

Diritti di riproduzione, banche dati e intelligenza artificiale “generativa”

Reproduction rights, databases and “generative” artificial intelligence

Carlo Meo*

ABSTRACT

Nel corso del 2023 numerosi editori giornalistici in tutto il mondo hanno deciso di impedire l'accesso ai propri contenuti ad OpenAI, l'operatore noto per aver sviluppato il servizio di intelligenza artificiale ChatGPT. Il lavoro prende spunto da questa vicenda per affrontare la questione se ed in quali circostanze le forme di sfruttamento delle opere poste in essere dall'IA generativa ricadano nell'ambito del diritto di riproduzione e dei diritti sulle banche dati. Il problema è esaminato con riferimento all'IA generativa di testo.

Several journalistic publishers around the world have recently decided to prevent access to their content by OpenAI, the operator known for developing the ChatGPT artificial intelligence service. The paper takes this case as a starting point to address the question whether and under what circumstances the forms of exploitation of works brought about by generative AI fall within the scope of reproduction rights and database rights. The issue is examined with reference to artificial intelligence generating text.

SOMMARIO:

1. Introduzione. – 2. L'IA generativa di testo. – 3. Estrazione di testo e di dati. – 4. Altre eccezioni e limiti al diritto d'autore. – 5. Riproduzioni temporanee. – 5.1. Temporalità. – 5.2. Utilizzo legittimo. – 5.3. Parte integrante ed essenziale del procedimento – 5.4. Transitorietà ed accessorialità. – 5.5. Mancanza di rilievo economico autonomo. – 5.6. Considerazioni di sintesi. – 6. I confini del diritto di riproduzione. – 7. Intelligenza artificiale e banche dati. – 7.1. Diritto d'autore e diritto *sui generis*. – 7.2. La decisione *Ryanair*. – 8. Conclusioni.

* Professore associato di diritto commerciale, Universitas Mercatorum; email: carlo.meo@unimercatorum.it.

1. Introduzione.

Nel corso del 2023 numerosi editori giornalistici in tutto il mondo hanno deciso di impedire l'accesso ai propri contenuti ad OpenAI, l'operatore noto per aver sviluppato il servizio di intelligenza artificiale ChatGPT. “Crawler” automatici “rastrellano” infatti da tempo i siti giornalistici per raccogliere dati con cui addestrare sistemi di intelligenza artificiale.

Diversi sono gli argomenti adoperati dagli editori per giustificare questa scelta. ChatGPT – si dice – manipola le informazioni e finisce per offrire agli utenti notizie false o imprecise. Il sistema si appropria, poi, dei contenuti prodotti da altri, senza includere citazioni o rinvii alle fonti. La principale preoccupazione dell'industria giornalistica sembra però un'altra. ChatGPT potrebbe diventare un diretto concorrente delle testate. Anziché leggere i giornali cartacei o digitali, gli utenti potrebbero interrogare il “chatbot” sulle notizie di interesse, ottenendo una sintesi delle informazioni contenute nei vari siti. Ciò azzererebbe il traffico sulle pagine dei giornali, privando gli editori degli introiti derivanti dalla pubblicità e dalla raccolta dei dati.

Lo strumento giuridico invocato dagli editori per bloccare OpenAI è il diritto d'autore. Gli articoli pubblicati sul sito – si dice – sono coperti da diritti esclusivi. Gli editori hanno dunque il diritto di impedirne l'uso ai sistemi di intelligenza artificiale¹.

Questa impostazione ha sollevato forti discussioni, anche al di fuori del campo dell'editoria giornalistica. L'esigenza di ottenere il consenso per l'uso di ogni contenuto adoperato nell'addestramento dell'IA rischia di rendere sostanzialmen-

¹ La ricerca oggetto del presente contributo si è conclusa nel mese di maggio 2024. A quanto risulta, le controversie in corso sul tema, specialmente negli USA, sono molto numerose, anche al di fuori del settore dell'editoria giornalistica. Si v. tra i casi più recenti *Basbanes et al. v. Microsoft Corp., et al.*, U.S. District Court for the Southern District of New York, 1:23-cv-10211 (filed Jan. 5, 2024); *The New York Times Co. v. Microsoft Corp., OpenAI*, U.S. District Court for the Southern District of New York, 1:23-cv-11195 (filed Dec. 27, 2023); *Sancton v. Microsoft Corp., OpenAI*, U.S. District Court for the Southern District of New York, 1:23-cv-10211 (filed Nov. 21, 2023); *Mike Huckabee v. Meta Platforms, Inc., Bloomberg L.P., Bloomberg Finance, L.P., Microsoft Corporation, and The EleutherAI Institute*, U.S. District Court for the Southern District of New York, 1:23-cv-09152 (filed Oct. 17, 2023); *Authors Guild v. Open AI*, U.S. District Court for the Southern District of New York, 1:23-cv-8292 (filed Sept. 19, 2023); *Chabon v. OpenAI Inc.*, U.S. District Court for the Northern District of California, 3:23-cv-04625-PHK (filed Sept. 8, 2023).

Per il momento, non risulta che sia stata ancora raggiunta una decisione definitiva negli USA sulla violazione dei diritti da parte dell'IA generativa. Alcuni profili del problema sono stati affrontati in maniera sommaria dalle corti in alcune ordinanze iniziali. Si v. ad es. *Kadrey v. Meta Platforms, Inc.*, 20 novembre 2023, 23-cv-03417-VC (N.D. Cal. Nov. 20, 2023); *Andersen v. Stability AI, Ltd.*, 30 ottobre 2023, 23-cv-00201-WHO (N.D. Cal. Oct. 30, 2023). V. anche l'ordinanza di rigetto della richiesta di giudizio sommario *Thomson Reuters, v. Ross Intelligence, Inc.*, 28 settembre 2023, 20-cv-613-SB (D. Del. Sep. 28, 2023).

te impossibile lo sviluppo di prodotti come ChatGPT. Sistemi in grado di elaborare immense quantità di dati e di offrire un quadro di sintesi su temi complessi devono, però, essere visti come risultati positivi dell'innovazione tecnologica. L'IA ha, del resto, già dimostrato le sue enormi potenzialità in alcuni campi, come la ricerca scientifica, l'analisi statistica e la programmazione elettronica.

Lo scontro tra editori e OpenAI ha acceso anche un dibattito in dottrina intorno alla questione se il diritto d'autore attribuisca effettivamente ai titolari dei diritti il potere di opporsi all'utilizzo delle opere da parte dell'intelligenza artificiale (d'ora in avanti, IA). Negli USA, una corrente di pensiero diffusa ritiene applicabile all'addestramento dell'IA il principio del "*fair use*"². Viceversa, in Europa sembra più diffusa la convinzione che l'uso dei testi da parte degli algoritmi sia incompatibile con l'attuale disciplina sul diritto d'autore. Gran parte dei contributi affronta quindi il tema principalmente da una prospettiva *de iure condendo*.

Il presente lavoro parte dall'idea che nel diritto dell'impresa si debba tenere conto, nei limiti del possibile, dell'internazionalizzazione dei rapporti e, quindi, anche dell'esigenza di evitare che, a livello comparatistico, si vengano a formare sistemi "a doppia velocità" su fenomeni di portata internazionale. Di qui l'idea di analizzare il problema non soltanto per individuare possibili riforme future del sistema europeo, ma ancor prima per chiedersi fin dove si possa arrivare in base alle regole attualmente in vigore nel nostro ordinamento. Ci si chiede, cioè, se mediante uno sforzo interpretativo non si possa già giungere, *de lege lata*, a soluzioni che consentano, almeno in certa misura, lo sviluppo dell'IA generativa.

Il lavoro affronta dunque la questione se ed in quali circostanze nel sistema europeo le forme di sfruttamento delle opere poste in essere dall'IA generativa ricadano nell'ambito dei diritti esclusivi. Il problema è esaminato con riferimento all'IA generativa di testo. Resta invece ai margini del lavoro l'analisi dei problemi sollevati dall'IA creatrice di immagini, per la quale si pongono questioni, in parte, diverse. Infine, il contributo non si occupa dell'ulteriore questione se i contenuti generati dall'IA siano protetti dal diritto d'autore³.

Per rispondere alla domanda da cui parte il presente lavoro occorre, in realtà, affrontare una pluralità di questioni. Le principali sembrano le seguenti: a) se la creazione di copie dei testi per l'addestramento dell'IA sia oggetto del diritto di riproduzione; b) se la raccolta di documenti dal *web* sia compatibile con i diritti sulle banche dati; c) se la generazione di testo da parte dell'IA violi i diritti di riproduzione e quelli di elaborazione sulle opere usate nella fase di addestramento.

² V. ad es. V. LINDBERG, *Building and using generative models under US copyright law*, in *Rutgers Business L. Rev.*, 2023, 1; M. SAG, *Copyright safety for generative AI*, in *Houston L. Rev.*, 2023, 104 ss.; M. A. LEMLEY, B. CASEY, *Fair Learning*, in *Texas L. Rev.*, 2021, 744 ss.

³ Anche questo problema è attualmente oggetto di discussione nel campo del giornalismo digitale. Si v. al riguardo A. TRAPOVA, P. MEZEL, *Robojournalism – A copyright study on the use of artificial intelligence in the European news industry*, in *GRUR Int.*, 2022, 589.

Il lavoro affronta le questioni a) e b). Per la questione c) si rinvia ad un successivo sviluppo del lavoro.

2. L'IA generativa di testo.

Nel concetto di IA generativa rientrano tutti i sistemi di IA che, a seguito di un addestramento realizzato su vaste quantità di dati, acquistano la capacità di produrre contenuti, come testi, immagini, sequenze di programmazione, brani musicali, ecc. ChatGPT si fonda su un c.d. “*large language model*” (LLM), cioè un sistema caratterizzato dalla capacità di generare linguaggio e di rispondere così agli *input* degli utenti.

Il processo che porta un'IA come ChatGPT a comunicare comincia con la raccolta di documenti dal *web*. Meccanismi automatici di ricerca (*bot* o *web-crawler*) passano in rassegna la rete, copiano testi presenti sulle pagine aperte al pubblico (articoli giornalistici, voci enciclopediche, raccolte di dati, libri, blog, ecc.) e archiviano le copie generate in banche dati (c.d. *dataset*).

I documenti contenuti nel *dataset* vengono poi sottoposti agli algoritmi dell'IA. In questa fase, i testi vengono utilizzati per effettuare esercizi di vario genere, come il riempimento di brani incompleti, la sistemazione di frasi disordinate o la continuazione di frasi tronche. Questi esercizi sono risolti dagli algoritmi per tentativi e, ad ogni passaggio, il sistema aggiorna i propri dati precedenti. In questo senso, l'IA “*apprende*” strada facendo la struttura ricorrente delle frasi e i principi fondamentali del linguaggio, cioè le regole grammaticali e sintattiche. L'IA registra poi anche il modo in cui vengono combinate le parole⁴.

Con questo procedimento, realizzato su una immensa quantità di testi, l'IA costruisce una mappa statistica pressoché completa delle possibili relazioni tra le parole che compongono una lingua. Essa è così in grado di orientarsi nel linguaggio e di generare testi originali, diversi da quelli utilizzati per l'addestramento⁵. In particolare, quando riceve una domanda dall'utente (c.d. *input* o *prompt*), l'IA in-

⁴ Per i sistemi in esame, l'addestramento avviene in genere senza supervisione, cioè su dati privi di “*etichettatura*” (c.d. *unsupervised learning*). Si v. sul punto T. GAO, Z. LIU, *Representation learning and NLP*, in *Representation learning for natural language processing*, edito da Liu, Lin, Sun, Springer, 2023, 8. Un intervento si ha comunque dopo il processo per apportare correttivi generali ed indirizzare la fase generativa, ad es., adeguando alcuni parametri, escludendo certi tipi di *output*, ecc. (c.d. “*fine tuning*”). V. C. CALLISON-BURCH, *Understanding artificial intelligence and its relationship to copyright*, Testimonianza scritta al US House of Representatives Judiciary Committee, 2023, <https://docs.house.gov/meetings/JU/JU03/20230517/115951/HHRG-118-JU03-Wstate-Callison-BurchC-20230517.pdf>.

⁵ Non c'è quindi un processo di vera e propria comprensione del testo da parte dell'IA: N. LUCCHI, *ChatGPT: a case study on copyright challenges for generative artificial intelligence systems*, in *Eur. J. of risk regulation*, 2023, 5.

dividua le parole chiave della domanda e costruisce la risposta con le parole che, sulla base dei modelli appresi, risultano statisticamente più adatte a proseguire il discorso iniziato dall'utente. L'IA parte, cioè, dall'*input* e prevede le parole successive più probabili. Poi ripete parola per parola questo processo, fino ad arrivare alla costruzione di un testo completo.

Fin qui, la generazione di testo sembra consistere sostanzialmente in un meccanismo di individuazione della parola più probabile data la parola precedente. Un sistema del genere, però, non sarebbe di per sé in grado di assicurare che si arrivi ad una frase di senso compiuto e tanto meno che la risposta sia complessivamente coerente con la domanda originaria⁶.

Questo problema è superato grazie al fatto che l'IA è anche in grado di tenere conto del "contesto" della conversazione nelle proprie risposte. In realtà, nella fase di addestramento, l'IA non si limita a contare le volte che una certa parola segue un'altra parola. Essa riesce anche ad individuare le espressioni chiave del testo che legge e a captare gli schemi linguistici che ricorrono in presenza di quelle espressioni. In altri termini, l'IA individua diverse tipologie di conversazioni possibili e adatta le statistiche registrate a questi diversi "contesti" comunicativi. Essa sa dunque modificare il proprio modo di parlare a seconda che le venga richiesta, ad es., un'informazione d'attualità, un racconto fantastico, una conversazione divertente oppure una teoria scientifica. L'IA è così in grado di conversare in maniera coerente sia da un punto di vista contenutistico che stilistico⁷.

La coerenza delle risposte è poi raggiunta anche grazie al fatto che, quando genera il testo, l'IA non individua il termine più probabile soltanto sulla base

⁶ Il sistema finirebbe poi anche per rispondere in maniera sempre uguale ad un medesimo *input*. Questo problema è risolto dall'IA, in genere, attraverso l'introduzione di una componente "random" nella generazione del testo. In altri termini, alcune delle parole usate vengono selezionate non sulla base di un calcolo statistico, ma attraverso meccanismi di individuazione casuale. Il che attribuisce varietà ed offre, in fase di *training*, anche la possibilità di ampliare le "conoscenze" del sistema.

⁷ In ciò sta una delle maggiori innovazioni dei sistemi di IA generativa, definiti sistemi "*transformer*", rispetto ai precedenti meccanismi. Si v. per una efficace descrizione sul punto M. SAG, (nt. 2), 2023: «one of the key differences between transformers and the prior state of the art, recurrent neural networks ("RNNs"), is that rather than looking at each word sequentially, a transformer first notes the position of the words. The ability to interpret these "positional encodings" makes the system sensitive to word order and context, which is useful because a great deal of meaning depends on sequence and context. Positional encoding is also important because it facilitates parallel processing: this in turn explains why throwing staggering amounts of computing power at LLMs works well for transformers, whereas the returns to scale for RNNs were less impressive. Transformers were also a breakthrough technology because of their capacity for "attention" and "self-attention." In simple terms, in the context of translation, this means that the system pays attention to all the words in source text when deciding how to translate any individual word. Based on the training data, the model learns which words in which contexts it should pay more and less attention to. Through "self-attention" the system derives fundamental relationships from input data and thus learns, for example, that "programmer" and "coder" are usually synonyms, and that "server" is a restaurant waiter in one context and a computer in another».

dell'ultima parola utilizzata, ma tiene conto di tutte le parole usate nella conversazione con l'utente (i.e. l'*input* e le parole già generate). Il testo prodotto dall'IA è quindi coerente anche con il "contesto" della specifica conversazione⁸.

3. Estrazione di testo e di dati.

Il procedimento fin qui sinteticamente descritto intercetta il diritto d'autore in varie fasi. Nella maggior parte dei casi, i testi presenti su Internet sono contenuti coperti dal diritto d'autore in qualità di «opere letterarie» (art. 2 l. aut.)⁹. Questo è senz'altro il caso degli articoli giornalistici raccolti dai siti *web* dei quotidiani. Gli articoli sono, peraltro, anche oggetto del diritto connesso degli editori sull'utilizzo online delle pubblicazioni giornalistiche, recentemente introdotto con l'art. 15 dir. 2019/790 (recepito in Italia nell'art. 43-bis l. aut.).

Per addestrare un'IA, in genere, i testi vengono scaricati da Internet, vengono tradotti in un formato adatto agli strumenti di lettura automatizzata e riversati in un *dataset*¹⁰. I *file* contenuti nel *dataset* possono poi eventualmente essere anche riprodotti, in tutto o in parte, per facilitare lo svolgimento degli esercizi di addestramento da parte dell'IA. Tutti questi passaggi comportano la creazione di copie dell'opera, e la creazione di copia è oggetto sia del diritto di riproduzione degli autori (art. 13 l. aut.) che del diritto connesso degli editori di giornale (art. 43-bis l. aut.)¹¹. A prima

⁸ Beninteso, tutto ciò non garantisce che le risposte fornite da un'IA siano corrette. L'IA riesce a rispondere alle domande che riguardano temi specifici soltanto grazie al fatto che durante l'addestramento ha incontrato le parole chiave relative al tema in questione e ha associato a queste parole un determinato contesto comunicativo, delle espressioni ricorrenti, ecc. Tutto ciò non garantisce però che la risposta sia corretta. Tanto più che, una volta concluso l'addestramento, l'IA perde accesso generalmente al *dataset* di riferimento. Sicché il sistema non ha neppure la possibilità di sottoporre le informazioni ad un controllo di attendibilità.

⁹ J.B. NORDEMAN, J. PUKAS, *Copyright exceptions for AI training data – will there be an international level playing field?*, in *J. IP Law and Practice*, 2022, 973, secondo cui la vasta maggioranza dei documenti che attualmente alimentano il processo di addestramento è coperta dal diritto d'autore. Fanno comunque eccezione le opere cadute in pubblico dominio e quelle prive di carattere "creativo", come i testi composti da formule matematiche, i testi normativi, i manuali di istruzioni per l'uso di prodotti, ecc. Si v. sul tema M. BERTANI, *Diritto d'autore europeo*, Torino, Giappichelli, 2011, 105 ss.

¹⁰ Si v. sul tema lo studio commissionato dalla Comm. Europea, *Study on copyright and new technologies: copyright data management and artificial intelligence*, SMART 2019/0038, 2022, 182 ss.

¹¹ L'esame del testo può anche avvenire con forme di "accesso diretto", vale a dire senza la intermediazione di una copia: il sistema individua il testo su Internet e lo sottopone ad analisi direttamente "alla fonte". In questo caso, non c'è una riproduzione, fatta salva forse la creazione di copie effimere nel processo di *training*. Comunque, a quanto risulta, questo metodo di analisi può trovare applicazione per sistemi semplici o che versano in una fase iniziale di sviluppo. Per la creazione di sistemi più complessi si segue, generalmente, un processo diverso, fondato sulla creazione di copie stabi-

vista, la raccolta dei testi è dunque un atto di sfruttamento soggetto all'esclusiva dei titolari dei diritti¹².

Nel dibattito pubblico sul tema si obietta che i testi vengono copiati da siti aperti al pubblico o, comunque, sono acquisiti tramite abbonamenti di lettura. In queste circostanze, gli autori non potrebbero pretendere di bloccare la successiva riproduzione dei testi da parte dei *crawler*. In realtà però, come è noto, il caricamento di un'opera su un sito *web* non produce l'esaurimento del diritto di riproduzione. Questo resta quindi esercitabile nei confronti delle forme di riutilizzo digitale dell'opera.

V'è anche chi sostiene che il titolare che accetta di caricare un articolo su un sito è consapevole che il suo testo potrà essere usato per attività di estrazione e accetta questa eventualità. In questo senso, si potrebbe dire che, con il caricamento, egli sta autorizzando implicitamente la copia della propria opera per fini di analisi. In realtà, però, gran parte dei titolari dei diritti è del tutto inconsapevole delle forme di utilizzazione computazionale che avvengono sul *web*. Quelli che ne sono consapevoli spesso si adoperano per bloccare l'estrazione, ad es., ponendo restrizioni al *download* seriale di contenuti dal sito. Peraltro, la mancanza di restrizioni tecniche alla copia del sito può derivare anche da un problema di costi o dalla difficoltà di costruire un efficace sistema di protezione. Non è dunque scontato che dietro il caricamento di un contenuto su un sito aperto al pubblico vi sia un'implicita accettazione dell'autore circa l'estrazione della propria opera. E non sembra quindi convincente leggere la pubblicazione digitale di un testo come una rinuncia dell'autore a far valere i propri diritti sul riutilizzo dell'opera¹³.

La creazione di copie a fini di addestramento è, dunque, in linea di principio, attività soggetta ai diritti esclusivi. Resta, però, da chiedersi se queste riproduzioni non possano rientrare in una delle eccezioni al diritto d'autore. Qui viene, innanzi-

li. Non è escluso comunque che la tecnica dell'analisi "diretta" possa presto trovare applicazione anche ai sistemi più complessi. Si v. sul tema le considerazioni svolte nello studio finanziato dalla Comm. Eur., J. TRIALLE, J. DE MEEÛS D'ARGENTEUIL, A. DE FRANQUEN, *Study on the legal framework of text and data mining*, 2014, 31 e 47. V. anche M. L. MONTAGNANI, G. AIME, *Il text and data mining e il diritto d'autore*, in *AIDA*, 2017, 382 e I. STAMATOUDI, *Text and data mining*, in *New Developments in EU and International Copyright Law*, edito da Stamatoudi, Wolters Kluwer, 2016, 1261.

¹² Il problema della creazione di copie per il *training* è al centro della maggior parte delle controversie statunitensi in corso in tema di IA. V. ad es. *Kadrey v. Meta Platforms, Inc.*, 20 novembre 2023, 23-cv-03417-VC (N.D. Cal. Nov. 20, 2023); *Andersen v. Stability AI, Ltd.*, 30 ottobre 2023, 23-cv-00201-WHO (N.D. Cal. Oct. 30, 2023); *Thomson Reuters, v. Ross Intelligence, Inc.*, 28 settembre 2023, 20-cv-613-SB (D. Del. Sep. 28, 2023).

¹³ In questo senso, si v. A. OTTOLIA, *Big data e innovazione computazionale*, Torino, Giappichelli, 2017, 33 ss. e, con riferimento alle copie "cache", B. HUGENHOLTZ, *Caching and copyright. The right of temporary copying*, in *EIPR*, 2000, 490 ss. V. anche CGUE, 16 novembre 2016, C-301/15, *Soulie e Døke*, par. 37 ss., in cui si afferma che un consenso implicito dell'autore può darsi soltanto laddove questi sia stato previamente informato della futura utilizzazione della sua opera da parte dei terzi e degli strumenti di cui dispone per vietarla.

tutto, in rilievo l'eccezione sull'estrazione di testo e di dati introdotta dalla dir. 2019/790 (art. 3 e art. 4, recepiti dagli artt. 70-ter ss. l. aut.)¹⁴. Con l'espressione «estrazione di testo e di dati» si intende «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» (art. 2, dir. 2019/790). Come già visto, l'addestramento dell'IA è un processo di analisi volto a costruire regressioni statistiche sull'uso del linguaggio. Sembra dunque poter rientrare nella definizione di «estrazione di testo e di dati»¹⁵. Il punto è attualmente confermato anche dall'art. 53 del testo finale del Regolamento europeo sull'intelligenza artificiale (*AI Act*). La disposizione impone infatti ai fornitori di modelli di IA il dovere di attuare politiche volte, tra l'altro, a rispettare le condizioni previste dalle disposizioni della dir. 2019/790 sull'estrazione di testo e di dati¹⁶.

Ai sensi di questa disciplina, gli organismi di ricerca e le istituzioni di tutela del patrimonio culturale possono liberamente riprodurre le opere per scopi di estrazione (art. 70-ter l. aut.), a condizione che l'operazione sia effettuata a fini di ricerca scientifica e che le riproduzioni riguardino solo documenti cui l'ente abbia legalmente accesso. Sono espressamente inclusi in questa categoria di documenti

¹⁴ L'eccezione si applica tanto al diritto d'autore quanto al diritto connesso degli editori di recente introduzione. V. art. 3, par. 1, e art. 4, par. 1, della dir. 2019/790.

¹⁵ In questo senso si esprime la maggior parte della dottrina. Si v. ad es. C. GEIGER, V. IAIA, *The forgotten creator: towards a statutory remuneration right for machine learning of generative AI*, in *Computer Law & Security review*, 2023; T. MARGONI, M. KRETSCHMER, *A deeper look into the EU text and data mining exceptions: harmonization, data ownership and the future of technology*, in *GRUR Int.*, 2022, 685. Qualche dubbio al riguardo è sollevato da J. B. NORDEMANN, J. PUKAS, (nt. 9), 974, secondo cui le correlazioni estratte dall'IA non sono accessibili all'uomo, ma possono essere usate soltanto dalla macchina stessa. Si dubita quindi che il sistema sia effettivamente volto a "generare informazioni". L'osservazione è condivisa anche dallo European Writers' Council: si v. lo *Statement on the trilogue negotiations of the AI Act proposal and on the urgently needed reform of the text and data mining exception Art. 4 of the CSDM directive 2019/790 (EU)*, 26 luglio 2023, disponibile al sito: https://europeanwriterscouncil.eu/23ewc_on_aiact/. Va detto però che la definizione di estrazione non contiene alcun riferimento all'utilizzo delle informazioni dopo la raccolta. Essa pare quindi prescindere dalla questione se le informazioni generate siano direttamente comprensibili dagli utenti oppure richiedano, a tal fine, un ulteriore passaggio tecnologico.

¹⁶ In questo senso, si v. P. KELLER, *A first look at the copyright relevant parts in the final AI Act compromise*, in *Kluwer Copyright Blog*, 11 dicembre 2023, disponibile al sito copyrightblog.kluweriplaw.com; I. EMANUILOV, T. MARGONI, *Forget me not: memorisation in generative sequence models trained on open source licensed code*, in *ssrn.com*, 2024, 20-21. Più precisamente, l'art. 53 afferma che il fornitore deve attuare politiche volte «ad adempiere al diritto dell'Unione in materia di diritto d'autore e diritti ad esso collegati e, in particolare, a 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» e deve mettere a disposizione del pubblico «una sintesi sufficientemente dettagliata dei contenuti utilizzati per l'addestramento del modello di IA per finalità generali».

quelli presenti in siti aperti al pubblico senza restrizioni¹⁷. Se necessario, le copie prodotte possono essere conservate con adeguati meccanismi di sicurezza. La raccolta di testi dal *web* e la creazione di un *dataset* per lo sviluppo dell'IA generativa sono dunque attività libere se poste in essere, ad es., da un'università per scopi meramente scientifici¹⁸.

La riproduzione a fini di estrazione è, in linea di principio, consentita anche agli organismi diversi da quelli di ricerca, come le imprese e gli enti pubblici (art. 4 dir. e art. 70-*quater* l. aut.). Anche qui, l'operazione deve riguardare documenti cui l'ente abbia legalmente accesso¹⁹. Per queste ipotesi di estrazione, però, si prevede che i titolari possano effettuare una "riserva". In tal caso, l'eccezione non si applica (sistema c.d. *opt-out*).

Per i testi caricati su siti aperti al pubblico, la riserva deve essere espressa in maniera tale che i sistemi di lettura automatizzata possano "catturarla", ad es., adoperando metadati o dando indicazione nelle condizioni d'uso del sito (art. 4, par. 3, dir. 2019/790 e cons. 18). Si tratta di un onere non molto gravoso, specialmente per le imprese che effettuano investimenti significativi per il funzionamento del proprio sito *web*, come gli editori giornalistici o letterari²⁰. V'è dunque la possibi-

¹⁷ V. considerando 14: «la nozione di accesso legale dovrebbe essere intesa nel senso che comprende l'accesso ai contenuti sulla base di una politica di accesso aperto o di accordi contrattuali, quali abbonamenti, tra i titolari dei diritti e gli organismi di ricerca o gli istituti di tutela del patrimonio culturale, o mediante altri mezzi legali».

¹⁸ La disposizione prevede anche che «i titolari dei diritti sono autorizzati ad applicare misure atte a garantire la sicurezza e l'integrità delle reti e delle banche dati in cui sono ospitate le opere o altri materiali. Tali misure non vanno al di là di quanto necessario per il raggiungimento di detto obiettivo». Le misure non devono, cioè, compromettere l'applicazione dell'eccezione (v. anche considerando 16). Per una panoramica sui possibili problemi applicativi sollevati da questa disposizione v. C. GEIGER, G. FROSIO, O. BULAYENKO, *Text and data mining in the proposed copyright reform: making the EU ready for an age of Big Data?*, in *IIC*, 2018, 836 ss.

Ai sensi del considerando 11, «in linea con l'attuale politica di ricerca dell'Unione, che incoraggia le università e gli istituti di ricerca a collaborare con il settore privato, gli organismi di ricerca dovrebbero beneficiare di una tale eccezione anche nel caso in cui le loro attività di ricerca siano svolte nel quadro di partenariati pubblico-privato. Gli organismi di ricerca e gli istituti di tutela del patrimonio culturale dovrebbero continuare a essere i beneficiari dell'eccezione, ma dovrebbero anche poter fare affidamento sui loro partner privati per effettuare l'estrazione di testo e di dati, anche utilizzando i loro strumenti tecnologici». Sembrerebbe dunque ammessa l'estrazione effettuata dai partner commerciali degli enti di ricerca. Naturalmente, a condizione che siano rispettati i limiti previsti dall'art. 3 e, quindi, che l'estrazione sia effettuata esclusivamente per scopi di ricerca scientifica e che sia effettuata su testi cui gli enti di ricerca abbiano legalmente accesso.

¹⁹ Le copie possono essere poi conservate per il tempo necessario ai fini dell'estrazione (art. 4, par. 2, dir. 2019/790).

²⁰ D'altra parte, l'*opt-out* può anche essere esercitato in maniera "collettiva" attraverso le *collecting societies*. È quanto accaduto con la francese SACEM, la quale ha recentemente esercitato l'*opt-out* per conto di tutti i titolari rappresentati. V. la dichiarazione del 12 ottobre 2023, al sito <https://societe.sacem.fr/en/news/our-society/sacem-favour-virtuous-transparent-and-fair-ai-exercises-its-right-opt-out>.

lità che si verifichi un ricorso generalizzato all'*opt-out* da parte di questi operatori. Tanto più che gli editori producono grandi quantità di testi ed hanno, quindi, più da guadagnare dalla concessione di eventuali licenze²¹. Per i sistemi di IA a carattere commerciale, la disciplina in tema di estrazione ha dunque un'efficacia piuttosto limitata²².

²¹ L. MANSANI, *Le eccezioni per estrazioni di testo e di dati, didattica e conservazione del patrimonio culturale*, in *AIDA*, 2019, 13; R. DUCATO, A. STROWEL, *Ensuring text and data mining: remaining issues with the EU copyright exceptions and possible ways out*, in *CRIDES Working Paper series*, 1/2021, 13. C'è da dire che, secondo alcuni, il controllo sul rispetto della riserva da parte dell'IA è molto difficile (se non addirittura impossibile) per i titolari. Sicché la tutela dell'*opt-out* rischia di essere sostanzialmente vanificata. Si v. *Study on copyright and new technologies: copyright data management and artificial intelligence*, SMART 2019/0038, 2022, 201. A questo proposito, il testo del regolamento europeo sull'IA (AI Act) contiene il dovere per i sistemi di IA di predisporre «una sintesi sufficientemente dettagliata dei contenuti utilizzati per l'addestramento del modello di IA per finalità generali» (art. 53). Secondo alcuni, si tratterebbe di una previsione volta proprio ad agevolare l'esercizio dell'*opt-out*. Si v. in tal senso C. GEIGER, V. IAIA, (nt. 15), 4 ss. e J. P. QUINTAIS, *Generative AI, copyright and the AI Act*, in *Kluwer Copyright Blog*, 9 maggio 2023.

²² Da questo punto di vista, l'UE ha adottato un approccio più restrittivo rispetto a quello seguito da altri ordinamenti, in cui il "*text and data mining*" è tendenzialmente consentito anche per fini commerciali. Si v. ad es. sulla soluzione adottata in Giappone, T. UENO, *The Flexible Copyright Exception for "Non-Enjoyment" Purposes – Recent Amendment in Japan and Its Implication*, in *GRUR Int.*, 2022. V. anche per una comparazione tra sistema giapponese e direttive europee: A. DERWAMAN, *Text and data mining exceptions in the development of generative AI models: what the EU member States could learn from the Japanese "nonenjoyment" purposes?*, in *J. of world IP*, 2023, 1. Nel Regno Unito, il governo ha avviato i lavori su un "*code of practice on copyright and AI*" volto ad elaborare soluzioni per semplificare l'acquisizione di licenze per l'estrazione a fini commerciali. Si v. le informazioni al sito: <https://www.gov.uk/guidance/the-governments-code-of-practice-on-copyright-and-ai>. Negli USA, la giurisprudenza ha talora qualificato l'uso delle opere per fini di estrazione come ipotesi di "*fair use*". L'uso è stato infatti considerato "trasformativo", i.e. «one that communicates something new and different from the original or expands its utility, thus serving copyright's overall objective of contributing to public knowledge»: *Authors Guild v. Google, Inc.*, 804 F. 3d 202 (2d Cir. 2015), 214. V. anche *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014). V. anche *Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 644–45 (4th Cir. 2009), relativo ad un sistema di analisi automatizzata del testo per finalità anti-plagio. Per un confronto della giurisprudenza statunitense con il diritto europeo si v. S. SCALZINI, *L'estrazione di dati e di testo per finalità commerciali dai contenuti degli utenti. Algoritmi, proprietà intellettuale e autonomia negoziale*, in *AGE*, 2019, 413 ss. Per una panoramica comparatistica sul *text and data mining* v. anche A. OTTOLIA, *L'opt-out commons nella nuova disciplina del data mining*, in *Il diritto d'autore nel mercato unico digitale*, a cura di COGO, in *Giur. it.*, 2022, 1255.

L'approccio dell'UE è considerato eccessivamente restrittivo dalla maggior parte della dottrina europea. Si v., ad es., il Position Statement del Max Planck di Monaco, *Artificial intelligence and intellectual property law*, 9 aprile 2021, 7. V. anche R. DUCATO, A. STROWEL, (nt. 21), 13 ss.; T. MARGONI, M. KRETSCHMER, (nt. 15), 685 ss.; C. GEIGER, G. FROSIO, O. BULAYENKO, (nt. 18); E. ROSATI, *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity*, in *Asia Pacific Law Review*, 2019, 199; G. GHIDINI, F. BANTERLE, *A critical view on the European Commission's proposal for a directive on copyright in the Digital Single Market*, in *Giur. comm.*, 2018, I, 961 ss.

Questa conclusione è rafforzata da un'altra considerazione. La creazione del *dataset* è un'attività complessa e può richiedere notevoli investimenti. Talora, ad occuparsi di questa fase del procedimento è un soggetto specializzato, diverso dall'impresa che sviluppa l'IA. È quanto accaduto nel caso di ChatGPT, che, almeno in una prima fase, è stato sviluppato da OpenAI sulla base di un *dataset* prodotto da Common Crawl. In un caso del genere, c'è quindi un'impresa che crea le copie, le immette in un *database* e poi trasferisce il *database* al programmatore dell'IA per l'addestramento. Come è noto, il trasferimento delle copie da un soggetto all'altro è oggetto di diritti diversi da quello di riproduzione, vale a dire dei diritti di distribuzione e di comunicazione al pubblico. L'eccezione sull'"estrazione" riguarda però esclusivamente il diritto di riproduzione. Essa consente, quindi, soltanto la creazione delle copie, non il loro successivo trasferimento, che resta soggetto agli altri diritti esclusivi. Nell'ipotesi, piuttosto comune, in cui il creatore del *dataset* sia un soggetto diverso dal programmatore, la disciplina sull'estrazione lascia comunque aperto il problema di raccogliere le autorizzazioni dei titolari delle opere coinvolte²³.

²³ In questo senso, si v. S. FLYNN, C. GEIGER, J. P. QUINTAIS, T. MARGONI, M. SAG, L. GUIBAULT, M.W. CARROLL, *Implementing user rights for research in the field of artificial intelligence*, in *American University Washington College of Law research paper*, 2020, 13; G. FRANCESCHELLI, M. MUSOLESI, *Copyright in generative deep learning*, in *Data & Policy*, 2022, 7 ss.; *Study on copyright and new technologies: copyright data management and artificial intelligence*, (nt. 10), 183. L'eccezione ai diritti sul trasferimento delle copie non pare potersi considerare implicita nell'eccezione al diritto di riproduzione. Il che si desume specialmente dall'art. 5, par. 4, della dir. 2001/29, in cui si afferma che «quando gli Stati membri possono disporre un'eccezione o limitazione al diritto di riproduzione in virtù dei paragrafi 2 e 3 del presente articolo, essi possono anche disporre un'eccezione o limitazione al diritto di distribuzione di cui all'articolo 4 nella misura giustificata dallo scopo della riproduzione permessa». Il che induce a pensare che, in mancanza di espressa previsione da parte del legislatore nazionale, l'eccezione al diritto di riproduzione si applichi soltanto alla creazione della copia. Peraltro, per l'"estrazione" la possibilità che gli Stati membri estendano l'eccezione alla distribuzione delle copie non è nemmeno contemplata nella direttiva.

Inoltre, come è noto, al "ritrasferimento" di un *file* digitale non si applica il principio dell'esaurimento. Il principio è infatti limitato dalla dir. 2001/29 al trasferimento di supporti "tangibili". V. considerando 29 dir. 2001/29: «la questione dell'esaurimento del diritto non si pone nel caso di servizi, soprattutto di servizi "on-line". Ciò vale anche per una copia tangibile di un'opera o di altri materiali protetti realizzata da un utente di tale servizio con il consenso del titolare del diritto. Perciò lo stesso vale per il noleggio e il prestito dell'originale e delle copie di opere o altri materiali protetti che sono prestazioni in natura. Diversamente dal caso dei CD-ROM o dei CD-I, nel quale la proprietà intellettuale è incorporata in un supporto materiale, cioè in un bene, ogni servizio "on-line" è di fatto un atto che dovrà essere sottoposto ad autorizzazione se il diritto d'autore o i diritti connessi lo prevedono». Il punto è stato recentemente confermato dalla CGUE. Si v. CGUE, 19 dicembre 2019, C-263/18, *Tom Kabinet*, par. 53 ss. L'applicazione dell'esaurimento alla diffusione digitale è comunque oggetto di discussione in dottrina. Si v. ad es. C. SGANGA, *Digital exhaustion after Tom Kabinet: a non-exhausted debate*, in *EU Internet law in the Digital Single Market*, edito da Synodinou, Jouglex, Markou, Prastitou-Merdi, Springer, 2021, 35 ss.; P. MEZEL, *Copyright exhaustion*, Cambridge, 2018, 92 ss.

4. Altre eccezioni e limiti al diritto d'autore.

Secondo alcune opinioni, le eccezioni sull'estrazione di testo e di dati sono le regole con cui l'ordinamento europeo ha inteso disciplinare il rapporto tra IA e diritto d'autore. L'uso delle opere da parte dell'IA sarebbe dunque consentito soltanto nel rispetto delle condizioni previste da queste eccezioni.

A prima vista, questa lettura è confermata dal considerando 105 del testo definitivo dell'*AI Act*, dove si legge che «qualsiasi utilizzo di contenuti protetti da diritto d'autore richiede l'autorizzazione del titolare dei diritti interessato, salvo se si applicano eccezioni e limitazioni pertinenti al diritto d'autore. La direttiva (UE) 2019/790 ha introdotto eccezioni e limitazioni che consentono, a determinate condizioni, riproduzioni ed estrazioni effettuate da opere o altri materiali ai fini dell'estrazione di testo e di dati. In base a tali norme, i titolari dei diritti hanno la facoltà di scegliere che l'utilizzo delle loro opere e di altri materiali sia da essi riservato per evitare l'estrazione di testo e di dati, salvo a fini di ricerca scientifica. Qualora il diritto di sottrarsi sia stato espressamente riservato in modo appropriato, i fornitori di modelli di IA per finalità generali devono ottenere un'autorizzazione dai titolari dei diritti, qualora intendano compiere l'estrazione di testo e di dati su tali opere». Il considerando pare, quindi, dare per scontato che l'IA possa fare uso di contenuti protetti soltanto alle condizioni previste dalle eccezioni di estrazione di cui alla direttiva 2019/790.

Dal canto suo, però, la stessa direttiva 2019/790 prevede che «vi possono essere anche casi di estrazione di testo e di dati che non comportano atti di riproduzione o in cui le riproduzioni effettuate rientrano nell'eccezione obbligatoria per gli atti di riproduzione temporanea di cui all'articolo 5, paragrafo 1, della direttiva 2001/29/CE, che dovrebbe continuare ad applicarsi alle tecniche di estrazione di testo e di dati che non comportino la realizzazione di copie al di là dell'ambito di applicazione dell'eccezione stessa» (considerando 9). In altre parole, secondo la direttiva 2019/790, le eccezioni di estrazione non sono le uniche vie percorribili per rendere compatibili i sistemi di analisi computazionale con il diritto d'autore. L'uso delle opere da parte di sistemi del genere è consentito anche nel caso in cui non vengano prodotte delle copie delle opere oppure nel caso in cui siano rispettate le condizioni per l'applicazione di altre eccezioni, come quella relativa alle riproduzioni temporanee.

Il considerando 105 dell'*AI Act* non pare allora in linea con le previsioni della direttiva 2019/790. In teoria, il considerando potrebbe essere letto come un tentativo di modificare la direttiva 2019/790, almeno per quanto riguarda i sistemi di IA. Questa lettura è, però, smentita dallo stesso *AI Act*, in cui si afferma che «il presente regolamento non pregiudica l'applicazione delle norme sul diritto d'autore previste dal diritto dell'Unione» (considerando 108). Inoltre, il regolamento elenca le fonti in cui si introducono modifiche e tra queste non figura la direttiva 2019/790.

Pare quindi più convincente ritenere che anche per i sistemi di IA valgano le considerazioni effettuate dal considerando 9 della direttiva 2019/790. Tali sistemi sono quindi compatibili con il diritto d'autore se: a) sono soddisfatti i requisiti per l'applicazione delle eccezioni di estrazione; b) sono soddisfatti i requisiti per l'applicazione di altre eccezioni, tra cui specialmente quella sulle riproduzioni temporanee; oppure c) non si verificano atti di riproduzione delle opere.

5. Riproduzioni temporanee.

Alle copie create per l'IA resta dunque potenzialmente applicabile l'eccezione di cui all'art. 5, par. 1, dir. 2001/29 (recepito in Italia con art. 68-*bis* l. aut.) in tema di riproduzioni temporanee. L'eccezione si applica se sono soddisfatte alcune condizioni cumulative. In particolare, le copie devono essere: a) temporanee; b) eseguite all'unico scopo di consentire la trasmissione in rete tra terzi con l'intervento di un intermediario oppure un utilizzo legittimo di un'opera; c) parte integrante ed essenziale di un procedimento tecnologico; d) transitorie o accessorie al procedimento; e) prive di rilievo economico proprio²⁴.

5.1. Temporaneità.

Il carattere “temporaneo” richiesto dalla condizione a) viene inteso in giurisprudenza nel senso che gli atti di riproduzione devono essere destinati alla cancellazione dopo un certo periodo di tempo²⁵. Nel caso dell'IA, il requisito è in genere soddisfatto dai primi atti di *download* dei testi dal *web*. Queste copie servono infatti soltanto come base per creare delle ulteriori copie da inserire nel *dataset* in un formato leggibile dagli algoritmi. Una volta realizzate queste seconde copie, le prime divengono obsolete e possono essere eliminate. Lo stesso vale per le eventuali copie prodotte durante gli esercizi di addestramento. Queste sono infatti quasi sempre riproduzioni effimere, volte a facilitare l'esercizio dell'IA e sono automaticamente eliminate alla conclusione dell'esercizio stesso o, comunque, alla chiusura dell'addestramento²⁶.

²⁴ La disposizione non è espressamente richiamata tra le eccezioni applicabili al diritto degli editori di giornali di cui all'art. 15 dir. 2019/790. Essa pare comunque applicabile al diritto in questione. L'art. 15 dir. 2019/790 afferma infatti che «gli Stati membri riconoscono agli editori di giornali stabilito in uno Stato membro i diritti di cui all'articolo 2 [...] della direttiva 2001/29/CE per l'utilizzo online delle loro pubblicazioni di carattere giornalistico da parte di prestatori di servizi della società dell'informazione». Il contenuto del diritto è definito con un rinvio all'art. 2 dir. 2001/29 (i.e. il diritto di riproduzione degli autori), il quale è soggetto all'eccezione di cui all'art. 5 par. 1.

²⁵ CGUE, 5 giugno 2014, C-360/13, *Public Relations*, par. 26.

²⁶ J. VESALA, *Developing artificial intelligence-based content creation: are EU copyright and antitrust law fit for purpose?*, in *IIC*, 2023, 361.

Più complessa è invece la situazione delle copie che compongono il *dataset*. Qui può, innanzitutto, accadere che il *dataset* venga interamente eliminato alla fine della procedura di addestramento dell'IA. Il requisito della "temporaneità" sarebbe, in tal caso, senz'altro rispettato²⁷. Potrebbe, però, anche accadere che il *dataset* sia conservato per un tempo più lungo, ad es., per consentire controlli o aggiornamenti periodici del sistema. La questione se copie come queste siano "temporanee" non è risolta espressamente dalla direttiva. La *ratio* del requisito di temporaneità è però quella di consentire le riproduzioni indispensabili allo sviluppo di un procedimento tecnologico ed evitare di "liberalizzare" copie che invece vanno oltre questo obiettivo. La fase di revisione è un passaggio indispensabile per assicurare il buon funzionamento di un procedimento. Se le copie necessarie per la revisione non fossero consentite, la disposizione non riuscirebbe a realizzare efficacemente la propria funzione. Tali copie sembrano dunque doversi considerare, a loro volta, compatibili con il requisito di "temporaneità", sempre che la loro durata sia strettamente collegata alla realizzazione di controlli e aggiornamenti.

Fin qui si è trattato di casi in cui la produzione del *dataset* è collegata alla realizzazione di un unico procedimento di addestramento. Si è detto, però, che la realizzazione di un buon *dataset* è un'attività complessa che può richiedere investimenti anche notevoli. Nella maggior parte dei casi, quindi, il *dataset* è costruito per essere utilizzato in una pluralità di procedimenti, anche futuri²⁸. Talora, esso è addirittura sviluppato da operatori specializzati che intendono offrirlo ai terzi sul mercato. In tutti questi casi, il *dataset* è prodotto per durare. E ciò esclude la possibilità di applicare l'art. 5, par. 1²⁹.

Un'eccezione a questo discorso può aversi, tutt'al più, nel caso in cui il contenuto del *dataset* venga periodicamente rinnovato. Può accadere, infatti, che i documenti all'interno del *dataset* siano sostituiti dopo un certo tempo con testi nuovi, per assicurare diversità al materiale d'addestramento. In queste circostanze, il *dataset* in quanto tale è stabile, ma le copie che lo compongono hanno durata limitata³⁰. In ogni caso, come l'eccezione di estrazione, anche l'art. 5, par. 1 si riferisce soltanto al diritto di riproduzione. La disposizione può quindi consentire la creazione del *dataset*, ma non il suo eventuale trasferimento da un'impresa che lo produce ad una che lo utilizza. Questo passaggio resta soggetto ai diritti di distribuzione o di comunicazione al pubblico.

²⁷ D. SCHÖNBERGER, *Deep copyright: Up – and downstream questions related to artificial intelligence (AI) and machine learning (ML)*, in *Droit d'auteur 4.0*, edito da De Werra, 2018, 145.

²⁸ M. LEMLEY, B. CASEY, (nt. 2), 753.

²⁹ In questo senso, T. MARGONI, M. KRETSCHMER, (nt. 15), 693. M. SENFTLEBEN, *Compliance of national TDM rules with international copyright law: an overrated nonissue?*, in *IIC*, 2022, 1483. V. anche G. FRANCESCHELLI, M. MUSOLESI, (nt. 23), 2022, e17.

³⁰ J. VESALA, (nt. 26), 361 ss.

5.2. Utilizzo legittimo.

Le copie temporanee devono poi essere volte soltanto a consentire una trasmissione in rete tra terzi oppure un utilizzo legittimo delle opere (condizione b). Il procedimento dell'IA non è un'ipotesi di trasmissione tra terzi. Resta dunque da vedere se le copie realizzate in questo procedimento possano dirsi funzionali a realizzare un utilizzo "legittimo" delle opere.

Secondo la Corte, un utilizzo è da considerare legittimo in due casi: se è autorizzato dal titolare o se non rientra tra le attività riservate dalla legge ai titolari³¹. Nel caso in esame, l'utilizzo che il procedimento punta a realizzare è l'analisi delle opere per l'apprendimento del linguaggio da parte dell'IA. Come già visto, in genere, questo utilizzo non viene autorizzato dai titolari. Non resta, allora, che domandarsi se l'analisi automatizzata di un testo in quanto tale rientri tra le attività che la legge sottopone all'esclusiva dell'autore.

La questione non è risolta espressamente dalla legge. In genere, la mera "fruizione" di un'opera non è considerata oggetto dei diritti d'autore. Così la lettura, l'analisi e lo studio di un'opera letteraria sono tradizionalmente considerate utilizzazioni libere. Secondo gran parte della dottrina, ciò dovrebbe valere anche per le forme di "lettura" automatizzata poste in essere da una macchina³².

Quest'impostazione è però criticata da alcuni autori. La lettura artificiale sarebbe radicalmente distinta dalla lettura umana. La macchina non si limita a "fruire" dell'opera come un qualunque essere umano, ma estrae dall'analisi un valore "informativo", che viene poi generalmente reimpiegato in attività commerciali. L'analisi in questione non potrebbe essere qualificata come ipotesi di "mero godimento", ma sarebbe a tutti gli effetti "utilizzazione economica" dell'opera³³.

Quest'ultima tesi appare oggi smentita dalla disciplina, già richiamata, sul-

³¹ CGUE, 26 aprile 2017, C-527/15, *Stichting Brein*, par. 65 ss.; CGUE, 4 ottobre 2011, cause riunite C-403/08, C-429/08, *FA Premier League*, par. 168; CGUE, 17 gennaio 2012, C-302/10, *Infopaq*, par. 44 ss.

³² Si v. R.M. HILTY, H. RICHTER, *Position statement of the Max Planck Institute on the modernisation of European copyright rules*, 2017, 2 ss.; A. MUSSO, *Eccezioni e limitazioni ai diritti d'autore nella direttiva UE n. 790/2019*, in *Dir. Informazione e dell'informatica*, 2020, 411 ss.; J. TRIAILLE, J. DE MEEÛS D'ARGENTEUIL, A. DE FRANQUEN, *Study on the legal framework of text and data mining*, (nt. 11), 31; J. LITMAN, *The exclusive right to read*, in *Cardozo Art & Ent. Law J.*, 1994, 29.

³³ In questo senso pare di poter leggere A. OTTOLIA, *L'utilizzo computazionale dell'opera dell'ingegno in Internet*, in *AIDA*, 2014, 386 ss. La tesi è poi sviluppata ulteriormente in A. OTTOLIA, (nt. 13), 19 ss. V. più recentemente A. OTTOLIA, (nt. 22), 1254 ss. Si v. anche sul punto D. SARTI, *Diritti esclusivi e circolazione dei beni*, Milano, Giuffrè, 1996, 359 ss., in cui l'A. sottolinea che l'argomento del "mero godimento" non può essere utilizzato per giustificare utilizzazioni che aumentano la concorrenzialità delle organizzazioni produttive. Il riferimento è specialmente allo sfruttamento del *software*. Secondo R. SERVANZI, *Le estrazioni di testo e di dati*, in *Nuove leggi civ. comm.*, 2022, 1152, questo ragionamento sarebbe applicabile anche all'uso delle opere nel *text and data mining*.

l'estrazione di testo e di dati di cui alla direttiva 2019/790. L'obiettivo dichiarato di queste eccezioni è quello di liberalizzare le ipotesi di analisi computazionale che rispettano una serie di condizioni. Le eccezioni si applicano, però, soltanto alle "riproduzioni" necessarie all'analisi. Non all'analisi in quanto tale. Da ciò si desume che, secondo il legislatore, l'analisi artificiale del testo non è di per sé oggetto di esclusiva e, pertanto, non richiede apposite eccezioni³⁴.

D'altra parte, il diritto d'autore si fonda sul principio per cui la protezione si estende solo alla "espressione", cioè al modo in cui l'idea creativa è espressa dall'autore, alla "forma" dell'opera. Non sono invece tutelate le idee e le informazioni contenute nell'opera³⁵. L'autore non può impedire l'utilizzo, anche per fini

³⁴ A conferma di questa lettura si v. il considerando 9 dir. 2019/790/UE: «L'estrazione di testo e di dati può essere effettuata anche in relazione a semplici fatti o dati non tutelati dal diritto d'autore, nel qual caso non è richiesta alcuna autorizzazione in base alla legislazione sul diritto d'autore. Vi possono essere anche casi di estrazione di testo e di dati che non comportano atti di riproduzione o in cui le riproduzioni effettuate rientrano nell'eccezione obbligatoria per gli atti di riproduzione temporanea [...]». Non convince dunque la critica, mossa da alcuni, secondo cui l'eccezione di estrazione avrebbe l'effetto di estendere indirettamente il diritto d'autore fino a coprire atti di mera "lettura" prima sottratti all'esclusiva. Sul punto si v. ad es. T. MARGONI, M. KRETSCHMER, (nt. 15), 693. Va detto che l'art. 3 dir. 2019/790 include espressamente tra le attività liberalizzate anche le "estrazioni". Questo però si spiega con il fatto che l'eccezione si applica anche al diritto *sui generis* sulle banche dati, il quale, come è noto, ha proprio ad oggetto l'estrazione di informazioni.

Su questi temi, una qualche ambiguità è creata dal considerando 105 del regolamento sull'intelligenza artificiale (*AI Act*), già richiamato nel testo, in cui si legge che «le tecniche di estrazione di testo e di dati possono essere ampiamente utilizzate in tale contesto per il reperimento e l'analisi di tali contenuti, che possono essere protetti da diritto d'autore e da diritti connessi. Qualsiasi utilizzo di contenuti protetti da diritto d'autore richiede l'autorizzazione del titolare dei diritti interessato, salvo se si applicano eccezioni e limitazioni pertinenti al diritto d'autore». Il che sembrerebbe presupporre che anche la mera analisi sia attività riservata. Questa lettura non convince, innanzitutto, per le ragioni già esposte nel par. 4. Inoltre, per le ragioni che sono espresse *infra* nel presente paragrafo, questa lettura "estensiva" dell'esclusiva non pare compatibile con i principi generali del diritto d'autore.

In Francia, è stata presentata una proposta di riforma volta, tra l'altro, a sottoporre l'addestramento dell'IA al diritto esclusivo dei titolari. V. art. 1 della *Proposition de loi* n. 1630 – *Visant à encadrer l'intelligence artificielle par le droit d'auteur*, pres. 12 settembre 2023: «l'intégration par un logiciel d'intelligence artificielle d'œuvres de l'esprit protégées par le droit d'auteur dans son système et a fortiori leur exploitation est soumise aux dispositions générales du présent code et donc à autorisation des auteurs ou ayants droit». La proposta prevede, addirittura, che i diritti sui contenuti generati dall'IA siano assegnati ai titolari dei diritti sulle opere che hanno reso possibile il risultato creativo. V. art. 2: «lorsque l'œuvre est créée par une intelligence artificielle sans intervention humaine directe, les seuls titulaires des droits sont les auteurs ou ayants droit des œuvres qui ont permis de concevoir ladite œuvre artificielle». La proposta è stata criticata duramente in dottrina: si v. ad es. C. GEIGER, V. IAIA, (nt. 15), 7 ss.

³⁵ Come è noto, il principio è codificato in diverse disposizioni dell'ordinamento. Nella l. aut. nazionale, si trova espresso all'art. 2, n. 8-9, con riferimento ai *software* e alle banche dati. A livello europeo, è richiamato all'art. 1 della dir. 91/250 sui programmi per elaboratore e agli artt. 3 e 5 della dir. 96/9 sulla protezione delle banche dati. Il principio è anche ribadito dal considerando 9 della dir.

commerciali, delle informazioni, degli insegnamenti e dei concetti tratti dalla propria opera³⁶. L'analisi computazionale ha la funzione di estrarre dall'opera informazioni che sarebbero inaccessibili attraverso la normale elaborazione umana³⁷. Riconoscere all'autore il potere di impedire l'analisi computazionale significa sostanzialmente impedire ai terzi lo sfruttamento di queste informazioni. Significa, quindi, estendere l'esclusiva a campi che la legge intende sottrarre al controllo dell'autore³⁸.

Si potrebbe obiettare che l'analisi computazionale estrae dati non tanto dal contenuto, ma dal testo dell'opera. Essa sarebbe quindi pur sempre un'utilizzazione della "forma espressiva" dell'opera. E, come tale, dovrebbe rientrare nell'ambito dell'esclusiva³⁹. Occorre intendersi, però, sul significato del concetto di "espressione" utilizzato dal legislatore. La forma espressiva è il veicolo che consente all'autore di comunicare le proprie idee e di trasmettere emozioni, sentimenti e riflessioni. Sono questi gli elementi che spingono il pubblico a "consumare" un libro o un articolo e che, quindi, determinano l'esistenza di un mercato dell'opera letteraria. Da questo punto di vista, la "forma espressiva" deve effetti-

2019/790, in cui si legge che «l'estrazione di testo e di dati può essere effettuata anche in relazione a semplici fatti o dati non tutelati dal diritto d'autore, nel qual caso non è richiesta alcuna autorizzazione in base alla legislazione sul diritto d'autore». Il principio che sottrae le idee alla tutela autoriale ha trovato poi applicazione anche nella giurisprudenza della Corte di Giustizia. Si v. recentemente CGUE, 11 giugno 2020, C-833/18, *Brompton Bicycle*, par. 27; CGUE, 22 dicembre 2010, C-393/09, *BSA*, par. 48 ss.; CGUE, 2 maggio 2012, C-406/10, *SAS*, par. 31 ss. Infine, il principio è presente nelle convenzioni internazionali in materia di diritto d'autore. Si v. art. 9 par. 2 TRIPs e art. 2 WCT.

³⁶ In questo senso, si v. M. BERTANI, (nt. 9), 109, secondo cui il principio per cui il diritto d'autore non si estende al sapere teorico si spiega, tra l'altro, con l'esigenza di consentire l'innovazione concorrente e il progresso culturale.

³⁷ M. GRANIERI, *Il data mining nella disciplina del diritto d'autore e la strategia europea sui dati*, in *AIDA*, 2022, 24; G. ROSSI, *Opere dell'ingegno come dati: il text and data mining nella direttiva 2019/790*, in *AIDA*, 2019, 235.

³⁸ In questo senso, si v. anche M. LEMLEY, B. CASEY, (nt. 2), 750; T. MARGONI, M. KRETSCHMER, (nt. 15), 689; V. MOSCON, *Data Access Rules, Copyright and Protection of Technological Protection Measures in the EU. A Wave of Propertisation of Information*, MPI Research Paper n. 23/14, 2023, 9-10; A. STROWEL, *Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals*, in *Copyright reconstructed*, edito da B. Hugenholtz, Wolters Kluwer, 2018, 226. Il punto sembra condiviso anche dalla giurisprudenza statunitense che si è occupata degli usi "trasformativi" delle opere dell'ingegno in campo digitale. Si v. ad es. il caso *Authors Guild v. Google* (2015) F. 3d 202, in cui si legge, con riferimento alle funzioni di ricerca nelle opere e agli *snippet*, che «the copyright resulting from the Plaintiffs' authorship of their works does not include an exclusive right to furnish the kind of information about the works that Google's programs provide to the public. For substantially the same reasons, the copyright that protects Plaintiffs' works does not include an exclusive derivative right to supply such information through query of a digitized copy». V. anche *Thomson Reuters, v. Ross Intelligence, Inc.*, 28 settembre 2023, 20-cv-613-SB (D. Del. Sep. 28, 2023), 24.

³⁹ B.L.W. SOBEL, *Artificial intelligence's fair use crisis*, in *Col. J. of law & the arts*, 2017, 46 ss.

vamente essere soggetta al controllo dell'autore. La forma di un'opera non è però soltanto veicolo di comunicazione della personalità del suo autore. Essa ha anche una ricca componente di informazioni: il testo contiene indicazioni sulla grammatica, sul significato delle parole, sul modo in cui esse vanno utilizzate, sulla frequenza delle soluzioni espressive, ecc. Se queste informazioni fossero oggetto dell'esclusiva, si giungerebbe al risultato assurdo di assegnare ad un autore il potere di opporsi all'uso altrui di una certa lingua. La componente "informativa" del testo cade dunque fuori dall'ambito dell'esclusiva.

Questa distinzione tra "forma" come veicolo di comunicazione e forma come veicolo di informazione è stata, a lungo, superflua. La raccolta di informazioni sulla lingua da parte del lettore di un libro è, infatti, attività inscindibile dalla fruizione della espressione creativa dell'autore. Ciò giustifica, ad es., che la vendita di un romanzo ad uno studente interessato ad usarlo per esercitarsi con la lingua sia attività soggetta all'esclusiva dell'autore. I due aspetti diventano invece scindibili nel caso dell'IA: gli algoritmi che "leggono" il testo non percepiscono l'espressione creativa dell'autore e non capiscono le sue scelte creative; essi si limitano ad estrapolare le informazioni statistiche sull'uso del linguaggio che sono racchiuse nelle scelte espressive. In sostanza, l'IA fa uso dell'espressione soltanto come veicolo di informazioni, non come veicolo di comunicazione creativa⁴⁰. La lettura artificiale del testo non sembra dunque potersi qualificare come attività riservata all'autore.

L'idea che l'addestramento, in sé, sia un utilizzo legittimo potrebbe essere, ancora, contestata sulla base del fatto che, una volta concluso il procedimento, l'IA potrebbe teoricamente utilizzare il linguaggio appreso per produrre testi in violazione del diritto d'autore⁴¹. Questo, ovviamente, non sarebbe un utilizzo "legittimo" delle opere dell'ingegno. Le copie temporanee non sarebbero allora qui realizzate "all'unico scopo" di consentire un utilizzo legittimo e l'art. 5, par. 1 non sarebbe applicabile all'IA generativa. In sostanza, l'obiezione in esame si fonda sull'idea che l'art. 5, par. 1 non possa essere applicato ad un utilizzo legittimo che può eventualmente portare ad un successivo utilizzo illegittimo.

La tesi non considera, però, che, in realtà, qualsiasi utilizzo dell'opera, anche legittimo, è astrattamente idoneo a causare successivi utilizzi illegittimi: la lettura di un testo dal sito può portare l'utente a scaricare una copia del testo sul compu-

⁴⁰ Considerazioni simili paiono espresse in M. BORGHI, S. KARAPAPA, *Non-display uses of copyright works: Google books and beyond*, in *Queen Mary J. of IP*, 2011, 44 ss.; M. LEMLEY, B. CASEY, (nt. 2), 749. V. anche le riflessioni sulla forma espressiva in G. SPEDICATO, *Interesse pubblico e bilanciamento del diritto d'autore*, Milano, Giuffrè, 2013, 156 ss.

⁴¹ L'argomento sembra sollevato dall'editore nel caso *Thomson Reuters, v. Ross Intelligence, Inc.*, 28 settembre 2023, 20-cv-613-SB (D. Del. Sep. 28, 2023), 19. V. anche sul punto G. FRANCESCHELLI, M. MUSOLESI, (nt. 23), 8; J. VESALA, (nt. 26), 362; S. KARAPAPA, *Defences to copyright infringement*, Oxford, Oxford UP, 2020, 112 ss.

ter da diffondere poi su altri siti; la trasmissione lecita di un videoclip su una piattaforma può essere utilizzata per effettuare registrazioni abusive; e via dicendo. In sostanza, ad un procedimento di utilizzo legittimo può sempre seguire un diverso procedimento illegittimo. Dire che l'art. 5, par. 1 è applicabile solo quando non vi sia alcun rischio di successivi utilizzi illegittimi significa allora sostanzialmente dire che la disposizione non può mai trovare applicazione.

È vero, per altro verso, che il caso dell'IA presenta una particolarità rispetto agli esempi sopra richiamati. L'eventuale violazione del diritto d'autore nella generazione di testo da parte dell'IA non è un procedimento del tutto autonomo rispetto a quello di addestramento, ma è uno sviluppo o, comunque, un'applicazione dello stesso procedimento. In altre parole, utilizzo legittimo e utilizzo illegittimo si verificano qui nell'ambito di procedimenti tecnologici collegati e consequenziali. Si pone dunque la questione se l'art. 5, par. 1 sia applicabile quando il successivo utilizzo illegittimo dell'opera provenga da un'applicazione o da uno sviluppo del procedimento legittimo.

La questione non è risolta espressamente dalla direttiva. Come riconosciuto dalla Corte di Giustizia, l'obiettivo dell'art. 5, par. 1 è però quello di «consentire e assicurare lo sviluppo ed il funzionamento di nuove tecnologie, nonché mantenere un giusto equilibrio tra i diritti e gli interessi dei titolari di diritti e degli utilizzatori delle opere protette che intendano beneficiare di tali tecnologie»⁴². L'IA può avere innumerevoli applicazioni utili e virtuose. Dire che l'art. 5, par. 1 non si applica in questo campo significa, di fatto, impedire la realizzazione dei risultati di benessere generale che queste linee di innovazione potrebbero realizzare. Se letta in questo senso, la disposizione finirebbe dunque per dare assoluta prevalenza agli interessi degli autori rispetto all'interesse generale. Il che non pare coerente con l'obiettivo di realizzare un giusto equilibrio tra gli interessi in gioco. Tanto più che, se l'IA, una volta sviluppata, violasse i diritti d'autore ricopiando un'opera, l'autore resterebbe comunque legittimato ad esercitare i suoi diritti di riproduzione e di comunicazione al pubblico nei confronti dei testi generati dall'IA. Per assicurare tutela ai titolari contro gli utilizzi illegittimi, non c'è allora bisogno di impedire del tutto lo sviluppo del procedimento.

L'art. 5, par. 1 pare dunque doversi piuttosto leggere nel senso che un procedimento, come quello dell'IA, che abbia come risultato un utilizzo "legittimo" delle opere può rientrare nell'eccezione, a prescindere dal fatto che lo stesso procedimento possa poi dare luogo anche ad alcune applicazioni illegittime, le quali ultime restano invece soggette alla comune disciplina protettiva del diritto dell'autore.

⁴² CGUE, 5 giugno 2014, C-360/13, *Public Relations*, par. 24.

5.3. Parte integrante ed essenziale del procedimento.

La condizione c) (i.e. che le copie siano parte integrante ed essenziale del procedimento) viene intesa nella giurisprudenza europea nel senso che gli atti di riproduzione devono risultare necessari affinché il procedimento funzioni efficacemente e devono essere interamente compiuti nell'ambito del procedimento stesso⁴³.

Le copie funzionali alla costruzione di un *dataset* sono necessarie per il buon funzionamento dell'addestramento, visto che questa non può prescindere dal contatto con i testi per apprendere il linguaggio umano⁴⁴.

Più articolata è invece la risposta alla domanda se il *dataset* sia realizzato interamente "all'interno" del procedimento. Il requisito pare doversi intendere nel senso che il procedimento non possa servirsi di copie prodotte in precedenza per altri fini, ma soltanto di copie prodotte in occasione del procedimento stesso. Pertanto, il requisito non è soddisfatto se il *dataset* viene creato in vista di una pluralità di applicazioni future o, addirittura, per essere venduto. Lo è invece nel caso in cui il *dataset* sia costruito appositamente per uno specifico processo di addestramento.

5.4. Transitorietà e accessorieta.

Ai sensi della condizione d), le copie devono poi essere, alternativamente, transitorie o accessorie al procedimento. La disposizione è piuttosto ambigua, in quanto il concetto di "transitorietà" sembra già implicito nel requisito della "temporaneità". La Corte sembra però intendere i concetti in maniera diversa: temporaneità significa che la copia deve avere una durata limitata; "transitorietà", invece, significa che le copie devono essere cancellate automaticamente una volta esaurito il proprio ruolo nel procedimento⁴⁵. In quest'ottica, possono esserci quindi copie "temporanee" e non "transitorie": è il caso in cui per le copie è previsto un termine di durata, ma la loro cancellazione richiede un intervento manuale dell'uomo.

⁴³ CGUE, 17 gennaio 2012, C-302/10, *Infopaq*, par. 30; CGUE, 5 giugno 2014, C-360/13, *Public Relations*, par. 28.

⁴⁴ Affinché il requisito sia rispettato, basta che le copie rendano il processo più efficace. Si v. CGUE, 5 giugno 2014, C-360/13, *Public Relations*, par. 35 ss. Sul tema, v. anche G. GUGLIELMETTI, *Riproduzione e riproduzione temporanea*, in *AIDA*, 2002, 35 ss.

⁴⁵ CGUE, 16 luglio 2009, C-5/08, *Infopaq*, par. 62; CGUE, 5 giugno 2014, C-360/13, *Public Relations*, par. 40: «un atto può essere qualificato come "transitorio" esclusivamente qualora la sua durata sia limitata a quanto necessario per il buon funzionamento del procedimento tecnologico utilizzato, restando inteso che tale procedimento deve essere automatizzato in modo tale da cancellare detto atto in maniera automatica, senza intervento umano, nel momento in cui è esaurita la sua funzione tesa a consentire la realizzazione di un siffatto procedimento».

Nel campo dell'IA, possono essere soggette ad eliminazione automatica le copie iniziali prodotte dal *download* dei testi dal *web* e quelle effimere eventualmente necessarie per lo svolgimento di esercizi di *training*. Lo stesso non vale sempre per il *dataset*, la cui eliminazione dipende dall'esito del procedimento ed è dunque, in genere, rimessa alle scelte dei programmatori.

Per queste copie, occorre dunque chiedersi se ricorra il requisito alternativo dell'"accessorietà". Anche il significato di "accessorietà" non è affatto chiaro. Il concetto sembra infatti sovrapporsi con il requisito per cui le copie devono essere parte integrante del procedimento. La Corte interpreta il requisito di "accessorietà" nel senso che le copie, in quanto tali, non devono avere "né un'esistenza né una finalità autonome rispetto al procedimento"⁴⁶. In quest'ottica, fra i due requisiti c'è allora una differenza. Come già visto, "essere parte integrante del procedimento" si riferisce alla "nascita" della copia, nascita che deve avvenire in occasione del procedimento. "Accessorietà" si riferisce invece alla "vita" successiva della copia. La copia prodotta deve poter essere usata soltanto per il procedimento e non deve poter esistere al di fuori del procedimento⁴⁷. Per soddisfare il requisito, il *dataset* deve dunque essere costruito con *standard* o altre misure tecnologiche che ne vincolino l'uso al solo addestramento dell'IA.

5.5. Mancanza di rilievo economico autonomo.

Infine, l'eccezione richiede che le copie siano prive di rilievo economico proprio (condizione e). Secondo la Corte, ciò significa che la copia temporanea non deve essere in grado di generare un vantaggio economico "distinto e separabile" rispetto al vantaggio economico realizzato con l'utilizzo legittimo delle opere che è l'esito del procedimento⁴⁸. La Corte, in genere, considera soddisfatto il requisito

⁴⁶ CGUE, 5 giugno 2014, C-360/13, *Public Relations*, par. 47 ss.

⁴⁷ G. GUGLIELMETTI, (nt. 44), 35.

⁴⁸ CGUE, 17 gennaio 2012, C-302/10, *Infopaq*, par. 50 ss. Qui la Corte afferma che ha rilievo economico proprio anche la riproduzione che comporta «una modifica dell'oggetto riprodotto, quale esistente al momento dell'avvio del procedimento tecnologico interessato, poiché i suddetti atti sono, in tal caso, diretti a facilitare non già il suo utilizzo, ma l'utilizzo di un oggetto diverso». L'affermazione della Corte potrebbe essere letta nel senso che qualsiasi modifica nel formato o nel "linguaggio" della copia sia una modifica dell'opera e ricada fuori dall'ambito dell'eccezione. In tal caso, la riproduzione temporanea realizzata ai fini dell'addestramento si collocherebbe fuori dall'ambito di applicazione dell'art. 5, par. 1, visto che il procedimento in questione comporta generalmente la traduzione delle opere in un linguaggio digitale, per consentire la lettura degli algoritmi.

Questa lettura della sentenza non pare però convincente. Quasi sempre i procedimenti tecnologici telematici si fondano sulla creazione di copie in formati diversi da quelli originari. D'altra parte, la creazione di una copia temporanea serve, in molti casi, al solo scopo di cambiare forma all'opera e di consentirne così usi altrimenti impossibili. Se ogni cambiamento di formato o di "linguaggio" fosse sufficiente ad escludere l'applicazione dell'art. 5 par. 1, il diritto d'autore finirebbe per precludere numerosi procedimenti tecnologici innovativi. Il che pare entrare in conflitto con la *ratio* di

se l'unica utilità che la copia può generare consiste in una maggiore efficienza del procedimento in cui è inserita. Il che si verifica quando la copia è inseparabile dal procedimento. In altri termini, la copia è priva di rilievo economico quando non può essere né condivisa con terzi né usata in altri procedimenti.

Inteso in questo senso, il requisito appare senz'altro soddisfatto nel caso in cui le copie siano transitorie. Queste copie sono automaticamente distrutte una volta esaurito il loro ruolo nel procedimento. Esse non si prestano quindi ad essere condivise con i terzi. Potrebbero, tutt'al più, essere adoperate per il funzionamento di altri sistemi tecnologici controllati dal creatore del procedimento principale. Questi sistemi andrebbero però sincronizzati con quello principale per essere contemporanei, avere la stessa velocità di calcolo e disporre della stessa durata. Appare piuttosto improbabile che ciò sia fattibile per procedimenti completamente diversi da quello principale. È più verosimile che si tratti di procedimenti ancillari e collegati a quel procedimento. In questo caso, il vantaggio economico prodotto dalla copia sarebbe quindi pur sempre riconducibile al procedimento principale.

Nel campo dell'IA possono allora essere considerati privi di rilievo economico i *download* iniziali che siano soggetti a forme di cancellazione automatica, nonché le copie effimere eventualmente prodotte durante il *training*.

Come già detto, le copie non transitorie possono ancora essere giustificate se risultano accessorie. E ciò si verifica quando le copie possono essere utilizzate soltanto all'interno del procedimento. Anche queste copie dovrebbero, allora, essere incapaci di produrre un ricavo autonomo rispetto al procedimento in cui sono usate. In questo senso, però, la "mancanza di rilievo economico proprio" sarebbe una mera ripetizione di requisiti già esistenti e sarebbe quindi una condizione del tutto inutile.

Per risolvere questo problema occorre una precisazione sul significato di "accessorietà". Si è detto che "accessorietà" significa che la copia deve poter essere usata soltanto per un procedimento. Uno stesso procedimento può essere, però, realizzato più volte o da più soggetti diversi. Per es., un'impresa potrebbe ripetere più volte l'addestramento della propria IA oppure potrebbero esserci più imprese che applicano lo stesso tipo di procedimento per addestrare IA diverse. Il requisito

fondo dell'art. 5, par. 1, cioè la tutela dell'efficienza dinamica del mercato digitale (CGUE, 5 giugno 2014, C-360/13, *Public Relations*, par. 23). D'altra parte, una discriminazione tra procedimenti fondati su copie identiche all'originale e procedimenti fondati su copie basate su formati diversi non pare giustificata dall'esigenza di tutelare gli interessi degli autori. Questi interessi giustificerebbero, anzi, conclusioni opposte: la produzione di copie identiche all'originale è infatti verosimilmente più problematica per gli autori rispetto alla creazione di una copia in un formato pensato per la lettura artificiale.

Sembra dunque preferibile ritenere che la Corte abbia inteso qui escludere l'applicazione dell'art. 5, par. 1 ai casi in cui la copia modifichi la "sostanza" dell'opera, producendo, cioè, un'elaborazione che incide sul suo contenuto espressivo. In questo caso, in effetti, consentire la copia potrebbe finire per "liberalizzare" atti di elaborazione che gli ordinamenti nazionali generalmente riservano all'autore.

può essere allora letto in due modi. L'accessorietà al procedimento può essere intesa, in primo luogo, nel senso che la copia deve poter essere utilizzata soltanto nello specifico procedimento in cui è creata. Se fosse intesa in questo senso, però, una copia accessoria non potrebbe proprio essere utilizzata al di fuori di un determinato procedimento e non ci sarebbe, quindi, alcuna possibilità che essa generi un ricavo autonomo rispetto al procedimento stesso. L'accessorietà implicherebbe necessariamente mancanza di rilievo economico autonomo delle copie. E quest'ultimo requisito sarebbe, appunto, inutile.

In alternativa, l'accessorietà al procedimento può essere intesa nel senso che la copia deve poter essere utilizzata soltanto in un certo tipo di procedimento. In quest'ottica, è accessoria una copia che può essere utilizzata sia nel procedimento in cui è creata che in altre esecuzioni dello stesso procedimento poste in essere dal produttore o da altri operatori. Una copia del genere è suscettibile di più applicazioni ed è, quindi, anche in grado di generare un vantaggio economico indipendente rispetto al singolo procedimento in cui è creata. In quest'ottica, c'è allora una differenza tra "accessorietà" e "mancanza di rilievo economico": una copia "accessoria" può essere usata in una pluralità di procedimenti dello stesso tipo; una copia "priva di rilievo economico" può essere utilizzata soltanto nello specifico procedimento in cui è creata.

Da tutto questo discorso deriva che, per rispettare il requisito e), un *dataset* non deve soltanto essere incompatibile con utilizzi diversi dall'addestramento dell'IA. Deve, invece, addirittura, avere caratteristiche tali da poter essere applicato soltanto nello sviluppo di una determinata IA. Non basta allora che il *dataset* sia realizzato con un formato standard applicabile, in generale, a tutti i processi di addestramento. Il produttore dovrà costruire il *dataset* con formati, matrici o misure tecnologiche che assicurino che le copie siano utilizzabili soltanto dai propri algoritmi e che diventino obsolete dopo il loro primo utilizzo.

5.6. Considerazioni di sintesi.

In sintesi, l'eccezione di cui all'art. 5, par. 1 può essere applicata alle copie funzionali all'addestramento dell'IA se ricorrono le seguenti circostanze:

- le copie iniziali, prodotte al momento del *download*, e quelle generate durante il *training* sono eliminate automaticamente dopo la creazione delle versioni dei testi da inserire nel *dataset*;
- il *dataset* è costruito in occasione di uno specifico procedimento di addestramento.
- il *dataset* non è utilizzabile al di fuori di tale procedimento e non può essere offerto sul mercato;
- il *dataset* viene conservato solo per il tempo strettamente necessario a consentire l'addestramento dell'IA ed eventuali revisioni del sistema.

La disposizione non consente dunque alle imprese di IA di “rifornirsi” dai terzi. L’eccezione, infatti, non copre l’ipotesi in cui un operatore sviluppi il *dataset* per offrirlo ai programmatori di IA. Resta quindi sostanzialmente ostacolata la nascita di un mercato dei *dataset*. Cade poi anche fuori dall’eccezione il caso in cui il *dataset* sia creato da un’impresa su commissione del programmatore. In sostanza, il programmatore di IA deve occuparsi di tutta l’attività “a monte” dell’addestramento, e cioè della ricerca delle fonti, dello scaricamento dei testi, dell’adattamento dei file e della costruzione del *database*. Il che impedisce di beneficiare delle efficienze che derivano dall’emersione di imprese specializzate nelle fasi iniziali della catena e comporta un’inefficiente duplicazione delle reti di raccolta dei dati e dei sistemi di elaborazione dei *dataset*.

Inoltre, l’impresa che sviluppa l’IA non può costruire un *dataset* stabile e compatibile con una pluralità di procedimenti diversi. L’operazione di raccolta e di adattamento dei dati deve quindi sostanzialmente essere ripetuta per ogni procedimento. Questa soluzione porta con sé un’enorme duplicazione di costi. E ciò a tacere del problema di impatto “ambientale” derivante dalla ripetizione di operazioni potenzialmente dispendiose anche sul piano energetico.

In questo campo, l’art. 5, par. 1 non riesce dunque a realizzare efficacemente la sua funzione di conciliare la tutela del diritto d’autore con le esigenze dell’innovazione tecnologica. Nel momento in cui l’art. 5, par. 1 viene adottato, all’inizio degli anni Duemila, la tecnologia telematica è sostanzialmente concepita come un’immensa biblioteca di informazioni da ricercare, elaborare e scambiare. In quest’ottica, per consentire l’innovazione digitale basta, in effetti, “liberalizzare” le copie effimere necessarie alla navigazione degli utenti, alla visualizzazione dei contenuti e alla trasmissione delle informazioni.

Lo scenario è però cambiato radicalmente negli ultimi anni. La tecnologia digitale non è più soltanto un “oceano” passivo di informazioni su cui navigare. Con l’IA, la tecnologia digitale è diventata, a sua volta, un protagonista attivo della comunicazione telematica. L’IA è capace di esaminare i dati esistenti e di elaborarli, producendo informazioni e soluzioni nuove. E questo consente di raggiungere risultati che sarebbero impensabili in un mondo in cui protagonisti attivi della comunicazione sono soltanto gli utenti umani. Affinché questo nuovo protagonista del digitale possa funzionare è, però, necessario che le informazioni esistenti siano ad esso “comprensibili”. Da un punto di vista tecnico, questo significa che i contenuti della rete devono essere copiati e trasformati in un linguaggio accessibile alla macchina. E, come si è visto, la liberalizzazione delle copie effimere non è più sufficiente a questo fine.

In sintesi, l’art. 5, par. 1 è una disposizione pensata per un mondo che è ormai cambiato ed è oggi incompatibile con la realtà del mercato digitale. Ci sarebbero allora, teoricamente, le premesse per un tentativo di interpretazione evolutiva della disposizione. Questa via non sembra, però, percorribile. V’è, in primo luogo, il problema che l’art. 5, par. 1 si fonda sul requisito che le copie siano “tempora-

nee”, mentre l’IA ha bisogno di copie (almeno in certa misura) “permanenti”. Per adeguare la disposizione a questa esigenza, si dovrebbe assegnare al termine “temporaneo” un significato che questo non può proprio assumere⁴⁹. Un’interpretazione evolutiva dell’eccezione richiederebbe poi che le esigenze della realtà attuale non siano state già prese in considerazione dal legislatore. Tuttavia, come già visto, la direttiva 2019/790 ha in qualche misura disciplinato la materia attraverso le eccezioni per l’estrazione di dati, che realizzano una liberalizzazione limitata degli atti di riproduzione. Gli ostacoli che l’IA affronta nell’attuale quadro normativo non dipendono allora tanto da un ritardo nell’aggiornamento delle regole, ma sembrano il frutto di una vera e propria scelta del legislatore. Una lettura dell’art. 5, par. 1 nel senso che siano consentite senza limiti tutte le riproduzioni necessarie per addestrare un’IA si porrebbe in contraddizione con questa scelta.

6. I confini del diritto di riproduzione.

L’idea che il sistema delle eccezioni al diritto di riproduzione non offra soluzioni soddisfacenti per l’IA è condivisa da gran parte della dottrina. E ciò ha contribuito a rafforzare la convinzione, da tempo diffusa, che, per riconciliare il diritto d’autore con gli obiettivi generali di efficienza dinamica del mercato digitale, sia necessaria una modifica delle direttive⁵⁰.

Alcuni autori propongono di espandere le eccezioni al diritto d’autore, vuoi introducendo una clausola generale di esenzione, sulla falsariga del “*fair use*” statunitense, vuoi aggiungendo una specifica eccezione che consenta tutte le forme di riproduzione necessarie per il funzionamento di sistemi tecnici innovativi, come l’IA⁵¹.

In base ad un’altra impostazione, anziché intervenire sulle eccezioni, occorrerebbe modificare il diritto di riproduzione, limitandolo alle copie che sono desti-

⁴⁹ Sul punto v. anche M. BORGHI, S. KARAPAPA, *Copyright and mass digitization*, Oxford, Oxford UP, 2013, 59; G. GHIDINI, *Proprietà intellettuale e innovazione digitale. Dalla “interferenza antitrust” a un nuovo paradigma?*, in *Giur. Comm.*, 2023, I, 367 ss.

⁵⁰ Si v. ad es. B. HUGENHOLTZ, *The new copyright directive: text and data mining (articles 3 and 4)*, in *Kluwer Copyright Blog*, 24 luglio 2019; G. FROSIO, *Should we ban generative AI, incentivize it or make it a medium for inclusive creativity?*, su *ssrn.com*, 2023, 12; G. ROSSI, (nt. 37), 247 ss.; M.L. MONTAGNANI, G. AIME, (nt. 11); K. CHRISTENSEN, *A European solution for text and data mining in the development of creative artificial intelligence*, in *Stockholm IP L. Rev.*, 2021, 18 ss.; R. CASO, *Il conflitto tra diritto d’autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, Trento Law and Technology Research Group, 2020.

⁵¹ Si v. tra gli altri C. GEIGER, V. IAIA, (nt. 15), 10 ss.; G. GHIDINI, (nt. 49), 367 ss. V., con riferimento in generale, agli usi digitali B. HUGENHOLTZ, M. SENFTLEBEN, *Fair use in Europe: in search of flexibilities*, IVIR Report, Amsterdam, 2011; I. HEARGRAVES, *Digital opportunity*, Report, Maggio 2011, 46 ss.; C. GEIGER, E. IZYUMENKO, *Towards a European “fair use” grounded on the freedom of expression*, in *American Univ. Int. L. Rev.*, 2019, 1 ss.

nate alla distribuzione ai consumatori finali. Dovrebbero essere, cioè, sottratte all'esclusiva tutte le copie "intermedie", cioè le copie che sono prodotte per essere adoperate all'interno di un processo tecnologico⁵².

Secondo alcuni autori, peraltro, questo risultato sarebbe in realtà già raggiungibile a livello interpretativo. Per riproduzione di un'opera si intende la moltiplicazione della stessa in copie. Il concetto di copia non è però definito nelle direttive e deve essere interpretato alla luce dell'obiettivo del diritto di riproduzione di assegnare all'autore un controllo sulla diffusione commerciale degli esemplari presso il pubblico. "Copie" sarebbero, allora, soltanto gli esemplari che sono destinati, in un modo o nell'altro, alla circolazione presso i consumatori finali; non quelli volti ad agevolare il funzionamento di procedimenti tecnici, che nulla hanno a che vedere con la "vita" commerciale dell'opera. La creazione di queste ultime copie e il loro eventuale trasferimento non richiederebbero quindi alcuna autorizzazione da parte dei titolari⁵³.

⁵² V. T. MARGONI, M. KRETSCHMER, (nt. 15), 693 ss.; R. DUCATO, A. STROWEL, (nt. 21), 24 ss., secondo cui l'espansione del diritto di riproduzione alle copie "intermedie" non è compatibile con gli obiettivi di fondo del diritto d'autore. V. in senso simile A. STROWEL, *Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals*, *Copyright reconstructed*, edito da Hugenholtz, Milano, Wolters Kluwer, 2018, 204; S. DUSOLIER, *Realigning Economic Rights with Exploitation of Works: The Control of Authors over the Circulation of Works in the Public Sphere*, in *Copyright reconstructed*, edito da Hugenholtz, Milano, Wolters Kluwer, 2018, 163 ss. e M. BORGHI, S. KARAPAPA, (nt. 49), 52 ss. e 153 ss., in cui gli A. esprimono proposte simili, effettuando anche un confronto dettagliato con la disciplina generale in tema di uso dei dati. In generale, l'idea che la tradizionale esclusiva sulla riproduzione richieda un ripensamento nel mondo digitale è da tempo diffusa a livello internazionale. Si v. ad es. J. LITMAN, *Real copyright reform*, in *Iowa L. Rev.*, 2010, 41 ss.; M. LEMLEY, *Dealing with overlapping copyrights on the Internet*, in *Univ. Dayton L. Rev.*, 1997, 22 ss.; P. SPADA, *La proprietà intellettuale nelle reti telematiche*, in *Riv. dir. civ.*, 1998, 636 ss.

⁵³ Si v. D. SCHÖNBERGER, (nt. 27), 13; *Study on copyright and new technologies: copyright data management and artificial intelligence*, (nt. 10), 182 ss.; M. SENFTLEBEN, (nt. 29), 1483 ss., in cui la questione è affrontata dal punto di vista delle convenzioni internazionali in materia di diritto d'autore. L'idea che le copie meramente tecniche cadano al di fuori dell'esclusiva è, comunque, presente nella dottrina europea anche prima della rivoluzione dell'IA. Si v. ad es. B. HUGENHOLTZ, (nt. 13), 482 ss.; L. SCHIUMA, *Diritto d'autore e normativa europea*, in *Treccani – diritto online*, 2009; A. MUSSO, *Diritto di autore sulle opere dell'ingegno letterarie ed artistiche*, in *Comm. Scialoja-Branca*, Zanichelli, 2008, 212 ss. Sull'esigenza di effettuare una valutazione di carattere "teleologico" in tema di riproduzioni digitali, v. R. ROMANO, *L'opera e l'esemplare nel diritto della proprietà intellettuale*, Padova, Cedam, 2001, 186 ss.

In tema di IA, la soluzione viene spesso sostenuta richiamando la distinzione che viene effettuata negli Stati Uniti tra utilizzi dotati di "expressive purposes" e utilizzi dotati di "non-expressive purposes". Questi ultimi vengono talora qualificati come "transformative uses". Da ciò parte della dottrina trae la conclusione che le copie intermedie realizzate per l'addestramento siano giustificate dalla clausola di "fair use". V. in questo senso, tra gli altri, M. SAG, (nt. 2), 2023, 104 ss.; M. LEMLEY, B. CASEY, (nt. 2), 744 ss. V. per un'opinione, in parte, diversa B. L. W. SOBEL, (nt. 39), 46 ss. La questione è oggetto dei procedimenti in corso. V. ad es. *Thomson Reuters, v. Ross Intelligence, Inc.*, 28 settembre 2023, 20-cv-613-SB (D. Del. Sep. 28, 2023). Una parte della dottrina statunitense

Quest'approccio è stato, a sua volta, criticato. Le tesi in esame fanno dipendere la qualifica di "copia" dallo scopo per cui un esemplare è prodotto. Il diritto di riproduzione è oggi definito dalle direttive come «il diritto esclusivo di autorizzare o vietare la riproduzione diretta o indiretta, temporanea o permanente, in qualunque modo o forma, in tutto o in parte». La definizione, molto ampia, non contiene alcun riferimento alla funzione per cui una copia viene prodotta. Il fatto che essa sia realizzata per uno scopo commerciale o per uno scopo tecnico sarebbe dunque irrilevante⁵⁴. Peraltro, le copie oggetto delle eccezioni di estrazione e di riproduzione temporanea sono sempre copie "intermedie" rispetto a procedimenti tecnici. Se fosse vero che queste copie si collocano fuori dall'esclusiva, le relative eccezioni sarebbero disposizioni inutili. L'idea che la nozione di "copia" dipenda dallo scopo della riproduzione sembra, dunque, incompatibile con le attuali regole sul diritto d'autore.

Di queste tesi appare però condivisibile l'idea che il concetto di "copia" debba essere interpretato tenendo conto degli obiettivi dell'esclusiva sulla riproduzione. Le ricostruzioni sopra richiamate partono dall'idea che il diritto di riproduzione abbia la funzione di assegnare all'autore il controllo sulla successiva circolazione delle copie sul mercato⁵⁵. Storicamente, il diritto di riproduzione si giustifica soprattutto con il fatto che è più semplice agire nei confronti delle stamperie abusive, rispetto ad agire nei confronti dei successivi atti di smercio delle copie. Atti, questi, che possono essere numerosi e difficili da individuare⁵⁶. Peraltro, questa funzione non riguarda soltanto la distribuzione delle copie materiali, ma vale anche per molte altre forme di sfruttamento commerciale dell'opera. Basti pensare, ad es., alla registrazione abusiva di un'opera musicale. Questa può essere distribuita sul mercato sotto forma di CD, può essere messa in onda sulla radio o sulla televisione, può essere diffusa nei pubblici esercizi, ecc. Agendo "a monte" con-

ha adoperato argomenti simili per sostenere che gli esemplari privi di "*expressive purposes*" si collocano del tutto al di fuori del diritto di riproduzione. M.W. CARROLL, *Copyright and the progress of science: why text and data mining is lawful*, in *U.C. Davis L. Rev.*, 2019, 894 ss., in cui si distingue tra "*copies that count*" e "*copies that don't count*". V. in senso simile J. QUANG, *Does training AI violate copyright law?*, in *Berkeley Tech. L. J.*, 2021, 1407 ss.

⁵⁴ A. OTTOLIA, (nt. 33), 394 ss. Il punto è confermato dalla giurisprudenza europea. V. CGUE, 4 ottobre 2011, cause riunite C-403/08, C-429/08, *FA Premier League*, 159; CGUE, 24 marzo 2022, C-433/20, *Austro Mechana*, par. 16 ss.

⁵⁵ Più precisamente, l'esclusiva sulla riproduzione è ricollegata in dottrina all'esigenza di assegnare all'autore il potere di decidere il numero degli esemplari in circolazione. Il che gli consente di esercitare un potere "monopolistico" sul mercato dell'opera, influenzando il prezzo degli esemplari. Si v. R. ROMANO, (nt. 53), 159; D. SARTI, (nt. 33), 358 ss., in cui, più precisamente, l'A. individua in via interpretativa l'esistenza di un diritto a stabilire la quantità di prodotti destinati al mercato. V. anche P. AUTERI, *Diritto di autore*, in AA.VV., *Diritto industriale*, Giappichelli, 2023, 717 ss.

⁵⁶ J.H. SPOOR, *The copyright approach to copying on the Internet: (over)stretching the reproduction right?*, in *The future of copyright in the digital environment*, edito da Hugenholtz, Wolters Kluwer, 1996, 77.

tro l'atto di registrazione, l'autore evita di dover agire contro tutte le possibili utilizzazioni della stessa.

Questo discorso potrebbe, in effetti, portare alla conclusione che le "copie", ai sensi del diritto di riproduzione, siano soltanto quelle destinate allo sfruttamento commerciale e, cioè, alla diffusione presso il pubblico. Tuttavia, il diritto di riproduzione ha subito un'evoluzione nel corso tempo.

A partire dagli anni '60, si diffondono le tecnologie di fotocopia e di registrazione analogica. Ciò semplifica la produzione di copie destinate al mercato, ma consente anche al grande pubblico di creare copie delle opere per uso personale a costi marginali. Il dilagare di queste copie "private" rischia di ridurre la domanda delle copie "commerciali". Con il tempo, il diritto di riproduzione finisce quindi per coprire anche le copie destinate ad usi "non commerciali". Data la difficoltà pratica di esercitare l'esclusiva nei confronti degli usi "privati", il diritto esclusivo in questo campo viene ben presto sostituito con diritti al compenso affidati alla gestione delle *collecting societies*. L'obiettivo è, in ogni caso, quello di proteggere il mercato dell'opera dalla concorrenza data dalla diffusione dell'attività privata di copia. Per il diritto d'autore sono dunque copie anche quelle destinate ad uno sfruttamento "domestico"⁵⁷.

Con la rivoluzione digitale, poi, l'attività di copia diventa ancora più accessibile. A questo punto, chiunque può partire dalla copia di un'opera per produrre un numero infinito di esemplari a costi pressoché nulli. Con Internet diventa poi anche molto semplice trasmettere le copie al pubblico. Ogni copia può facilmente essere caricata e diffusa su diversi siti-*web*. Gli utenti di questi siti possono, a loro volta, creare nuove copie da diffondere, e così via. In sostanza, sul *web* una singola copia può dare luogo ad un'infinita serie di trasmissioni. Trasmissioni, peraltro, molto difficili da individuare e da contrastare. Ogni copia digitale porta quindi con sé un pericolo per i mercati "ufficiali" dell'opera. In questo contesto, il diritto di riproduzione viene esteso fino a coprire, come si è visto, anche le copie "intermedie". Queste copie hanno funzioni meramente tecniche e non sono destinate né alla diffusione commerciale né alla fruizione privata degli utenti. Eppure, nel mondo digitale esse possono essere facilmente distolte dalla loro funzione originaria e trasmesse al pubblico. Di qui l'esigenza di sottoporle al controllo dell'autore⁵⁸. Queste copie vengono consentite soltanto in presenza di condizioni che

⁵⁷ Su questa evoluzione si v. M. RICOLFI, *Il diritto d'autore*, in *Tr. dir. comm.*, diretto da G. COTTINO, II, Padova, Cedam, 2001, 415; D. SARTI, *Copia privata e diritto d'autore*, in *AIDA*, 1992, 35 ss. P. AUTERI, (nt. 55), 717 ss.; J. H. SPOOR, (nt. 56), 70 ss.

⁵⁸ M. RICOLFI, (nt. 57), 418 ss. L'estensione del diritto di riproduzione alle copie incidentali è oggetto di accese discussioni durante la negoziazione delle convenzioni internazionali in materia. Si v. per una ricostruzione M. SENFTLEBEN, (nt. 29), 1483 ss. Sull'evoluzione della nozione di riproduzione con l'avvento del digitale si v. R. ROMANO, (nt. 53), 159 ss. e 181 ss., in cui si mette in luce il fatto che, con la tecnologia telematica, la distinzione tradizionale tra creazione della copia e sua successiva distribuzione viene meno. Fenomeno, questo, che diventa particolarmente evidente in mate-

ne rendano impossibile lo sfruttamento commerciale. È, appunto, il caso dell'art. 5, par. 1 sulle copie temporanee.

Il diritto di riproduzione ha dunque assunto nel tempo funzioni diverse. Il problema di fondo che esso affronta è, però, sempre lo stesso. Se fosse libera la produzione di copie, sarebbe impossibile per il titolare esercitare un controllo sul mercato dell'opera. Il mercato verrebbe facilmente "inondato" di esemplari non autorizzati e i diritti esclusivi sulle varie forme di sfruttamento commerciale sarebbero, di fatto, svuotati di efficacia. Il diritto di riproduzione serve quindi a proteggere i mercati che la legge riserva all'autore. In questo senso, si tratta di un diritto strumentale rispetto a tutti gli altri diritti esclusivi dell'autore.

È alla luce di questa funzione che pare doversi leggere il concetto di "copia": per il diritto d'autore "copia" non è soltanto l'esemplare prodotto per la distribuzione commerciale, ma qualsiasi esemplare che abbia le caratteristiche per entrare in concorrenza con le forme di sfruttamento economico che la legge riserva al titolare, indipendentemente dallo scopo per cui l'esemplare è originariamente creato⁵⁹.

Tornando all'IA, si tratta dunque di capire se le copie che vengono prodotte per l'addestramento siano idonee a "minacciare" i mercati oggetto di esclusiva.

A questo proposito, va detto che i sistemi di IA sono molto numerosi e presentano significative differenze. Il loro funzionamento è spesso segreto e non è dunque possibile conoscere con precisione ogni aspetto del procedimento. È perciò impossibile dare alla domanda una risposta definitiva, valida per ogni sistema di IA. Si può però svolgere qualche osservazione di taglio generale, da cui trarre criteri per l'applicazione al caso concreto.

Se nel *dataset* i testi sono conservati nel linguaggio e nel formato originario o, comunque, in un formato facilmente utilizzabile per fini diversi dall'addestramento, i testi sono potenzialmente utilizzabili sul mercato della diffusione

ria di protezione dei programmi per elaboratore. Si v. anche L. SCHIUMA, (nt. 53), secondo cui la disintermediazione dell'attività di copia resa possibile dalle tecnologie digitali giustifica un ripensamento nell'estensione del diritto di riproduzione.

⁵⁹ Si v. sul punto M. RICOLFI, (nt. 57), 415 ss.; A. MUSSO, (nt. 53), 205; B. HUGENHOLTZ, (nt. 13), 482 ss., anche per riferimenti ad opere precedenti. Si v. anche M. SENFTLEBEN, (nt. 29), 1497, secondo cui simili proposte di definizione del concetto di "copia" sono più volte emerse nel dibattito internazionale.

A prima vista, la conclusione raggiunta nel testo si scontra con l'art. 68 l. aut., in cui si legge che «è libera la riproduzione di singole opere o brani di opere per uso personale dei lettori, fatta a mano o con mezzi di riproduzione non idonei allo spaccio o diffusione dell'opera nel pubblico». Il che pare presupporre che, in linea di principio, anche le riproduzioni inidonee alla distribuzione commerciale siano "copie" per il diritto d'autore. In realtà, l'eccezione non pare incoerente con quanto sopra detto. Le copie dell'art. 68 non sono idonee al commercio per il mezzo con cui sono realizzate, ma sono pur sempre esemplari leggibili da parte del pubblico. In teoria, esse possono circolare tra gli utenti ed interferire, così, con la domanda delle copie commerciali. In questo senso, esse possono effettivamente rientrare nel concetto di "copia" sopra richiamato. La potenzialità lesiva di queste copie è poi, di fatto, molto ridotta e ciò giustifica l'esistenza di un'apposita eccezione.

dell'opera al pubblico. In questo caso, non c'è dubbio quindi che il *dataset* sia composto da "copie" ai sensi del diritto d'autore.

Nella maggior parte dei casi, però, l'addestramento dell'IA non avviene su *file* di questo genere. Il testo nella sua forma originaria contiene molti elementi che "disturbano" la comprensione degli algoritmi dell'IA, come gli articoli, la punteggiatura, le espressioni non chiare, gli errori, ecc. Peraltro, i formati leggibili per l'uomo non si prestano ad un'analisi efficace da parte degli algoritmi. Per ottimizzare il processo di estrazione di dati, innanzitutto, i testi devono essere suddivisi in unità più piccole, ad es., parole o espressioni. È il c.d. processo di "tokenizzazione" del testo. Le singole unità sono poi talora "asciugate" da alcune variabili che complicano la raccolta di dati. È ad es. il caso del verbo coniugato che può essere riportato ad una versione base, come il verbo all'infinito. Inoltre, a seconda dei casi, le unità di testo possono essere mantenute nell'ordine originario oppure "mescolate" con unità simili provenienti da altri testi. Generalmente, le parole vengono poi tradotte in vettori matematici per la lettura algoritmica. Questa complessa attività di intervento sul testo viene chiamata "normalizzazione" o "*pre-processing*"⁶⁰ ed è volta ad ottimizzare l'elaborazione dei dati da parte dell'algoritmo.

Beninteso, il "*pre-processing*" può avvenire in modi diversi. Ci sono casi in cui la lavorazione è limitata e il testo resta quindi, almeno in certa misura, utilizzabile per fini diversi dall'apprendimento artificiale. In altri casi, invece, il testo è sottoposto ad un profondo processo di manipolazione, fino a trasformare la forma originaria delle opere in immense serie numeriche. Queste forme di intervento sottraggono alla copia i connotati tipici di un testo destinato alla fruizione del pubblico. C'è chi parla, al riguardo, di una vera e propria trasformazione dell'opera in "dato". Ed in effetti il *pre-processing* può anche essere visto come una prima forma di produzione di metadati sull'opera.

In questi ultimi casi, il file non può essere immediatamente utilizzato dagli utenti per accedere all'opera originaria. A tal fine, l'utente dovrebbe disporre della matrice di decodifica e dovrebbe conoscere con precisione tutte le fasi di manipolazione che il testo ha subito. Peraltro, anche se l'utente disponesse di queste informazioni, la ricompilazione di un *dataset* sarebbe difficile da realizzare con i comuni programmi operativi. Teoricamente, si potrebbe realizzare utilizzando specifici programmi di riconversione. Questa operazione, richiederebbe, comunque, notevole spazio di memorizzazione e strumenti tecnici dotati di elevata capacità di calcolo. Per l'applicazione di sofisticati sistemi di decodificazione è anche necessaria una preparazione tecnica in tema di programmazione e di analisi dei dati. Infine, i processi di normalizzazione sono piuttosto vari, sicché l'uso di programmi di riconversione standard non può sempre garantire l'efficace trasforma-

⁶⁰ M.L. MONTAGNANI, G. AIME, (nt. 11), 378.

zione del file in un testo fruibile. Dal punto di vista del pubblico, un file soggetto a profondi processi di normalizzazione non sembra quindi potersi considerare, di per sé, sostituibile alle normali copie digitali.

Una volta assunta la forma “*pre-processed*”, il file può essere utilizzato soltanto per l’analisi da parte degli algoritmi dell’IA. Questa ovviamente è, a sua volta, una forma di utilizzazione commerciale del testo. Come già detto, però l’analisi computazionale ha ad oggetto soltanto l’opera come veicolo di informazioni, non l’espressione creativa dell’autore. Si tratta quindi di un’attività che cade fuori dall’ambito dell’esclusiva⁶¹. I file normalizzati, in quanto tali, non interferiscono allora con i mercati che la legge riserva all’autore. Le tesi dottrinali sopra richiamate potrebbero allora essere lette nel senso che siano questi particolari file a cadere fuori dal concetto di “copia” e dall’ambito dell’esclusiva autoriale⁶².

Questa conclusione si espone ad alcune obiezioni. Se è vero che il file “normalizzato” non è direttamente accessibile per gli utenti, l’algoritmo dell’IA è però tendenzialmente in grado di risalire al testo originario. Esso ha infatti la capacità computazionale per individuare le matrici e per ricostruire le varie fasi di manipolazione e di interpolazione subite dal testo di base. Se fosse programmata per questo scopo, un’IA sarebbe quindi in grado di ritrasformare un file normalizzato in un testo fruibile e di trasmetterlo al pubblico⁶³.

La situazione delle copie normalizzate non è allora poi tanto diversa dal caso di altri esemplari che sono di per sé inutilizzabili, ma che diventano fruibili se sottoposti ad ulteriori passaggi tecnici. È il caso, ad es., della diffusione dei contenuti *torrent*, in cui la copia subisce una scomposizione in fase di trasmissione e viene poi “riasmblata” per consentire il *download* e l’accesso all’opera da parte degli utenti. D’altra parte, anche il CD audio, da solo, non è in grado di comunicare l’opera; se combinato con un apposito strumento di “lettura” diventa però a tutti gli effetti veicolo dell’opera. Tutte queste situazioni sono considerate riproduzioni

⁶¹ In questo senso, pare di poter leggere il ragionamento dell’ordinanza *Thomson Reuters, v. Ross Intelligence, Inc.*, 28 settembre 2023, 20-cv-613-SB (D. Del. Sep. 28, 2023): «if Ross’s characterization of its activities is accurate, it translated human language into something understandable by a computer as a step in the process of trying to develop a “wholly new,” albeit competing, product — a search tool that would produce highly relevant quotations from judicial opinions in response to natural language questions. This also means that Ross’s final product would not contain or output infringing material».

⁶² Indicazioni in questo senso (riferite però al diritto connesso del produttore di fonogrammi) paiono potersi leggere in CGUE, 29 luglio 2019, C-476/19, *Pelham*, par. 31: «quando un utente, nell’esercizio della libertà delle arti, preleva un campione sonoro da un fonogramma al fine di utilizzarlo, in una forma modificata e non riconoscibile all’ascolto, in una nuova opera, si deve ritenere che un utilizzo del genere non costituisca una «riproduzione», ai sensi dell’articolo 2, lettera c), della direttiva 2001/29».

⁶³ Considerazioni simili sono espresse, seppur con riferimento alle prime forme di sviluppo del mercato digitale, in G. GUGLIELMETTI, (nt. 44), 17 ss.

ai sensi del diritto d'autore⁶⁴. Il che depone a favore della tesi secondo cui anche un file “*pre-processed*” rientra nel concetto di copia.

In realtà, però, i file normalizzati dell'IA presentano delle peculiarità rispetto a queste situazioni. Le copie *torrent* e le riproduzioni meccaniche sono pur sempre copie prodotte allo scopo di realizzare la diffusione dell'opera al pubblico. La forma “intermedia” che si assegna alla copia è, anzi, proprio funzionale a rendere più efficace o più rapida la comunicazione del suo contenuto. In questi casi, si può dunque presumere che alla produzione e alla diffusione della copia facciano seguito atti di sfruttamento dell'opera riservati all'autore.

Lo stesso non pare potersi dire per i file normalizzati dell'IA. Questi file non sono costruiti per consentire la trasmissione al pubblico dei contenuti. Al contrario, il loro formato rende la comunicazione del contenuto più difficile. Questa comunicazione è pure tecnicamente possibile, ma richiede un'azione ulteriore che è avulsa rispetto alla finalità per cui il file è originariamente generato e che è volta proprio ad indirizzare la copia verso una finalità diversa. In queste circostanze, si pone, allora, quanto meno il problema se sia corretto intervenire al momento della produzione del file o piuttosto al momento in cui si verifica il passaggio tecnico che cambia la “vocazione” del file, rendendolo idoneo a ledere gli interessi dei titolari. E, in ogni caso, la riconversione del file è un passaggio del tutto eventuale, tanto più che allo stato si tratta di un'operazione che non sembra alla portata di un utente qualsiasi. Qui non si può quindi presumere che alla creazione del file segua il suo sfruttamento per fini di trasmissione dell'opera al pubblico. Infine, anche se l'utilizzo per scopi di trasmissione del testo fosse fattibile per tutti gli utenti della rete, resta il fatto che questi file sono prodotti per uno scopo diverso e legittimo, vale a dire quello di consentire l'analisi algoritmica. Rispetto alle altre copie, le quali, salvo rare eccezioni, sono sempre destinate ad utilizzi coperti da esclusiva, questi file sono suscettibili di diversi utilizzi, alcuni riservati e altri legittimi.

⁶⁴ In realtà, la trasmissione *torrent* è stata affrontata dalla giurisprudenza europea soprattutto dal punto di vista della comunicazione al pubblico interattiva. Si v. CGUE, 14 giugno 2017, C-610/15, *Stichting Brein* e soprattutto CGUE, 17 giugno 2021, C-597/19, *Mircom*. In quest'ultimo caso si affrontava, tra l'altro, la questione se l'utente che partecipa ad una trasmissione *torrent* commetta atti di violazione del diritto d'autore. La questione pregiudiziale faceva riferimento esclusivamente al diritto di comunicazione al pubblico. Nella sua motivazione, la Corte afferma inizialmente che «alla Corte spetta, se necessario, riformulare le questioni che le sono sottoposte. Infatti, la Corte ha il compito di interpretare tutte le disposizioni del diritto dell'Unione che possano essere utili ai giudici nazionali al fine di dirimere le controversie di cui sono investiti, anche qualora tali disposizioni non siano espressamente indicate nelle questioni a essa sottoposte da detti giudici». È interessante notare che la Corte non richiama a questo proposito le disposizioni sul diritto di riproduzione. Essa conclude nel senso che «costituisce una messa a disposizione del pubblico, ai sensi di tale disposizione, il caricamento, a partire dall'apparecchiatura terminale di un utente di una rete tra pari (peer-to-peer) verso apparecchiature terminali di altri utenti di tale rete, dei segmenti, previamente scaricati da detto utente, di un file multimediale contenente un'opera protetta, benché tali segmenti siano utilizzabili da soli soltanto a partire da una determinata percentuale di scaricamento».

C'è poi anche un'altra considerazione che vale a distinguere il caso dell'IA dalle copie *torrent* e dalle riproduzioni meccaniche sopra richiamate. Proprio perché queste ultime sono copie destinate a trasmettere l'opera al pubblico, gli applicativi tecnici necessari per "leggerle" sono in genere alla portata di tutti. Pertanto, se la creazione delle copie fosse libera, per il titolare dei diritti sarebbe poi estremamente difficile agire nei confronti dei successivi atti di sfruttamento, che potrebbero anche essere posti in essere direttamente dagli utenti finali in maniera "decentrata". In sintesi, il pericolo che copie del genere ledano gli interessi dell'autore è piuttosto elevato e ciò giustifica l'anticipazione dell'esclusiva al momento della creazione della copia. Anche questo argomento non sembra valido, però, in tema di IA. Qui l'atto che rende fruibile l'opera per il pubblico richiede sempre l'intervento "centrale" di un'IA o, comunque, di un servizio digitale in grado di riconvertire il testo. Questo atto di conversione del file da parte di un'IA comporta la creazione di una versione fruibile dell'opera ed è quindi sicuramente da riguardare come un atto di "copia" coperto dall'esclusiva. Peraltro, se alla decodifica segue la diffusione del testo agli utenti si ha anche un atto di comunicazione al pubblico, anch'esso oggetto di esclusiva. In altre parole, la situazione di un'IA che fosse programmata per restituire agli utenti opere da leggere non sarebbe diversa da quella di una piattaforma "pirata" che mette opere a disposizione del pubblico senza autorizzazione⁶⁵. A fronte di un uso "abusivo" dei file norma-

⁶⁵ Questa conclusione è valida soprattutto nel caso in cui l'IA sia programmata per replicare testi "memorizzati" in fase di addestramento. L'IA si comporterebbe qui in maniera non molto diversa da un qualunque sito di trasmissione di opere "pirata". Si tratta di un problema che è emerso nell'ambito delle controversie tra editori e ChatGPT. In particolare, certi editori affermano che i meccanismi statistici su cui si fonda ChatGPT fanno sì che il sistema replichi integralmente gli articoli di giornale usati in fase di addestramento, se l'utente fornisce come input le prime righe del testo. In questo modo, ChatGPT diventa un meccanismo per aggirare le restrizioni poste dagli editori per i consumatori non abbonati al sito.

Qualche dubbio sulla responsabilità dell'operatore di IA potrebbe porsi invece nell'eventualità in cui il servizio fosse programmato semplicemente per convertire *file* normalizzati forniti dagli utenti in testi facilmente fruibili dal pubblico. A prima vista, in questo caso, l'IA si limita a fornire la infrastruttura tecnica, mentre l'atto di *upload* viene effettuato dagli utenti del servizio. La fattispecie sembrerebbe allora avvicinarsi alle situazioni in cui un operatore offre in maniera meramente passiva un mezzo di comunicazione che viene poi adoperato dagli utenti per violare diritti d'autore. In realtà, però, se è vero che in questo caso l'iniziativa dell'atto di trasmissione è presa dall'utente, resta fermo il fatto che la comunicazione finale del testo fruibile viene effettuata dal sistema di IA. Comunque, appare anche discutibile qualificare come infrastruttura meramente passiva un sistema di IA appositamente programmato per offrire servizi di riconversione di file in testi fruibili dal pubblico. Sembra più convincente ritenere che un operatore che offre un'infrastruttura del genere al pubblico sia, a sua volta, responsabile della comunicazione resa possibile dal proprio servizio. In questo senso, con riferimento alle piattaforme "*peer to peer*" v. CGUE, 14 giugno 2017, C-610/15, *Stichting Brein*, par. 36 ss.: «le opere così messe a disposizione degli utenti della piattaforma di condivisione online TPB sono state messe online su tale piattaforma non dagli amministratori di quest'ultima, bensì dai suoi utenti. Tuttavia detti amministratori, mediante la messa a disposizione e la gestione di una piattaforma di condivisione online, come quella di cui al procedimento principale,

lizzati, il titolare non sarebbe allora costretto ad agire nei confronti di una massa dispersa di soggetti, ma potrebbe agire direttamente nei confronti dell'operatore dell'IA⁶⁶.

In definitiva, l'analogia con le copie sopra richiamate non sembra sufficiente, di per sé, a concludere che le copie normalizzate dell'IA siano riproduzioni ai sensi del diritto d'autore. In ogni caso, resta ferma l'obiezione di fondo: questi file possono dare luogo indirettamente a forme di fruizione dell'opera. Si tratta quindi di capire se sia copia ai sensi del diritto d'autore una versione dell'opera che: a) non è immediatamente in grado di trasmettere l'espressione creativa dell'autore al pubblico; b) può dare luogo indirettamente ad alcuni utilizzi concorrenti con lo

intervengono con piena cognizione delle conseguenze del proprio comportamento, al fine di dare accesso alle opere protette, indicizzando ed elencando su tale piattaforma i file torrent che consentono agli utenti della medesima di localizzare tali opere e di condividerle nell'ambito di una rete tra utenti (peer-to-peer). A tale riguardo [...] senza la messa a disposizione e la gestione da parte dei suddetti amministratori di una siffatta piattaforma, le opere in questione non potrebbero essere condivise dagli utenti o, quantomeno, la loro condivisione su Internet sarebbe più complessa. Occorre pertanto considerare che, con la messa a disposizione e la gestione della piattaforma di condivisione online TPB, gli amministratori di quest'ultima offrono ai loro utenti un accesso alle opere di cui trattasi. Si può quindi ritenere che essi svolgano un ruolo imprescindibile nella messa a disposizione delle opere in questione». V. anche in tema di piattaforme CGUE, 22 giugno 2021, cause riunite C-682/18 e C-683/18, *Peterson*.

⁶⁶ A questo ragionamento si potrebbe replicare che, in realtà, in futuro potrebbero essere sviluppati applicativi IA in grado di riprodurre le opere o di decodificare le copie a livello "locale", cioè direttamente dagli utenti sui propri dispositivi personali. Il che aprirebbe effettivamente la porta al rischio che gli atti di fruizione avvengano anche qui in maniera decentrata. D'altra parte, ai sensi del considerando 27 della direttiva 2001/29, «la mera fornitura di attrezzature fisiche atte a rendere possibile o ad effettuare una comunicazione non costituisce un atto di comunicazione ai sensi della presente direttiva». E ciò potrebbe teoricamente essere letto nel senso che l'operatore che offre al pubblico l'applicativo non sia poi responsabile dell'eventuale sfruttamento abusivo dello stesso da parte degli utenti. Questa conclusione non pare però condivisibile. Il considerando 27 ha l'obiettivo di evitare che si possa contestare una violazione dei diritti d'autore a soggetti che offrono normali infrastrutture tecniche suscettibili di molteplici utilizzi leciti. Diverso è invece il caso di un soggetto che distribuisce un applicativo appositamente programmato per dare abusivamente accesso alle opere. Qui sembra più convincente ritenere che l'operatore sia, a sua volta, responsabile dello sfruttamento illecito delle opere. Si v. CGUE, 26 aprile 2017, C-527/15, *Stichting Brein*, par. 53, in cui si legge che il diritto di comunicazione al pubblico «ricomprende la vendita di un lettore multimediale, come quello di cui al procedimento principale, nel quale sono state preinstallate estensioni, disponibili su Internet, contenenti collegamenti ipertestuali a siti web liberamente accessibili al pubblico sui quali sono state messe a disposizione del pubblico opere tutelate dal diritto d'autore senza l'autorizzazione dei titolari di tale diritto». Lo stesso vale probabilmente per l'operatore che, essendo a conoscenza dell'uso illecito dell'applicativo posto in essere dagli utenti, abbia deliberatamente omesso di adottare dei presidi tecnici per evitare la violazione dei diritti. Del resto, l'adozione di presidi del genere non pare ormai tecnicamente impossibile: basti pensare ai sistemi automatici "anti-plagio" in grado di individuare la coincidenza di un certo testo con i testi presenti in un *database* di riferimento. Su questo punto v. la giurisprudenza sulla comunicazione al pubblico delle opere da parte degli intermediari della rete, citata alla nt. 66.

sfruttamento commerciale dell'opera; c) ha però come funzione principale quella di consentire attività di raccolta e di analisi di dati, che cadono fuori dal campo di applicazione del diritto d'autore.

Una prima possibile soluzione del problema è quella di interpretare il diritto di riproduzione nel senso che il rischio, anche remoto, che la creazione di un file possa portare ad una successiva interferenza con i mercati dell'opera basta a concludere che il suddetto file rientra nel campo dell'esclusiva.

Questa interpretazione solleva però qualche perplessità. L'analisi computazionale dell'IA consente di raccogliere e di utilizzare informazioni che non sarebbero verosimilmente accessibili con le normali tecniche di analisi. Anticipare la tutela dei titolari al momento della creazione dei file normalizzati significa rendere impossibile o, comunque, ostacolare enormemente l'utilizzo di queste informazioni. La tesi che include tali file tra le riproduzioni oggetto di esclusiva parte dunque dal presupposto che, in questo contesto, l'interesse di proteggere i titolari dei diritti sia l'unico obiettivo meritevole di tutela o, comunque, che l'interesse dei titolari sia prevalente rispetto all'interesse all'uso delle informazioni.

Si è già detto, però, che il diritto d'autore si fonda sul principio per cui la protezione riguarda solo l'espressione creativa dell'autore e non si estende alle idee e alle informazioni contenute nell'opera. Anche l'interesse alla libera circolazione delle informazioni è, quindi, considerato rilevante dalla disciplina del diritto d'autore ed è individuato come un limite alla protezione degli autori. L'interpretazione delle regole sul diritto d'autore deve dunque tenere conto di entrambi gli interessi e tentare di realizzare un bilanciamento in caso di conflitto. In quest'ottica, la lettura che estende il diritto di riproduzione a qualunque oggetto che possa interferire, anche solo lontanamente, con i mercati dell'opera andrebbe, quanto meno, sottoposta ad un vaglio di "proporzionalità". Si tratta, cioè, di chiedersi se non vi siano delle interpretazioni alternative altrettanto efficaci dal punto di vista dei titolari, ma meno restrittive nei confronti della raccolta e dell'utilizzo delle informazioni.

La lettura qui proposta, secondo cui i file fortemente normalizzati non devono essere inclusi tra le "copie" ai sensi del diritto di riproduzione, è sicuramente meno restrittiva dal punto di vista della circolazione delle informazioni. Resta, però, da chiedersi se essa sia anche in grado di assicurare ai titolari una tutela adeguata nel caso in cui la produzione dei file porti a forme di sfruttamento commerciale della espressione creativa dell'autore. Come accadrebbe, appunto, nel caso in cui un'IA venga programmata per decodificare dataset normalizzati e per trasmetterne i contenuti al pubblico. A questo proposito, si è già visto, però, che la lettura in esame non priva affatto di tutela i titolari contro gli eventuali usi "abusivi" dei file normalizzati. Il titolare resta infatti libero di esercitare i propri diritti di riproduzione e di comunicazione contro i successivi atti di divulgazione del testo. La differenza rispetto alla lettura più restrittiva sta nel fatto che qui il titolare non può agire al momento iniziale della produzione dei file normalizzati. A differenza di

quello che accade in altri campi, però, qui lo spostamento in avanti dell'intervento non pare rendere più difficile l'esercizio dei diritti. Il titolare infatti non è costretto ad agire nei confronti di una massa dispersa di utenti ma può rivolgere le sue pretese anche nei confronti dell'operatore di IA che offra agli utenti l'infrastruttura di decodifica e di trasmissione dei testi. La lettura che sottrae i dataset normalizzati all'esclusiva sembra allora realizzare un migliore bilanciamento degli interessi in gioco ed appare quindi preferibile⁶⁷.

Questa conclusione appare del resto confermata anche se si guarda al problema dal punto di vista degli obiettivi generali della disciplina sul diritto d'autore. La tesi secondo cui anche i file normalizzati sono copie si fonda sull'idea che l'obiettivo del diritto d'autore sia quello di assicurare la massima tutela possibile agli autori. Questa però non è l'unica possibile lettura della disciplina. Secondo un'altra parte della dottrina, il diritto d'autore e la proprietà intellettuale, in generale, sono discipline funzionalizzate al conseguimento di obiettivi di benessere collettivo⁶⁸.

⁶⁷ Va detto che qui si potrebbe anche porre la questione se la trasformazione del testo in un codice di stampo matematico non sia piuttosto una traduzione dell'opera. Attività, questa, che è oggetto di un autonomo diritto esclusivo (art. 18, co. 1, l. aut.). La questione impone di chiedersi quale sia il significato del concetto di "traduzione" in questo campo. Il diritto di tradurre l'opera ha tradizionalmente lo scopo di consentire all'autore di beneficiare dell'espansione territoriale dei mercati dell'opera. La legge sembrerebbe riferirsi, quindi, alla traduzione in una lingua che consenta di raggiungere un nuovo pubblico; cioè, in sostanza, ad una lingua "parlata". Il punto sembra confermato dalla lettera dell'art. 18, primo comma, il quale riserva all'autore la traduzione dell'opera «in altra lingua o dialetto». Il che dovrebbe escludere l'applicazione del diritto alla normalizzazione del testo. D'altra parte, gli argomenti espressi nel testo con riferimento alla portata del diritto di riproduzione sembrano validi anche per il diritto di traduzione.

⁶⁸ M. LIBERTINI, *Tutela e promozione delle creazioni intellettuali e limiti funzionali della proprietà intellettuale*, in *AIDA*, 2014, 299 ss., in cui l'A. critica la tesi secondo cui la proprietà intellettuale trova il suo fondamento nelle stesse ragioni che giustificano la tutela delle altre forme di proprietà privata. Queste tesi sono sostenute da una parte della dottrina, tra l'altro, facendo leva sull'art. 17, par. 2 della Carta dei diritti fondamentali dell'UE, dove si legge che «la proprietà intellettuale è protetta». V. sul punto A. OTTOLIA, *The Public Interest and Intellectual Property Models*, Torino, Giappichelli, 2010; M. BERTANI, (nt. 9), 148, in cui l'A. afferma, comunque, che la proprietà intellettuale può subire compressioni nelle ipotesi in cui l'interesse della collettività alla circolazione dell'opera abbia rango pari o superiore all'interesse del titolare. L'argomento relativo all'art. 17, par. 2 è stato ridimensionato dalla CGUE. Si v. CGUE, 29 luglio 2019, C-476/19, *Pelham*, par. 33: «la Corte ha, in tal senso, già dichiarato che non risulta in alcun modo dall'articolo 17, paragrafo 2, della Carta né dalla giurisprudenza della Corte che il diritto di proprietà intellettuale sancito da tale disposizione sia intangibile e che la sua tutela debba essere garantita in modo assoluto».

Nel senso che il fondamento "costituzionale" del diritto d'autore imponga un bilanciamento tra gli interessi degli autori e l'interesse generale all'innovazione, v., tra gli altri, L. SCHIUMA, (nt. 53); M. RICOLFI, (nt. 57), 461, in cui l'A. afferma che, nella disciplina sul diritto d'autore, l'incentivo proprietario degli autori deve essere coordinato con l'interesse generale alla diffusione dei contenuti protetti. V. anche con specifico riferimento all'IA l'analisi di R.M. HILTY, J. HOFFMANN, S. SCHEUERER, *Intellectual property justifications for artificial intelligence*, in *Artificial Intelligence and Intellectual Property*, edito da Lee, Hilty, Liu, 2021, 50 ss.

Un ampliamento della protezione autoriale deve quindi pur sempre essere sorretto da reali esigenze di sviluppo economico e non deve imporre sacrifici sproporzionati all'efficienza dinamica in altre direzioni. Si è detto che l'estensione dell'esclusiva alle copie "normalizzate" non si giustifica con esigenze di protezione dei mercati dell'opera, visto che l'autore resta comunque legittimato ad esercitare i propri diritti nei confronti della riproduzione "decodificata". Inoltre, questa lettura "espansiva" dell'esclusiva equivarrebbe sostanzialmente a bloccare nel mercato europeo lo sviluppo di una tecnologia altamente innovativa e suscettibile di innumerevoli applicazioni utili, molte delle quali nulla hanno a che vedere con i mercati dei prodotti culturali. Anche da questo punto di vista, la lettura appare dunque sproporzionata e, come tale, incompatibile con gli obiettivi generali del diritto d'autore.

In sintesi, assorbita la funzione della riproduzione normalizzata, tecnica e interna, nel processo volto all'addestramento dell'IA generativa, non pare violare il diritto di riproduzione la mera creazione di copie necessarie al procedimento destinate ad essere conservate in formati e linguaggi utilizzabili esclusivamente per l'analisi degli algoritmi.

Beninteso, questo non significa che il *dataset* sottoposto ad attività di normalizzazione "piena" metta completamente fuori gioco il diritto di riproduzione. Non è detto infatti che la normalizzazione avvenga già in fase di creazione del *dataset*. Può accadere, ad es., che il *dataset* contenga testi identici agli esemplari di base e che il *pre-processing* avvenga nella fase del *training* vero e proprio, cioè dopo la condivisione del *dataset* con i programmatori dell'IA. In questo caso (e in tutti i casi in cui il *dataset* si compone di copie "fruibili" da parte del pubblico) vale quanto detto nei paragrafi precedenti: la creazione del *dataset* comporta creazione di "copie" ed entra quindi in conflitto con il diritto di riproduzione, fatta salva l'eventuale applicazione delle eccezioni in tema di estrazione o di riproduzione temporanea.

Se si adotta la lettura qui proposta, però, la creazione di un *dataset* che sia fin dall'inizio sottoposto a processi di forte normalizzazione non comporta la produzione di "copie". L'operazione cade dunque del tutto al di fuori del diritto di riproduzione ed è da considerare libera. In questo caso, il problema di applicare le eccezioni non si pone. Un *dataset* "normalizzato" potrebbe essere costruito anche andando al di là delle strette maglie dell'art. 5, par. 1: potrebbe, cioè, essere composto da *file* permanenti, suscettibili di trasferimento ad altri programmatori ed utilizzabili in una pluralità di processi di analisi computazionale. In questi casi, dovrebbe poi considerarsi libera anche la successiva circolazione del *dataset*. L'atto di trasferimento non ha infatti ad oggetto "copie" dell'opera e si colloca, pertanto, al di fuori del diritto di distribuzione. D'altra parte, l'operazione non pare neppure rientrare tra le forme di comunicazione al pubblico, visto che, come già detto, i *file* in questione sono incapaci di trasmettere direttamente l'opera agli utenti. Nel caso del *dataset* pienamente normalizzato, l'esclusiva potrebbe venire

in rilievo, tutt'al più, per le copie iniziali create al momento del *download* che consistono in repliche pedissequae dei contenuti raccolti dal sito. Come già visto, però, se sottoposte a meccanismi di cancellazione automatica, queste copie possono rientrare nell'ambito di applicazione dell'art. 5, par. 1.

7. *Intelligenza artificiale e banche dati.*

Come è noto, il diritto europeo prevede una forma di protezione anche per le “banche dati” (direttiva 96/9/CE). In questo campo, per “banca dati” si intende «una raccolta di opere, dati o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili grazie a mezzi elettronici o in altro modo» (art. 1, par. 2, dir. 96/9). Quasi tutti i siti di contenuti, come quelli giornalistici, sembrano poter rientrare in questa definizione: essi infatti raccolgono “opere” (i.e. gli articoli o i testi pubblicati), le dispongono in maniera sistematica e offrono agli utenti meccanismi di accesso individuale⁶⁹.

Si pone quindi, a questo punto, anche la questione se la raccolta di testi dal web per l'addestramento dell'IA sia compatibile con i diritti sulle banche dati.

7.1. *Diritto d'autore e diritto sui generis.*

Per le banche dati esistono due diverse forme di tutela. In primo luogo, il diritto d'autore protegge le banche dati dotate di carattere “creativo”, cioè quelle in cui la selezione o la disposizione dei materiali è il frutto di scelte originali (art. 3, dir. 96/9). La maggior parte delle risorse web adotta criteri di sistemazione standard e sembra dunque incapace di soddisfare questo requisito⁷⁰. Il panorama di

⁶⁹ La questione se un sito internet sia da inquadrare tra le “banche dati” ai sensi della direttiva è discussa in dottrina. Si v. A. MUSSO, (nt. 53), 114; M.S. SPOLIDORO, *Il sito web*, in *AIDA*, 1998, 187; L. MANSANI, *La protezione dei database in Internet*, in *AIDA*, 1996, 151; G. BONELLI, *Il sito web quale opera dell'ingegno*, in *Dir. dell'informazione e dell'informatica*, 2002, I, 199 ss.

⁷⁰ Nella maggior parte dei casi, infatti, i siti internet sono organizzati secondo criteri volti a velocizzare la comprensione degli utenti e la navigazione tra i contenuti. Così, una pagina web tende a disporre i contenuti secondo criteri cronologici, per temi generali (es. cronaca, politica, sport, ecc.) o in base all'indice di gradimento degli utenti (si pensi al criterio delle visualizzazioni nelle piattaforme). In tutti questi casi, la disposizione dei contenuti non pare potersi considerare “originale”. Si v. su questo punto R. MEYS, *Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market: Are European Database Protection Rules Still Threatening the Development of Artificial Intelligence?*, in *GRUR Int.*, 2021, 459; V. FALCE, *L'“insostenibile leggerezza” delle regole sulle banche dati nell'unione dell'innovazione*, in *Riv. dir. ind.*, 2018, 385 ss. Quanto alla scelta dei contenuti, il panorama dei siti internet è quanto mai variegato ed è dunque certamente possibile che ve ne siano alcuni che adottano scelte creative nel definire i propri contenuti. Tuttavia, la creatività che rileva qui non attiene alla scelta del tema di fondo del sito, ma unicamente alla selezione degli specifici materiali che compongono il sito. In altre parole, un sito che sceglie un tema

Internet è comunque estremamente variegato ed è ben possibile che tra i siti adoperati dall'IA ve ne siano anche alcuni "creativi".

Il diritto d'autore qui protegge comunque soltanto la struttura originale della banca dati, non i contenuti della stessa. I contenuti restano in linea di principio liberamente utilizzabili, fatto salvo naturalmente il caso in cui si tratti di opere a loro volta protette dal diritto d'autore (art. 3, par. 2, dir. 96/9)⁷¹. Dal canto loro, i crawler dell'IA in genere non replicano la struttura dei siti Internet. Essi riprendono soltanto i testi presenti sul sito, li traducono in un formato particolare e li riversano in una apposita banca dati ordinata secondo criteri diversi da quelli usati nel sito-fonte. Sotto questo profilo, il procedimento dell'IA non pare quindi entrare in conflitto con il diritto d'autore⁷².

Le banche dati possono però anche essere protette dal diritto c.d. *sui generis* (art. 7 dir. 96/9 e art. 102-bis ss. l. aut.). A differenza del diritto d'autore, questo diritto attribuisce al titolare un controllo anche sul contenuto della banca dati. Il titolare ha infatti il potere di vietare, tra l'altro, le attività di estrazione delle informazioni presenti nella banca dati⁷³. Qui per estrazione si intende «il trasferimento permanente o temporaneo della totalità o di una parte sostanziale del contenuto di una banca di dati su un altro supporto con qualsiasi mezzo o in qualsivoglia forma». La raccolta automatica di testi può indubbiamente rientrare in questa definizione, quanto meno nei casi in cui essa avvenga mediante creazione di copie da trasferire in un apposito *dataset*⁷⁴.

Resta comunque ancora da chiedersi se le fonti adoperate per l'addestramento dell'IA rientrino tra quelle protette dal diritto *sui generis*. Questa tutela si applica soltanto se sono stati effettuati investimenti rilevanti per il conseguimento, la verifica o la presentazione dei contenuti della banca dati⁷⁵.

originale (es. raccogliere le sentenze che riguardano la protezione degli animali) e tende poi a raccogliere in maniera esaustiva tutte le informazioni e i contenuti relativi a quel tema non sarebbe il frutto di una "selezione" creativa ai sensi della disciplina sulle banche dati. V. su questo punto S. VON LEWINSKI, *Database Directive*, in *European copyright law*, edito da Walter, von Lewinski, 2010, Oxford, 705. Per una panoramica sulle decisioni nazionali in questo campo v. E. DERCLAYE, *The legal protection of databases. A comparative analysis*, Cheltenham, Edward Elgar, 2008, 45 ss.

⁷¹ In questa ipotesi resta fermo quanto detto in precedenza sulla possibile violazione del diritto d'autore sui singoli testi da parte dell'IA. Si v. P. AUTERI, *Note introduttive*, in *Attuazione della direttiva 96/9/CE relativa alla tutela giuridica delle banche dati*, in *Nuove leggi civ. comm.*, 2003, 1175 ss.

⁷² A. MUSSO, (nt. 53), 112 ss.; P. AUTERI, *sub art. 2*, in *Attuazione della direttiva 96/9/CE relativa alla tutela giuridica delle banche dati*, in *Nuove leggi civ. comm.*, 2003, 1190 ss.; P. SPADA, *Banche di dati e diritto d'autore (il "genere" del diritto d'autore sulle banche dati)*, in *AIDA*, 1997, 8.

⁷³ P. AUTERI, (nt. 71), 1177.

⁷⁴ In tal senso v. I.B. ØRSTAVIK, *Access to data for training algorithms in machine learning – copyright law and "right-stacking"*, in *Artificial Intelligence and the Media*, edito da Pihlajarinne e Alén-Savikko, Cheltenham, Edward Elgar, 2022, 272 ss.

⁷⁵ Sulla fattispecie costitutiva del diritto *sui generis* v. in generale V. DI CATALDO, *Banche-dati e diritto sui generis: la fattispecie costitutiva*, in *AIDA*, 1997, 24.

Per la creazione dei grandi archivi *web*, come i registri elettronici di dati raccolti da Internet o i *dataset*, è necessario costruire una infrastruttura di memorizzazione, individuare i dati rilevanti, classificarli ed adattarli agli standard di archiviazione, ecc. Tutte attività, queste, che possono richiedere investimenti notevoli. È dunque molto probabile che tali fonti rientrino tra i *database* protetti. Con la conseguenza che la raccolta di materiali in questo caso non è libera, ma richiede l'autorizzazione dei titolari dei diritti *sui generis*⁷⁶.

Più complessa è invece la questione se rientri nell'ambito della disciplina sulle banche dati un normale sito di contenuti, come una pagina *web* giornalistica.

Anche lo sviluppo di un sito del genere può comportare investimenti notevoli. Bisogna però intendersi sugli investimenti che rilevano ai fini della tutela *sui generis*: questa forma di protezione scatta soltanto se vi sono investimenti rilevanti sotto il profilo del “conseguimento”, della “verifica” o della “presentazione” dei dati.

Il significato di “conseguimento” dei dati in questo campo ha dato luogo ad un certo dibattito. Le discussioni hanno riguardato soprattutto l'ipotesi in cui un'impresa svolga un'attività economica diversa dalla raccolta e dalla elaborazione di informazioni. È possibile infatti che, in questi casi, l'attività principale dell'impresa produca comunque dei dati e che tali dati siano poi riuniti in un *database* funzionale allo svolgimento dell'attività principale⁷⁷. Ci si è dunque chiesti se, in questi casi, gli investimenti sostenuti dall'impresa per realizzare l'attività principale rientrino tra gli investimenti volti al “conseguimento” di dati da prendere in considerazione nella valutazione sull'applicabilità della tutela *sui generis* al *database*⁷⁸.

Secondo la Corte di Giustizia, l'obiettivo della tutela *sui generis* è quello di incentivare gli investimenti nei sistemi di archiviazione e di gestione dei dati. Investimenti, questi, che non verrebbero sostenuti senza apposite forme di protezione giuridica⁷⁹. La creazione di dati nuovi è invece un effetto delle attività economi-

⁷⁶ Restano, comunque, applicabili le eccezioni previste per il diritto *sui generis*, tra cui l'eccezione di *text and data mining*, di cui si è già detto, laddove ne ricorrano i presupposti. Si v. sopra par. 3.

⁷⁷ È, ad es., il caso della lega sportiva. La lega investe risorse significative per il coordinamento tra le squadre, la pianificazione delle partite, l'individuazione degli stadi e dei partecipanti, ecc. Tutto ciò è indispensabile per svolgere l'attività principale di organizzazione del campionato, ma comporta anche inevitabilmente la costruzione di una banca dati, cioè il calendario degli incontri. V. CGUE, 9 novembre 2004, C-444/02, *Fixtures marketing*.

⁷⁸ Secondo una parte della dottrina, qualsiasi *database* connesso ad un'attività principale dovrebbe cadere fuori dalla tutela *sui generis* (c.d. teoria del *database spin-off*). Si v. B. HUGENHOLTZ, *Abuse of database right. Sole source information banks under the EU database directive*, in *ssrn.com*, 2004, 6.

⁷⁹ CGUE, 9 novembre 2004, C-444/02, *Fixtures marketing*, par. 40; CGUE, 9 novembre 2004, C-46/02, *Fixtures marketing*, par. 33; CGUE, 9 novembre 2004, C-338/02, *Fixtures Marketing*, par.

che svolte dalle imprese sul mercato e si verifica a prescindere dalla previsione di apposite forme di tutela⁸⁰. Pertanto, secondo la Corte, l'investimento rivolto al "conseguimento" dei dati comprende soltanto i mezzi destinati alla ricerca di dati pre-esistenti e alla loro riunione in una banca dati; non i mezzi impiegati, direttamente o indirettamente, per la creazione dei dati⁸¹. Da ciò deriva anche che gli investimenti sostenuti per un'attività economica produttiva di dati non rientrano tra quelli rilevanti ai fini della tutela *sui generis*. La banca dati che sia il mero "sotto-prodotto" di un'attività principale è tutelata solo se ha comportato degli in-

23; CGUE, 9 novembre 2004, C-203/02, *The British Horseracing Board*, par. 30; CGUE, 9 ottobre 2010, C-304/07, *Directmedia Publ.*, par. 33. La Corte richiama a tal proposito il considerando 12 dir. 96/9, in cui si legge che «tale investimento nei moderni sistemi di memorizzazione e gestione delle informazioni non sarà effettuato all'interno della Comunità a meno che non venga introdotta una tutela giuridica stabile ed uniforme per tutelare i costitutori di banche di dati».

⁸⁰ Questo ragionamento emerge in CGUE, 9 novembre 2004, C-444/02, *Fixtures marketing*, par. 39, in cui si legge che la finalità della direttiva è quella di «incentivare e tutelare gli investimenti nei sistemi di "memorizzazione" e "gestione" dei dati che contribuiscono allo sviluppo del mercato delle informazioni in un contesto caratterizzato da una crescita esponenziale della massa di informazioni prodotte e elaborate annualmente in tutti i settori di attività». V. nello stesso senso la nota alla sentenza di A. COGO, in *AIDA*, 2005, 416; A. OTTOLIA, (nt. 13), 79-80 e M. HUSOVEC, *The end of (meta)search engines in Europe?*, in *Chicago-Kent J. of Int. Prop.*, 2015, 153. Secondo alcuni, le conclusioni della Corte sarebbero anche implicitamente motivate dall'esigenza di evitare di attribuire ad un soggetto il monopolio sui dati da esso generati. Si v. in particolare il commento di M. LEISTNER alla decisione C-203/02, in *IIC*, 2005, 581 ss. e A. KUR, R. M. HILTY, C. GEIGER, M. LEISTNER, *First evaluation of directive 96/9/EC on the legal protection of databases*, in *IIC*, 2006, 552 ss.

⁸¹ La Corte non pare quindi aderire del tutto alla teoria dei *database spin-off*. Si v. sul punto M. BERTANI, *Banche dati ed appropriazione delle informazioni*, in *Eur. Dir. priv.*, 2006, 323. Secondo diversi autori, sarebbe comunque molto difficile provare l'esistenza di investimenti autonomi rispetto all'attività principale. Di fatto, quindi, i *database spin-off* finirebbero per cadere fuori dall'ambito della tutela *sui generis*. Si v. E. DERCLAYE, *The Database Directive*, in *EU Copyright Law*, edito da Stamatoudi, Torremans, 2021, 233.

Il problema dei *database* connessi ad un'attività economica principale è stato poi più volte affrontato dalla giurisprudenza degli Stati membri, talora proprio con riguardo al *web scraping* di siti internet. V. i casi olandesi Rb. Arnhem, 16 marzo 2006, in *AMI*, 2006, 93; Hof Arnhem, 4 luglio 2006, in *Computerrecht*, 2007/5; Hof Amsterdam, 13 marzo 2012, in *Rechtspraak.nl*, in cui si affronta la questione se il sito di vendita di biglietti di una compagnia aerea comporti investimenti tali da giustificare l'applicazione della tutela *sui generis*. La corte nella decisione afferma che il sito è il principale canale di vendita della compagnia. Gli investimenti nel sito sono sostenuti per la propria attività principale di vendita. Tali investimenti non possono quindi essere integralmente presi in considerazione per la tutela *sui generis*. In senso simile si v. la decisione spagnola Trib. Supremo, 9 ottobre 2012, *Ryanair c. Atrapalo*, in *vlex.es*. V. anche la decisione francese Cour de Cassation, 5 marzo 2009, in *legifrance.gouv.fr*. V. per una decisione, in parte, diversa BGH, 25 marzo 2010, I ZR 47/08, su una banca dati avente ad oggetto i dati sul traffico delle autovetture in autostrada. Per una panoramica comparatistica sul trattamento delle banche dati *spin-off* si v. A.C. BEUNEN, *Protection for databases: the European database directive and its effects in the Netherlands, France and the United Kingdom*, in *Wolf Legal Publ.*, 2007, 108 ss.

vestimenti autonomi aggiuntivi rispetto a quelli che l'impresa sostiene in funzione della propria attività principale⁸².

Si è detto che, nel caso dei siti giornalistici, i “dati” sono i testi, gli articoli e i contenuti del sito stesso. L'impresa che gestisce un sito internet investe risorse spesso notevoli nella produzione di questi contenuti. Questi sono però, per l'appunto, investimenti volti alla “creazione” di dati, non alla raccolta e alla riunione di contenuti pre-esistenti. D'altra parte, l'attività principale delle imprese che gestiscono questi siti, specialmente in campo giornalistico, è quella di diffondere articoli, testi o altro tipo di prodotti. L'investimento nella produzione di contenuti in grado di attirare l'attenzione degli utenti viene presumibilmente sostenuto a prescindere dalla protezione del sito come banca dati. In base all'impostazione seguita dalla Corte, questi investimenti sono pertanto irrilevanti ai fini della tutela *sui generis*⁸³.

Qui un investimento rilevante per il “conseguimento” di dati si potrebbe avere, tutt'al più, nell'ipotesi in cui il gestore del sito investa risorse nella raccolta di contenuti ulteriori, prodotti da altri operatori. Non sembra il caso però della maggior parte dei siti giornalistici, che tendono a diffondere principalmente i propri prodotti editoriali. Lo stesso pare potersi dire per i *blog*, le enciclopedie digitali, i siti di discussione o di recensione, in cui in genere il gestore del sito pubblica contenuti prodotti appositamente per il sito stesso. Ancora, nel caso dei *forum* e dei *social network*, i contenuti sono direttamente creati dagli utenti e non c'è quindi nessun investimento da parte del gestore nella raccolta di materiali.

Resta, comunque, ancora da verificare se il sito richieda investimenti rilevanti sotto gli altri due profili richiamati dalla disciplina, cioè la “verifica” o la “presenziazione” dei dati.

Sotto il primo profilo, la Corte afferma che l'investimento nella “verifica” dei dati è irrilevante per la tutela *sui generis* se avviene contestualmente alla creazione dei dati stessi⁸⁴. Un sito giornalistico sottopone gli articoli a forme di control-

⁸² Il punto è confermato dalla stessa direttiva 96/9 al considerando 19, in cui si legge che «di norma, la compilazione di varie registrazioni di esecuzioni musicali su CD non rientra nel campo d'applicazione della presente direttiva sia perché, in quanto compilazione, non soddisfa le condizioni per essere tutelata dal diritto d'autore, sia perché non rappresenta un investimento sufficientemente rilevante per beneficiare del diritto sui generis». In realtà, l'attività di produzione fonografica richiede investimenti molto elevati. Il CD in quanto banca dati non è però tutelato proprio perché si tratta in un mero “sottoprodotto” dell'attività principale, cioè quella di produzione e di distribuzione delle copie fonografiche.

⁸³ In questo senso, con riferimento ai siti *web* giornalistici, si v. Rb. Rotterdam, 22 agosto 2000, in *Rechtspraak.nl*.

⁸⁴ CGUE, 9 novembre 2004, C-203/02, *The British Horseracing Board*, parr. 39-40: «Certo, il processo di iscrizione di un cavallo su un elenco di corsa richiede un certo numero di verifiche preliminari, relative all'identità di colui che effettua l'iscrizione, alle caratteristiche del cavallo, nonché alle qualificazioni del cavallo, del suo proprietario e del fantino. Tuttavia, questo lavoro di verifica

lo, ad es., sotto il profilo dell'attendibilità delle notizie o della presenza di errori formali. Questa verifica si colloca, però, in genere proprio nella fase di lavorazione finale dei contenuti prima della pubblicazione e pare dunque doversi considerare irrilevante ai fini della protezione del sito come *database*.

Più complesso è invece il tema dell'investimento nella "presentazione" dei dati. Secondo la Corte, con questa espressione la direttiva intende fare riferimento ai «mezzi intesi a conferire alla detta banca di dati la sua funzione di gestione dell'informazione, ossia quelli destinati alla disposizione sistematica o metodica degli elementi contenuti in questa banca di dati nonché all'organizzazione della loro accessibilità individuale».

I siti giornalistici investono risorse significative nello sviluppo dell'interfaccia grafica, nei sistemi di navigazione interna, nella eventuale creazione di metadati, nella catalogazione dei contenuti, ecc. Questi investimenti sono, però, a loro volta, collegati all'attività principale del sito, cioè la comunicazione al pubblico di articoli e notizie. La Corte non ha mai risolto la questione se un investimento nella "presentazione" di dati che sia collegato ad un'attività economica principale diversa dalla mera gestione di dati sia rilevante ai fini della tutela *sui generis*. Come già detto, però, l'obiettivo della direttiva è quello di incentivare gli investimenti che potrebbero non essere sostenuti senza apposite forme di protezione.

L'organizzazione di un sito giornalistico è indispensabile per consentire agli utenti di fruire degli articoli pubblicati. D'altra parte, la sistemazione dei contenuti serve anche a catturare l'attenzione del pubblico e a massimizzare il traffico. Lo dimostra il fatto che spesso i siti ordinano i contenuti in base alle preferenze degli utenti o in base al grado di "tendenza" di certi temi. Criteri, questi, che rispondono più all'esigenza di stimolare la curiosità degli utenti che a quella di consentire una rapida ed efficace ricerca dei dati all'interno del sito.

In sostanza, per svolgere la propria attività efficacemente, il gestore di un sito giornalistico non può fare a meno di presentare i contenuti prodotti in una maniera ordinata, efficace e comprensibile. Gli investimenti nella "presentazione" dei contenuti sono quindi verosimilmente effettuati dagli editori a prescindere dalla previsione di forme di protezione giuridica per il sito internet. Anch'essi sembrano quindi doversi considerare irrilevanti per l'applicazione della tutela *sui generis*⁸⁵.

preliminare interviene nella fase di creazione dell'elenco relativo alla corsa di cui trattasi. Esso costituisce quindi un investimento collegato alla creazione di dati, e non alla verifica del contenuto della banca di dati».

⁸⁵ In questo senso, sugli investimenti effettuati per la presentazione di un sito *web*, v. Cour d'Appel de Lyon, 27 aprile 2010, in *legifrance.gouv.fr.*, in cui si legge che «les factures [...] concernent une charte graphique dont l'objet est d'homogénéiser l'identité visuelle des boutiques en ligne de la marque shop avenue [...]. Elles ne peuvent être retenues comme investissement en vue de procéder à la recherche, la vérification et au rassemblement des données alors qu'elles ne concernent que la ligne visuelle des sites. Les factures [...] comportent entre autres objets la création de bannières publicitaires, la création de formulaires newsletters, la gestion du projet, de sorte qu'elles

L'impostazione qui seguita sembra, del resto, in linea con le posizioni recentemente assunte sul tema da parte della Commissione⁸⁶. In occasione dell'ultima valutazione sull'impatto della direttiva 96/9, effettuata nel 2018, la Commissione ha affermato che «it needs to be pointed out that due to the 2004 CJEU decisions, which clarified the scope of the sui generis right, it is assumed that the sui generis right does not apply to databases that are the by-products of the main activity of an organisation. This means that the sui generis right does not apply broadly to the data economy (machine-generated data, IoT devices, big data, AI, etc.)»⁸⁷. La Commissione sembra qui dare per scontato che i *database* funzionali ad un'attività principale siano esclusi dalla tutela *sui generis*. E ciò anche se, come in

ne peuvent être considérées comme participant dans leur entier à la création de la base de données». Si v. anche Trib. Milano, 4 giugno 2013, in *dejure.it*, in cui il Tribunale raggiunge conclusioni simili con riguardo al sito di una compagnia aerea, adottando però un'impostazione diversa. Il Tribunale afferma che l'investimento nella presentazione di un sito di vendita di biglietti è un investimento rilevante ai fini della tutela *sui generis*. Poi però afferma anche che l'estrazione dei dati dalla banca dati non pregiudica l'investimento nella presentazione dei dati, ma danneggia soltanto l'attività principale del gestore del *database*. Sulla base di questo ragionamento la sentenza esclude che vi sia una violazione del diritto *sui generis*: «la convenuta mette a disposizione delle agenzie, anche online, una licenza [...] che consente al licenziatario l'accesso e l'estrazione di tutti i dati, che possono essere visualizzati sul suo sito, al prezzo simbolico di euro 100 all'anno (pure devolute in beneficenza), con l'impegno a non intermediare nell'acquisto dei voli (che deve essere fatto sul sito Ryanair dall'utente finale, eventualmente connesso tramite un link sulla pagina web dell'OTA). Da tale circostanza risulta che l'odierna convenuta non considera che l'estrazione e reimpiego dei dati sui voli da parte della OTA siano tali di per sé da pregiudicare gli sforzi di investimento ed innovazione posti in essere per la costituzione della banca dati, ma intenda escludere l'utilizzazione degli stessi per l'attività di intermediazione nell'acquisto dei voli Ryanair, a garanzia delle politiche commerciali di esclusiva vendita diretta. Alla luce di siffatte elementi (comprensivi del comportamento concreto del costituente), la condotta di estrazione/reimpiego attuata da Viaggiare sulla banca dati della convenuta – anche, se attivata attraverso un espediente informatico e scontando tutte le parzialità del servizio così fornito ai suoi utenti (con conseguenti responsabilità nei loro confronti) – non appare di per sé inammissibilmente invasiva e come tale illecita, considerato che lo ius excludendi sembra essere stato attivato da Ryanair al fine preminente di tutelare la propria scelta di politica commerciale di vendita esclusiva attraverso i propri canali».

⁸⁶ Si v. sul punto COMMISSIONE, *Una strategia europea per i dati*, 19 febbraio 2020, COM (2020) 66 final. Per una descrizione dei conflitti tra la direttiva 96/9/CE e i nuovi obiettivi europei in tema di dati si v. C. SGANGA, *Ventisei anni di direttiva database alla prova della nuova strategia europea dei dati: evoluzioni giurisprudenziali e percorsi di riforma*, in *Dir. Inf.*, 2022, 651 ss.

⁸⁷ V. COMMISSIONE, *Executive Summary of the Evaluation of Directive 96/9/EC on the legal protection of databases*, SWD (2018) 146 final, 2 e COMMISSIONE, *Evaluation of Directive 96/9/EC on the legal protection of databases*, 25 aprile 2018, SWD (2018) 147 final, 24: «it is generally understood that, in the case of spin-off databases, companies would have produced these databases anyway without further incentives. It is evident that football leagues keep creating fixture lists as much as horse racing companies create racing data, despite not having a sui generis right». V. anche 36: «this would also apply to webpages created by most companies which, while falling within the definition of database contained in the Database Directive, would be made of 'created' data rather than 'obtained' data, and thus would not be protected».

genere accade, vengono effettuati investimenti nell'organizzazione sistematica dei dati all'interno del *database*⁸⁸.

A queste conclusioni si potrebbe forse obiettare che le fonti *web* sono molto spesso collegate ad attività principali diverse dalla mera gestione dei dati. Se si afferma che gli investimenti necessari per la "presentazione" di questi *database* sono irrilevanti, si finisce per ridurre notevolmente l'ambito di applicazione della tutela *sui generis* in campo digitale. In realtà, secondo la lettura qui proposta, cadono fuori dalla tutela *sui generis* soltanto quegli investimenti che risultano funzionali alla realizzazione di una certa attività principale e che quindi sono di norma effettuati a prescindere dalla esistenza di forme di protezione del *database*. È però anche possibile che un'impresa decida di effettuare degli investimenti aggiuntivi non strettamente collegati alla propria attività principale. Basti pensare all'editore giornalistico che costruisce un archivio digitale delle pubblicazioni per consentire ricerche tra gli articoli più risalenti. Ancora, si pensi al caso in cui l'editore raccolga in un *database* i dati sulle preferenze degli utenti per condividerli con i terzi⁸⁹.

In sintesi, per le ragioni fin qui esposte gli investimenti di norma sostenuti da un'impresa giornalistica per il proprio sito non sembrano rientrare tra quelli rilevanti ai sensi della direttiva sulle banche dati. Il sito giornalistico in quanto tale pare cadere fuori dal campo del diritto *sui generis*. L'estrazione dei testi per l'IA effettuata su siti del genere dovrebbe quindi essere libera, fatti salvi naturalmente i diritti d'autore sugli articoli, per i quali vale quanto detto nei paragrafi precedenti⁹⁰.

⁸⁸ Questo approccio pare poi seguito anche dalle disposizioni del *Data Act* (Reg. 2023/2854/UE). L'art. 43 afferma infatti che «il diritto "sui generis" di cui all'articolo 7 della direttiva 96/9/CE non si applica quando i dati sono ottenuti o generati da un prodotto connesso o un servizio correlato che rientra nell'ambito di applicazione del presente regolamento». Pertanto, la tutela *sui generis* non si applica ai dati generati da prodotti e, in generale, a tutti quelli *machine-generated* (considerando 112). E, anche qui, la disposizione non fa salva l'ipotesi in cui sia stato effettuato un investimento nella presentazione dei dati.

⁸⁹ In questo senso, pare di poter leggere A. OTTOLIA, (nt. 13), 80 ss. e M. HUSOVEC, (nt. 80), 151. Sul tema degli investimenti aggiuntivi rispetto a quelli sostenuti per la creazione dei dati si v. BGH, 30 aprile 2009, I ZR 191/05, in *GRUR*, 2009, 852.

⁹⁰ La conclusione espressa nel testo sembra, del resto, in linea con la sentenza della CGUE, 3 giugno 2021, C-762/19, *CV-Online Latvia*, par. 41, in cui, nel chiedersi se l'utilizzo di una banca dati da parte di un aggregatore di contenuti costituisca una violazione del diritto *sui generis*, la Corte afferma che «occorre stabilire un giusto equilibrio tra, da un lato, il legittimo interesse dei costitutori di banche di dati di essere in grado di ammortizzare il loro investimento rilevante e, dall'altro, quello degli utenti e dei concorrenti di tali costitutori di avere accesso alle informazioni contenute in dette banche di dati nonché la possibilità di creare prodotti innovativi basati su tali informazioni». A tal fine, secondo la Corte, occorre tenere in considerazione, tra l'altro, il fatto che un aggregatore di contenuti contribuisce «alla creazione e alla distribuzione di prodotti e servizi aventi un valore aggiunto nel settore dell'informazione. Offrendo ai propri utenti un'interfaccia unificata che consente di effettuare ricerche in diverse banche di dati secondo criteri pertinenti dal punto di vista del loro contenuto, essi concorrono ad una migliore strutturazione dell'informazione e facilitano la ricerca in Internet». In sostanza, nel ricostruire la portata della protezione occorre tenere conto del fatto che il diritto *sui ge-*

7.2. La decisione *Ryanair*.

Un ostacolo alla raccolta dei testi per l'IA dai siti internet può comunque ancora derivare dalla nota decisione *Ryanair* della Corte di Giustizia⁹¹.

In questo caso si affronta la questione se, ai sensi della dir. 96/9, il costituente di una banca dati (e, più precisamente, di un sito internet) non tutelata né dal diritto d'autore né dal diritto *sui generis* possa imporre delle limitazioni contrattuali all'utilizzo dei contenuti da parte degli utenti.

Secondo la Corte, le banche dati in questione cadono completamente fuori dal perimetro della disciplina europea. Pertanto, la direttiva non osta a che il creatore di una banca dati del genere regoli l'accesso alla stessa tramite clausole contrattuali.

In sostanza, in base a questa decisione, un sito può vietare l'estrazione dei testi attraverso apposite clausole inserite nelle condizioni del servizio, a prescindere dall'applicabilità del diritto *sui generis*.

La decisione è stata fortemente criticata in dottrina. Secondo molti autori, essa porta a conseguenze paradossali: una banca dati priva dei requisiti della direttiva sarebbe potenzialmente protetta nello stesso modo, se non addirittura in maniera più forte, rispetto alle banche dati dotate dei requisiti⁹². Il che finirebbe per svuotare la direttiva di ogni utilità⁹³. Questa critica pare, peraltro, condivisa anche dalla Commissione europea⁹⁴.

neris non deve tradursi in una restrizione sproporzionata per forme di utilizzo innovativo dei dati in grado di produrre benefici di efficienza. Il ragionamento pare potersi applicare anche all'IA.

⁹¹ CGUE, 15 gennaio 2015, C-30/14, *Ryanair*.

⁹² Le banche dati protette dalla direttiva sono infatti pur sempre soggette alle eccezioni imperative ivi previste. V. specialmente art. 6, art. 8 e art. 9 dir. 96/9/CE. L'art. 15 della stessa direttiva afferma che «qualsiasi disposizione contrattuale in contrasto con l'articolo 6, paragrafo 1 e con l'articolo 8 è nulla e priva di effetti».

⁹³ V. sul punto T. SYNODINOU, *Databases and screen scraping: lawful user's rights and contractual restrictions do not fly together*, in *EIPR*, 2016, 312 ss.; LEISTNER, *Big data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform*, in *ssrn.com*, 2018, 9; M. BORGHI, S. KARAPAPA, *Contractual restrictions on lawful use of information: sole-source databases protected by the back door?*, in *EIPR*, 2015, 505; F. GIORDANELLI, *IPRs and big data: a proposal for a fair balance between businesses' legitimate interests and data sharing in the light of the EU Data Act*, in *ADI*, 2022, 124. V. anche lo studio realizzato per la Commissione *Study in support of the evaluation of directive 96/9/EC on the legal protection of databases*, SMART 2017/0084, 2018, 22.

In realtà, le restrizioni contrattuali hanno comunque, a loro volta, dei limiti rispetto alla tutela *sui generis*. Il diritto derivante dalla clausola contrattuale ha efficacia soltanto *inter partes* e non consente di agire nei confronti dei terzi che facciano uso del contenuto informativo del database. Il diritto *sui generis* protegge invece il titolare anche nei confronti delle utilizzazioni indirette del database: CGUE, 9 novembre 2004, C-203/02, (nt. 79), par. 43 ss. Si v. poi M. W. MONTEROSSO, *Estrazione e (ri)utilizzo di informazioni digitali all'interno della rete internet. Il fenomeno del c.d. web scraping*, in *Dir. Inf.*, 2020, 345 ss., anche per un'analisi della giurisprudenza nazionale sulla questione se i termini di utilizzo del sito web siano effettivamente idonei a produrre un vincolo contrattuale in base al diritto degli Stati membri.

⁹⁴ COMMISSIONE, *Evaluation of Directive 96/9/EC on the legal protection of databases*, 25 aprile 2018, SWD (2018) 147 final, 32.

C'è da dire poi che negli ultimi anni la politica europea in tema di dati è cambiata. A partire dal 2018, la Commissione ha lanciato una nuova strategia fondata sull'idea di promuovere la circolazione e l'accesso ai dati nel mercato europeo digitale⁹⁵. In più di un'occasione, la Commissione ha individuato la protezione dei *database* come uno dei potenziali ostacoli a questa strategia ed ha quindi espresso l'intenzione di ridurre la portata del diritto *sui generis*⁹⁶. La sentenza *Ryanair*, che consente sostanzialmente ai gestori dei *database* telematici di rafforzare la protezione dei dati con un semplice avviso sul sito, non pare in linea con questi obiettivi.

D'altra parte, la Commissione ha anche espresso l'esigenza di costruire un vero e proprio mercato unico europeo dei dati, eliminando gli ostacoli derivanti dall'esistenza di regole diverse negli ordinamenti nazionali⁹⁷. In base alla sentenza *Ryanair*, l'uso dei *database* può essere oggetto di restrizioni contrattuali che sono disciplinate dal diritto nazionale. Questa impostazione lascia allora aperta la possibilità che l'uso di una banca dati offerta su Internet sia disciplinato in maniera diversa nei vari Stati membri. Anche sotto questo profilo, la sentenza pare quindi in conflitto con le recenti posizioni della Commissione⁹⁸.

⁹⁵ COMMISSIONE, *Una strategia europea per i dati*, 19 febbraio 2020, COM (2020) final, 1: «i cittadini dovrebbero disporre dei mezzi per prendere decisioni migliori sulla base delle informazioni ottenute dai dati non personali, e tali dati dovrebbero essere disponibili a tutti, siano essi soggetti pubblici o privati, piccoli o grandi, start-up o colossi. In questo modo la società trarrà il massimo vantaggio dall'innovazione e dalla concorrenza e tutti beneficeranno di un dividendo digitale».

⁹⁶ COMMISSIONE, *Una strategia europea per i dati*, 19 febbraio 2020, COM (2020) final, 16, in cui tra gli obiettivi della politica in tema di dati si menziona la realizzazione di una «valutazione del quadro in materia di protezione dei diritti di proprietà intellettuale al fine di potenziare ulteriormente l'accesso ai dati e il loro utilizzo (compresa una possibile revisione della direttiva sulle banche dati e un eventuale chiarimento dell'applicazione della direttiva in materia di segreti commerciali quale quadro abilitante)». V. anche COMMISSIONE, *Piano d'azione sulla proprietà intellettuale per sostenere la ripresa e la resilienza dell'UE*, 25 novembre 2020, COM (2020) 760 final, 16: «la direttiva sulle banche dati prevede una tutela *sui generis* per le banche dati, che sono il risultato di ingenti investimenti. Da una recente valutazione è emerso che, sebbene apportino un valore aggiunto, la direttiva sulle banche dati potrebbe essere rivista per facilitare l'accesso ai dati e il loro utilizzo. Facendo seguito alla strategia europea per i dati, la Commissione intende pertanto riesaminare la direttiva, in particolare per agevolare la condivisione e lo scambio di dati generati da macchine e dati generati nel contesto dell'introduzione dell'IoT». V. poi l'art. 43 del *Data Act* (Reg. 2023/2854/UE), in cui si prevede che sono sottratti alla tutela *sui generis* i *database* che contengono dati generati da «prodotti connessi» e «servizi correlati», cioè sostanzialmente dati derivanti dalle tecnologie «*Internet of things*». Sul rapporto tra le politiche europee in tema di dati e il diritto d'autore v. tra gli altri V. MOSCON, (nt. 38).

⁹⁷ COMMISSIONE, *Una strategia europea per i dati*, 19 febbraio 2020, COM (2020) final, 13: «la frammentazione tra gli Stati membri rappresenta un grave rischio per la visione di uno spazio comune europeo di dati e per l'ulteriore sviluppo di un autentico mercato unico per i dati». Di qui l'idea di adottare misure orizzontali volte a «creare il quadro globale necessario per l'economia agile basata sui dati, evitando così una dannosa frammentazione del mercato interno causata da azioni incoerenti tra i settori e gli Stati membri».

⁹⁸ Nello stesso senso si v. R. MEYS, (nt. 70), 464. Per la verità questa critica alla sentenza *Ryanair* era valida anche prima del lancio delle nuove politiche europee. La direttiva 96/9 aveva

Infine, la Corte è recentemente tornata a pronunciarsi sugli obiettivi della direttiva sulle banche dati nel caso *CV-Online Latvia*⁹⁹. Questa decisione riguarda una questione diversa da quella affrontata in *Ryanair*, cioè se determinati atti di sfruttamento di una banca dati rientrino tra le attività riservate al titolare del diritto *sui generis*. A tal proposito, la Corte afferma che la direttiva coinvolge due interessi: «da un lato, il legittimo interesse dei costitutori di banche di dati di essere in grado di ammortizzare il loro investimento rilevante e, dall'altro, quello degli utenti e dei concorrenti di tali costitutori di avere accesso alle informazioni contenute in dette banche di dati nonché la possibilità di creare prodotti innovativi basati su tali informazioni». Secondo la Corte, nell'interpretare la direttiva occorre realizzare un bilanciamento tra questi due interessi e non si può quindi attribuire al diritto *sui generis* una portata eccessivamente ampia¹⁰⁰.

La sentenza *Ryanair* ha sostanzialmente l'effetto di sacrificare l'interesse del pubblico ad acquisire e ad utilizzare le informazioni, anche quando non vi sia alcun investimento rilevante da proteggere. Essa si fonda dunque sull'implicito presupposto che l'unico interesse rilevante per la direttiva 96/9 sia quello di proteggere le banche dati. Il punto è, però, smentito dalla successiva sentenza *CV-Online Latvia*. Con quest'ultima decisione, la Corte sembra allora, almeno in certa misura, aver cambiato posizione rispetto al passato. E ciò si deve forse anche alle nuove politiche europee in tema di dati¹⁰¹.

l'espresso obiettivo di eliminare tutti gli ostacoli al funzionamento del mercato interno derivanti dall'esistenza di discipline nazionali differenti in tema di banche dati. Il punto emerge chiaramente al considerando 3, in cui si legge che «è opportuno eliminare le differenze esistenti che producono distorsioni al funzionamento del mercato interno ed impedire che ne sorgano di nuove, mentre non occorre eliminare o impedire che sorgano quelle differenze che non pregiudicheranno il funzionamento del mercato interno oppure lo sviluppo di un mercato dell'informazione all'interno della Comunità». V. anche considerando 4: «la tutela delle banche di dati sulla base del diritto d'autore esiste in forme diverse negli Stati membri, in base alla legislazione o alla giurisprudenza, e che la mancata armonizzazione dei diritti di proprietà intellettuale può avere per effetto di ostacolare la libera circolazione di beni o servizi all'interno della Comunità fintantoché esistano differenze tra le varie legislazioni degli Stati membri per quanto riguarda la portata e le condizioni della tutela dei diritti» e considerando 48: «l'obiettivo della presente direttiva [...] consiste nell'assicurare un livello adeguato e uniforme di tutela alle banche di dati». Queste osservazioni sono confermate dalla stessa CGUE: v. CGUE, 18 ottobre 2012, C-173/11, *Football Dataco*, par. 25. V. sul tema E. DERCLAYE, M. HUSOVEC, *Sui generis database protection 2.0: judicial and legislative reforms*, in *EIPR*, 2022, 323; C. SGANGA, (nt. 86), 671-672.

⁹⁹ CGUE, 3 giugno 2021, C-762/19, *CV-Online Latvia*.

¹⁰⁰ Pertanto, la Corte conclude nel senso che gli atti di utilizzo di una banca dati rientrano tra le attività riservate al titolare del diritto *sui generis* soltanto se arrecano un pregiudizio all'investimento da questi sostenuto. V. anche l'opinione dell'Avvocato Generale Szpunar del 14 gennaio 2021, par. 40 ss., cui la Corte fa rinvio nella sentenza.

¹⁰¹ Sulla portata innovativa della sentenza *CV-Online Latvia* v. L. ZOBOLI, *Il diritto sui generis sulle banche dati e l'economia digitale: perché il Data Act non basta*, in *Riv. dir. ind.*, 2021, 332-333; E. DERCLAYE, M. HUSOVEC, (nt. 98), 4 ss.; C. SGANGA, (nt. 86), 680 ss.

Per tutte queste ragioni, la validità del precedente *Ryanair* pare attualmente piuttosto discutibile. Sembra invece più convincente ritenere che le clausole contrattuali con cui un sito non protetto impedisce qualsiasi forma di estrazione agli utenti siano incompatibili con la disciplina europea in tema di banche dati.

8. Conclusioni.

Bisogna riconoscere che alcune delle proposte interpretative qui formulate si pongono, in parte, in controtendenza rispetto agli orientamenti prevalenti in Europa e, soprattutto, rispetto alla tendenza “espansiva” generalmente adottata dalla Corte di Giustizia nell’interpretare la portata dell’esclusiva autoriale. Del resto, lo stesso *AI Act*, pur non apportando modifiche alle direttive sul diritto d’autore, sembra dare per scontato che, in linea di principio, l’uso dei contenuti protetti da parte dell’IA sia vietato. Su un piano realistico, non si può dunque prescindere dall’affrontare il problema del conflitto tra produzione culturale e IA anche in una prospettiva *de iure condendo*.

D’altra parte, quest’analisi si rende necessaria anche per un’altra ragione. Analizzando il problema *de iure condito*, si è detto che ci sono casi in cui l’IA sfrutta le opere soltanto come fonte di informazioni e che queste situazioni dovrebbero collocarsi fuori dall’ambito del diritto d’autore. Questo non significa però che lasciare del tutto libera l’IA di servirsi delle informazioni presenti sul *web* sia, in assoluto, la maniera più equilibrata di regolare il fenomeno.

L’IA generativa può trovare utile applicazione in numerosi settori. Può essere adoperata, ad es., per scopi di carattere scientifico, per la soluzione di problemi tecnici o per automatizzare certi aspetti di un processo produttivo. Ancora, può svolgere compiti di ricerca statistica o assistere l’utente nella comprensione, nell’analisi o nella stesura di documenti. In tutti questi casi, il servizio non si pone in diretta competizione con le attività delle imprese editoriali e la soluzione migliore resta probabilmente quella di consentire l’utilizzo libero di articoli e contenuti per l’addestramento dell’IA.

Come già visto, però, ci sono anche sistemi di IA offerti al grande pubblico per rispondere alle domande più varie. In questo caso, l’utente potrebbe rivolgersi al *chatbot* per avere informazioni sulle notizie del giorno, su eventi di attualità, su temi storici, ecc. Allo stato questa funzionalità dell’IA non sembra essere arrivata ancora al punto da produrre un concreto impatto negativo sull’industria editoriale. Con l’affinamento dei servizi, però, l’IA potrebbe effettivamente cominciare a sviluppare contenuti del tutto sostituibili ai prodotti editoriali. Il rischio si pone soprattutto per le forme più recenti di IA, nelle quali i meccanismi generativi sono combinati con la possibilità di ricercare su Internet in tempo reale le informazioni richieste dagli utenti. In queste circostanze, l’IA sostanzialmente compete sul mercato dei contenuti editoriali senza sostenere gli investimenti necessari per la

raccolta delle notizie, per la selezione delle fonti e per la verifica delle informazioni. Qui si rischia allora effettivamente di disincentivare gli investimenti editoriali tradizionali, che sono però pur sempre necessari.

Per queste applicazioni dell'IA generativa, un qualche intervento normativo a protezione degli editori appare opportuno¹⁰². Si tratta, però, di cercare un compromesso tra protezione degli incentivi alla attività editoriale e protezione degli incentivi all'investimento nello sviluppo di innovazioni socialmente utili. Per le ragioni già espresse, il meccanismo dell'esclusiva non pare in grado di realizzare questo bilanciamento. Una soluzione più convincente è forse quella di attribuire, da un lato, agli operatori di IA il diritto di servirsi dei testi per l'addestramento e, dall'altro, alle imprese editoriali il diritto di ricevere un compenso ragionevole da parte degli operatori che offrono al pubblico servizi di *chatbot* o altre forme di IA che possono entrare in competizione con i prodotti editoriali. Per un approfondimento di questi aspetti si deve però, a questo punto, rinviare ad un futuro sviluppo del lavoro.

¹⁰² D'altra parte, l'interesse a tutelare i giornali contro attività parassitarie che ne mettano a rischio la tenuta è considerato meritevole di tutela dall'ordinamento. Basti pensare all'art. 101 l. aut., che vieta «la riproduzione sistematica di informazioni o notizie, pubblicate o radiodiffuse, a fine di lucro, sia da parte di giornali o altri periodici, sia da parte di imprese di radiodiffusione».