici, Sociali e Culturali, inteso come il diritto a godere dei benefici del progresso scientifico e delle sue applicazioni. Dunque, il diritto alla scienza è sinonimo di divulgazione della conoscenza, attualmente veicolata anche attraverso l'IA, che rientra tra le tecnologie dell'informazione (TIC), ossia strumenti a mezzo dei quali gli individui si informano. Ebbene, se l'IA è uno strumento veicolare della conoscenza, allora le attività di TDM, prodromiche all'addestramento dell'IA, sono parte del processo attraverso il quale si garantisce la divulgazione della stessa; quindi, nel momento in cui gli Stati prevedono eccezioni di TDM all'interno dei loro ordinamenti giuridici, stanno in realtà ottemperando agli obblighi derivanti dall'art. 15 ICESCR.

Infine, riprendendo quanto affermato dalla dottrina del *fair use*, riteniamo che non sia possibile ravvisare una violazione del *copyright* quando nell'*output* prodotto dall'intelligenza artificiale non sia riconoscibile un'opera dalla quale essa ha tratto ispirazione, proprio per lo stesso principio per cui non c'è plagio se, per la stesura di un contributo accademico, l'autore si avvale di molteplici fonti, che rielabora e trasforma aggiungendovi il suo pensiero critico.



**ABSTRACT (ITA)**

Le eccezioni di *Text and Data Mining* (TDM) previste dagli articoli 3 e 4 della direttiva 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale consentono di effettuare attività di riproduzione ed estrazione di opere coperte da *copyright* sia per finalità di ricerca che per scopi commerciali. Tuttavia, poiché anche l'addestramento dell'intelligenza artificiale si fonda, come il TDM, su un'analisi automatizzata dei dati volta a generare informazioni – tra cui modelli, tendenze e correlazioni – alcuni fornitori di modelli di IA hanno iniziato a invocare le predette eccezioni per giustificare l'utilizzo, senza licenza o autorizzazione, di opere protette da *copyright* ai fini dell'addestramento. In questo contributo si analizza una fattispecie decisa recentemente dal Tribunale di Amburgo, in cui i giudici tedeschi erano chiamati a stabilire se le attività di estrazione e analisi dei dati svolte a fini di ricerca potessero integrare una violazione del diritto d'autore qualora tra i dati utilizzati per l'addestramento dell'IA fosse compresa un'opera protetta. L'analisi prosegue con l'approfondimento di ulteriori elementi di prassi, tenendo conto delle linee di tendenza emergenti negli ordinamenti statali e nelle norme internazionali e dell'Unione europea. L'obiettivo è verificare se le eccezioni di *text and data mining* – in quanto strumenti funzionali alla ricerca scientifica e, al contempo, tecnologia alla base dell'addestramento dell'IA, oggi impiegata come mezzo di informazione – possano trovare fondamento nel diritto alla scienza, inteso quale diritto alla diffusione della conoscenza. A tal fine, si riflette anche sulla possibilità che tale diritto possa costituire un limite ai diritti di proprietà intellettuale riconosciuti e tutelati a livello internazionale.

**ABSTRACT (ENG)**

The exceptions for text and data mining (TDM) provided for in Articles 3 and 4 of Directive 2019/790 on copyright and related rights in the digital single market allow for the reproduction and extraction of copyright-protected works for both research and commercial purposes. However, since artificial intelligence training, like TDM, is based on automated data analysis aimed at generating information – including

patterns, trends and correlations – some AI model providers have begun to invoke the above exceptions to justify the use, without licence or authorisation, of copyright-protected works for training purposes. This article analyses a case recently decided by the Hamburg Regional Court, in which German judges were called upon to determine whether data mining and analysis activities carried out for research purposes could constitute copyright infringement if the data used for AI training included a protected work. The analysis continues with an in-depth examination of other significant elements of practice, taking into account emerging trends in national legal systems and in international and European Union regulations. The aim is to verify whether text and data mining exceptions – as tools for scientific research and, at the same time, the technology underlying the training of AI now used as a means of information – can be based on the right to science, understood as the right to the dissemination of knowledge. To this end, consideration is also given to the possibility that this right may serve as a limitation on intellectual property rights recognised and protected at international level.