

Per l'Ufficio di presidenza della VII Commissione
(Cultura, scienza e istruzione) della Camera dei deputati

Sulla tutela del diritto d'autore
con particolare riferimento all'uso delle nuove tecnologie

I DATI COME ANTIDOTO E LEVA CONCORRENZIALE

Valeria Falce*

SOMMARIO: 1. I nuovi giganti– 2. Apertura dei dati e privative industriali: prospettive di bilanciamento e convergenza. – 2.1. Verso il governo dei dati consapevole. – 2.2. Il perno del bilanciamento tra interessi. – 3. Tensioni tra accesso e controllo dei dati per i sistemi e modelli di intelligenza artificiale – 4. La terza bozza del Codice di condotta per l'IA – 5. Conclusioni.

1. I NUOVI GIGANTI

Nell'ambito del duplice percorso di trasformazione economica (ecologica e digitale)¹ intrapreso in sede europea, quella digitale si caratterizza per il **primato dei dati come asset commerciale strategico**, la prevalente disintermediazione e il decentramento nei rapporti², anche rispetto alla fruizione dei contenuti, la

* Professore ordinario di diritto dell'economia e Jean Monnet (ad personam) Professor in Digital Transformation and AI Policy, Direttore dell'ICPC Innovation Regulation and Competition Policy Centre nell'Università Europea di Roma (valeria.falce@unier.it). Il presente scritto si inserisce nelle attività di ricerca promosse nell'ambito di un PRIN-Progetto di Ricerca di Interesse Nazionale dedicato alle piattaforme e riprende e aggiorna due lavori in via di pubblicazione rispettivamente in AGE-Analisi Giuridica dell'Economia e in Trattato – le piattaforme digitali. E-Agorà.

¹ EUROPEAN COMMISSION-DIRECTORATE-GENERAL FOR COMMUNICATION-U. LEYEN, *Political guidelines for the next European Commission 2019-2024*; Opening statement in the European Parliament plenary session 16 July 2019; Speech in the European Parliament plenary session 27 November 2019.

² In argomento, si rinvia a V. FALCE, *Strategia dei dati e Intelligenza artificiale. Verso un nuovo ordine giuridico del mercato*, Torino, 2023; V. FALCE-J. CANNATACI-O. POLLICINO, *Legal Challenges of Big data*, Cheltenham, UK, 2020; V. FALCE-G. GHIDINI-G. OLIVIERI, *Informazione e Big Data tra Innovazione e Mercato*, in *Quad. romani dir. comm.*, 2018; nonché V. FALCE, *Financial Innovation*

contaminazione e conseguente integrazione tra attività, prodotti e servizi, e di qui nell'emersione di nuovi attori, le piattaforme³.

Piattaforme, **dati e intelligenza artificiale** vengono (non a caso) elevati a **cifra distintiva della rivoluzione industriale**: nuova (la quinta) e unica nel suo genere (perché, mentre le precedenti sono state “accese” da una singola tecnologia – macchina a vapore, energia elettrica, computer –, la nuova rivoluzione si caratterizza per un insieme di tecnologie che, grazie ad Internet, si aggregano in modo sistemico). Sul mercato si affermano **nuovi intermediari** (le piattaforme), nuovi prodotti e servizi (interconnessi tramite Internet) e nuovi modelli di business, incentrati su dati e tecnologie abilitanti (*big data*, *data analytics* e sistemi cognitivi, ma anche *Internet of Things*, *cloud*, realtà aumentata, tecnologie abilitanti e robotica avanzata) che consentono la creazione e la condivisione di contenuti in via disintermediata.

Le piattaforme digitali, in questo senso, incidono ogni ambito dell'economia e della società, si propongono come il nuovo “interfaccia” del (e sul) mercato e, attraverso servizi e prodotti che soddisfano capillarmente le esigenze della domanda, offrono un enorme potenziale di crescita e di aumento del benessere.

Si inaugura così la prima fase, in cui i dati sono il nuovo petrolio, le piattaforme lo raffinano e da esse estraggono valore. Grazie ai *data analytics* e sistemi IA, le piattaforme vengono sempre più e meglio addestrate così da offrire servizi migliori. La concorrenza è “tra” e “all'interno” delle piattaforme, i mercati sono dinamici, i processi innovativi, disruptive e “veloci”. Per effetto di un processo di inarrestabile distruzione creativa, contesto e dinamiche evolvono rapidamente⁴. Le piattaforme spuntano rapidamente e rapidamente vengono spazzate via.

tra disintermediazione e mercato, Torino, 2021; V. FALCE-A. GENOVESE, *La portabilità dei dati in ambito finanziario*, Quaderno FinTech, CONSOB, 2021; V. FALCE-G. FINOCCHIARO, *Fintech: Diritti, Concorrenza, Regole*, Bologna, 2019; V. FALCE, *Competition Law Enforcement in Digital Markets*, Torino, 2021.

³ OECD (2024), *Competition Policy in Digital Markets. The Combined Effect of Ex Ante and Ex Post Instruments in G7 Jurisdictions*; OECD, *Big data: bringing competition policy to the digital era, Background note by the Secretariat*, 8; OECD, Hearing on Competition Economics of Digital Ecosystems held during the 134th Meeting of the Competition Committee on 1-3 December 2020.

⁴ Per una prima ricognizione della letteratura sul tema, si veda V.H.S.E. ROBERTSON, *Antitrust Market Definition for Digital Ecosystems*, Concurrences N. 2-2021/On-Topic/Competition policy in the digital economy, 3-9; M.J. JACOBIDES-I. LIANOS, *Ecosystems and competition law in theory and practice*, in *Industrial and Corporate Change*, 30(5), 2021, 1199-1229; D.A. CRANE, *Ecosystem Competition and the Antitrust Laws*, in *Nebraska Law Review*, 98(2), 2019, 412 ss.; M. JACOBIDES, *How to Compete When Industries Digitize and Collide: An Ecosystem Development Framework*, in *California Management Review*, 64(3), 2022, 99 ss.; M. JACOBIDES-C. CENNAMO-A. GAWER, *Towards a theory of ecosystems*, in *Strategic Management Journal*, 39(8), 2018, 2255 ss.; A. FLETCHER, *Digital competition policy: Are ecosystems different?*, DAF/COMP/WD(2020)96, 2; G. PETROPOULOS, *Competition Economics of Digital Ecosystems*, DAF/COMP/WD(2020)91, 5; M. BOURREAU, *Some Economics of Digital Ecosystems*, DAF/COMP/WD(2020)89, 5; M. BOURREAU-A. DE STREEL, *Digital Conglomerates and EU Competition Policy* (CERRE Report, 2019), 9-10; D.A. CRANE, *Ecosystem Competition*, DAF/COMP/WD(2020)67.

A questa fase ne succede una seconda che si caratterizza per la **persistenza di pochi ecosistemi digitali** che offrono prodotti e servizi tra loro tecnologicamente connessi e funzionalmente complementari⁵. Esempi di ecosistemi “compiuti” sono Google, che, in aggiunta al motore di ricerca, si espande su prodotti e servizi collegati, come i *browsers*, cioè i software per navigare su Internet, i sistemi operativi e il video streaming, o anche Facebook, che dal social network si “allarga” a prodotti e servizi contigui ma anche distanti, dal gaming alla messaggistica, dalla rivendita al dettaglio ai *devices*.

Nel nuovo contesto **l’ecosistema primeggia e dimensiona il mercato**, perché addestrato in maniera ineguagliabile. L’innovazione che viene introdotta è quella che sostiene l’ecosistema e non anche quella capace di interferire se non di scardinare i modelli di *business*. Le ragioni del primato degli ecosistemi sono diverse ma tutte convergenti. L’ecosistema, innanzitutto, per il fatto di offrire prodotti vuoi integrati vuoi complementari, è in grado di generare consistenti economie di scala perché con gli stessi fattori produttivi si possono produrre beni e servizi diversi e i risparmi aumentano all’aumentare delle quantità. Inoltre, quanto più si estende su mercati diversi, più si fortifica, perché il valore dell’ecosistema e dei suoi servizi aumenta all’aumentare del numero di utilizzatori. E consolida la propria posizione grazie ai dati, che, trattati e raffinati con sofisticate tecniche di IA, sono in grado di intercettare, prima, e creare, poi, nuovi bisogni, classificare emozioni, indirizzare orientamenti. Ancora, poiché le forme di “occupazione preventiva” di spazi e ambiti di possibile interesse commerciale assicurano un vantaggio competitivo, predilige collaborazioni intersettoriali e acquisizioni strategiche, ma anche forme innovative di appropriazione di beni immateriali⁶.

Sono queste caratteristiche “genetiche” che facilitano l’integrazione tra servizi e la convenienza a permanere all’interno dell’ecosistema, sollecitando una nuova riflessione⁷. Riflessione che si focalizza su una doppia dimensione di mercato:

⁵ Si permette di rinviare anche a N. FARAONE-V. FALCE, *Digital ecosystems in the wake of a legislative/regulatory turmoil: a first (tentative) antitrust assessment of the Italian (and European) experience in the AGCM case law*, in *World Competition*, vol. 46, Issue 1, 2023, 37.

⁶ Per una ricognizione di ordine generale, si veda M.A. CUSUMANO-A. GAWER-D.B. YOFFIE, *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*, HarperBusiness, 2019. Si rimanda a N. FARAONE-V. FALCE, *Digital ecosystems in the wake of a legislative/regulatory turmoil: a first (tentative) antitrust assessment of the Italian (and European) experience in the AGCM case law*, cit., 37 ss.

⁷ Nell’Unione, si rimanda, tra gli altri, alla decisione sul caso AT.40099 – Google Android, resa in data 18 luglio 2018 e alla successiva pronuncia di appello resa dal Tribunale, sentenza del 14 settembre 2022, causa T-604/18, Google e Alphabet/Commissione (Google Android), § 268 ss.; o anche Commissione Europea, caso AT.40437, Apple – App Store Practices (music streaming), 16 giugno 2020; Commissione Europea, caso AT.40652, Apple – App Store Practices (e-books/audiobooks), 16 giugno 2020; Commissione Europea, caso AT.40452, Apple – Mobile payments, 16 giugno 2020. Negli USA, si rimanda a Epic Games, inc. v. Apple Inc., case No. 4:20-cv-05640-YGR (N.D. Cal.), Judge Y. Gonzalez Rogers, 10 September 2021, 1 (“Epic v. Apple”), successivamente

l'una si svolge tra ecosistemi e l'altra si dipana al loro interno. Mentre questa ricorre quando le imprese che offrono un ecosistema di prodotti e/o servizi fronteggiano un vincolo competitivo esercitato anche da imprese "specializzate", che si pongono su un altro livello della catena del valore, la prima, rimanda ad una corsa competitiva tra sistemi di prodotti e servizi tra loro incompatibili, che inevitabilmente fidelizza il consumatore⁸, che preferisce rimanere al suo interno⁹.

2. APERTURA DEI DATI E PRIVATIVE INDUSTRIALI: PROSPETTIVE DI BILANCIAMENTO E CONVERGENZA

Come risposta alle fasi economiche, di emersione delle piattaforme prima e di imposizione degli ecosistemi poi, **l'Europa fa leva sui dati non solo in quanto "motore" della trasformazione¹⁰ ma anche come "antidoto" alle**

appellata: Epic Games, inc. v. Apple Inc., case 21-16506-21-16695, 15 July 2022 (9th Cir). Per un approfondimento, Competition and Markets Authority, Mobile Ecosystems Market Study, 10 June 2022. Si veda anche F. BOSTOEN-D. MĂNDRESCU, *Assessing abuse of dominance in the platform economy: a case study of app stores*, in *European Competition Journal*, 431(16), 2020, 7 ss.

⁸ C. MATUTES-P. REGIBEAU, "Mix and Match": product compatibility without network externalities, in *RAND Journal of Economics*, 19(2), 1988, 221-234.

⁹ Si veda P. AGHION-N. BLOOM-R. BLUNDELL-R. GRIFFITH-P. HOWITT, *Competition and innovation: an inverted-U relationship*, in *Quarterly Journal of Economics*, 120(2), 2005, 701 ss. e F. MARTY-T. WARIN, *Innovation in Digital Ecosystems: Challenges and Questions for Competition Policy*, CIRANO Cahier Scientifique 2020S-10, 4, secondo cui un'impresa dominante, anche se protetta da barriere all'ingresso, può avere incentivi a innovare, specialmente al di fuori del mondo digitale, allorché, ad esempio, commercializzano beni durevoli e l'innovazione sia necessaria per indurre i clienti a rinnovare le loro apparecchiature. In secondo luogo, può avere incentivi ad innovare specialmente se in concorrenza con altre piattaforme digitali. A tal proposito, tali aziende sono incentivate a innovare per garantire la continuità dei flussi di dati di cui dispongono e, quindi, aumentare le capacità predittive dei loro algoritmi. Le capacità predittive degli algoritmi sono tanto più elevate quanto più aggiornati e diversificati sono i dati. Più una piattaforma diversifica i propri servizi, migliore è la sua "performance predittiva". Infine, l'innovazione può essere un mezzo per sostenere ed espandere la propria posizione dominante. Inoltre, l'innovazione è un fattore essenziale anche per l'impresa fornitrice che opera all'interno dell'ecosistema, giacché il successo della sua strategia di *business* e della sua inclinazione dipende dall'appartenenza a un ecosistema dinamico. Allo stesso tempo, un basso tasso di innovazione può esporli a essere rimpiazzati da altre imprese che operano all'interno della piattaforma o, addirittura, escluse dalla piattaforma. Da ultimo, si rinvencono tre ulteriori ragioni per cui le imprese "complementari" che operano all'interno dell'ecosistema hanno un incentivo a innovare. Il primo motivo è la concorrenza tra i fornitori "complementari" stessi. La seconda ragione risiede nel rischio che la piattaforma integri il servizio fornito nella propria offerta se non soddisfatta della qualità del prodotto o delle prestazioni del servizio. La terza ragione è la necessità di diversificare il rischio se un cambiamento strategico guidato dalla piattaforma/ecosistema è astrattamente in grado di mettere in discussione la sostenibilità della sua tecnologia o del suo modello di *business*.

¹⁰ Sul ruolo dei dati e dei big data: J. CANNATACI-O. POLLICINO-V. FALCE, *Legal Challenges of Big data*, cit., *passim*; M. DELMASTRO-A. NICITA, *Big Data. Come stanno cambiando il nostro mondo*, Bologna, 2019; sulla concentrazione di dati da parte degli ecosistemi: M. POLO-A. SASSANO, *DMA:*

concentrazioni monopolistiche¹¹⁻¹².

Oggetto di programmi di investimento¹³ e perno di una profonda revisione regolatoria, è proprio sui dati e sulle tecnologie abilitanti che l'Unione decide di "scommettere" prima e "accelerare" poi per recuperare, se non assicurare, apertura dei mercati, autonomia digitale e sovranità tecnologica, così da affrancarsi dallo stato di sudditanza e dipendenza tecnologica dalle economie extra-europee¹⁴ e di qui rilanciare il modello europeo.

La partenza parrebbe in salita: il mercato dei dati genera, secondo le stime della Commissione europea, un valore economico (pari al 2,6% del PIL dell'UE-27) decisamente insoddisfacente per incertezze sulla titolarità dei diritti, squilibri nel potere negoziale, disponibilità limitata di servizi *cloud* affidabili e mancanza di interoperabilità intersettoriale dei dati nell'UE. A ciò si aggiunga che, nonostante il suo valore abbia sfiorato gli 82 miliardi di euro nel 2023, così sovrastando la Cina (il cui mercato si è assestato intorno ai 50 miliardi di euro), l'Europa sconta un significativo ritardo rispetto sia agli Stati Uniti (il cui mercato dei dati ha raggiunto un valore di 350 miliardi di euro) che alla Cina in termini di sviluppo e sfruttamento per ragioni riconducibili ad un più lento tasso di adozione delle tecnologie digitali, ad ostacoli strutturali agli investimenti nella digitalizzazione, alla frammentazione degli approcci politici, alla mancanza di riconoscimento dei vantaggi del digitale e al continuo divario di competenze digitali.

Il percorso e il traguardo da centrare sono definiti: invertire la marcia e superare le criticità riscontrate anche attraverso l'introduzione di regole facilitanti e abilitative ed insieme all'iniezione di capitali ed investimenti ingentissimi, per consentire al mercato dei dati di raggiungere un valore di quasi 120 miliardi di euro nel

Digital Markets Act o Digital Markets Armistice?, in *Merc. conc. reg.*, 2021, 529, i quali ritengono che "in questa architettura i dati raccolti rimangono nella disponibilità della piattaforma, accelerano la dinamica di rafforzamento della sua posizione dominante e costituiscono una barriera all'entrata"; S. MANNONI-G. STAZI, *Sovranità.com. Potere pubblico e privato ai tempi del cyberspazio*, Napoli, 2021; G. GHIDINI-D. MANCA-A. MASSOLO, *La nuova civiltà digitale. L'anima doppia della tecnologia*, Milano, 2020, 114.

¹¹ Mi si permetta di rinviare in proposito a V. FALCE (a cura di), *Digital Ecosystems. Market Challenges and pro-competitive solutions*, Torino, 2024, *passim*.

¹² Mi si permetta di rinviare, oltre ai contributi riportati alla nota 2, a V. FALCE, *Le regole sulle banche dati nella strategia europea: (molti) diritti e (poche) responsabilità*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Bologna, 2022; ID., *Intelligenza artificiale e Fintech. Consulenza decentrata tra algogovernance ed elemento fiduciario*, *ivi*; V. FALCE, *Financial Innovation tra Disintermediazione e Mercato*, Torino, 2021; V. FALCE-G. FINOCCHIARO, *Fintech: Diritti, Concorrenza, Regole*, Bologna, 2019.

¹³ Si rimanda per un approfondimento al link: https://eic.ec.europa.eu/eic-2025-work-programme_en.

¹⁴ M.A. CUSUMANO-A. GAWER-D.B. YOFFIE, *The business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*, Harper Business, 2019; N. FARAONE-V. FALCE, *Digital ecosystems in the wake of a legislative/regulatory turmoil*, *cit.*, 87 ss.

2030¹⁵ e a quello dell'IA e delle nuove tecnologie un valore di oltre 200 miliardi di dollari nel 2030¹⁶.

Le azioni programmate e attuate sul fronte regolatorio si inquadrano nella *European Data Strategy*. Si punta su regole uniformi (orientate al principio di armonizzazione massima) sia per **orientare mercati e servizi digitali alla trasparenza, correttezza e contendibilità**¹⁷, sia per **favorire l'accesso e il riuso di dati pubblici e privati**. In particolare, con i Regolamenti Data Governance Act¹⁸ e Data Act¹⁹ l'Europa opta per un sistema regolatorio orizzontale e intersettoriale per l'accesso, il riutilizzo e la condivisione dei dati, promuovendo la creazione di *Spazi Comuni* in settori strategici e ambiti di interesse pubblico. Con l'AI Act disciplina l'immissione sul mercato UE dei sistemi di intelligenza artificiale, seguendo un approccio *ex ante* basato sul rischio²⁰. Con Eidas 2, intende garantire a imprese e cittadini una identità digitale europea volta a facilitare rapporti e consentire l'instaurazione di nuove relazioni su base transnazionale²¹. Con il pacchetto sulla *cybersicurezza*, crea i presupposti per salvaguardare la sicurezza e l'integrità dei dati all'interno dei confini europei.

Tuttavia, al contempo ed in **senso apparentemente diacronico**, l'Europa insiste per l'introduzione di forme di tutela efficienti, trasparenti ed equilibrate a favore di chi, investendo in innovazione e creatività, contribuisce al progresso tecnologico e

¹⁵ IDC, Studio sul mercato europeo dei dati 2021-2023, 2024, accessibile all'indirizzo <https://digital-strategy.ec.europa.eu/it/library/results-european-data-market-study-2021-2023>.

¹⁶ Si rinvia per una prima indicazione alle informazioni e materiali reperibili su <https://www.assonime.it/Eventi/gruppi/Pagine/Gruppo-di-giunta-Data-Strategy-Europea.aspx>.

¹⁷ Da ultimo: Regolamento sul Digital Market Act 2022/1925/UE; Regolamento sul Digital Service Act 2022/2065/UE. Su cui, si v., per un'introduzione, G. CAGGIANO-G. CONTALDI-P. MANZINI (a cura di), *Verso una legislazione europea su mercati e servizi digitali*, Bari, 2022, 205 ss., nonché essenzialmente sul primo, V. FALCE (a cura di), *Competition law enforcement in digital markets*, cit.

¹⁸ Assonime, Circolare 17/2024 – Data Governance Act: ambiti di attuazione e spunti di policy per le imprese, disponibile all'indirizzo <https://www.assonime.it/attivita-editoriale/circolari/Pagine/Circolari.aspx>.

¹⁹ Assonime, Circolare 9/2024 – Le nuove regole europee sull'accesso ai dati e sul loro utilizzo (“Data Act”). Opportunità e sfide per le imprese, disponibile all'indirizzo <https://www.assonime.it/attivita-editoriale/circolari/Pagine/Circolari.aspx>.

²⁰ Assonime, Circolare 14/2024 – Il regolamento europeo sull'intelligenza artificiale: analisi ragionata della nuova disciplina e prospettive di policy per le imprese, disponibile all'indirizzo <https://www.assonime.it/attivita-editoriale/circolari/Pagine/Circolari.aspx>. Per tutti, A tal riguardo cfr. A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione*, vol. 1. *Diritti fondamentali, dati personali e regolazione*; vol. 2. *Amministrazione, responsabilità, giurisdizione*; vol. 3. *Proprietà intellettuale, società, finanza*, Bologna, 2022.

²¹ Assonime, Circolare 20/2024 – Dall'EUDI all'IT wallet: le prospettive per le imprese dall'identità digitale al portafoglio di attributi, disponibile all'indirizzo <https://www.assonime.it/attivita-editoriale/circolari/Pagine/Circolari.aspx>.

alla diffusione della cultura. Con le conclusioni del Consiglio del 2020²² la Commissione UE, infatti, promuove la creazione di diritti europei (come il brevetto europeo), la modernizzazione degli istituti esistenti (come quelli sul diritto d'autore), la curvatura pro-concorrenziale *by design* dei diritti proprietari (come la proposta di regolamento sui brevetti essenziali) e la promozione di un *enforcement* capace di contrastare pirateria e abusi (anche attraverso il Digital Service Act) a livello unionale²³.

In estrema sintesi e a prezzo di qualche semplificazione, **la Strategia si propone come innovativa e a tratti propulsiva**: ora forza il principio della libera negoziazione obbligando l'apertura di talune categorie di dati ed amministrando i rapporti, ora riconduce ad unità lo **statuto del dato che, indipendentemente dalla natura, viene qualificato come commodity e, per questa via, codifica il principio della monetizzazione**, ora eleva ad un nuovo status l'intermediario dei dati, che al ricorrere di specifici requisiti, gode di una speciale legittimazione per dare impulso a mercati e *data space*²⁴, ora indirizza i mercati nel senso di un maggiore ordine, di una più marcata tutela dei diritti e di una più chiara responsabilizzazione dei rapporti, ora infine li regola in linea con i principi della correttezza e della concorrenza.

Alla luce di questa doppia spinta, l'Europa, nel contenere gli spazi di autonomia degli Stati, parrebbe seguire un andamento ondivago se non schizofrenico sul fronte della *policy*. Per un verso, sembra voler sovvertire i paradigmi classici, imponendo o comunque orientando Stati membri e stakeholder all'apertura, la condivisione e la riutilizzazione dei dati, così da alimentare spazi europei e relativi mercati, favorire collaborazioni intra e intersettoriali e facilitare il recupero della distanza competitiva rispetto a USA e Cina. Per altro verso, pare rinnovare una certa deriva (iper)protezionistica, rafforzando la tutela proprietaria, le sue prerogative e le forme di protezione dei beni immateriali, nel presupposto che, nella nuova economia, siano al cuore del posizionamento delle imprese sul mercato.

A guardar meglio e come si dirà subito, l'aporia è solo apparente e le due forze convergono in un **nuovo ordine giuridico del mercato**²⁵ in cui la nuova società digitale²⁶, create le condizioni per ridurre la distanza in termini di forza economica

²² Accessibile all'indirizzo Internet https://single-market-economy.ec.europa.eu/industry/strategy/intellectual-property/intellectual-property-action-plan-implementation_en.

²³ Per un esame delle traiettorie e stagioni europee della disciplina della proprietà intellettuale, mi si permetta di rinviare a V. FALCE-G. GHIDINI, *Disciplina della proprietà intellettuale: un volto spiccatamente europeo*, in *Europa*, a cura di Amato-Moavero Milanesi-Pasquino-Reichlin, Roma, 2018, 335.

²⁴ Cfr. V. FALCE, *Regolazione delle piattaforme e tutela antitrust*, in R. BOCCHINI (a cura di), *Manuale di diritto privato dell'informatica*, 2024, 61.

²⁵ V. FROSINI, *L'orizzonte giuridico dell'Internet*, in *Dir. inf.*, 2000, 271 ss.; T.E. FROSINI, *L'orizzonte giuridico dell'intelligenza artificiale*, in *Dir. inf.*, 2022, 5 ss.

²⁶ Cfr. V. CODELUPPI, *Mondo digitale*, Roma-Bari, 2022.

e commerciale nei confronti delle superpotenze extraeuropee, possano intanto sfruttare i benefici di mercati dei dati, aperti e contendibili, se ed in quanto ne riconoscano il valore specifico e complessivo²⁷.

2.1. Verso il governo dei dati consapevole

Volgendo lo sguardo verso la prima direttrice, il rinnovamento muove dalla cornice, cioè dalla definizione di regole che traghettino gli Stati membri verso un modello unico e condiviso di governo e circolazione dei dati²⁸.

Anticipato dalla direttiva *Open Data*, il *Data Governance Act* è un Regolamento che aggiunge un ulteriore tassello alla creazione del mercato unico per lo scambio e il riutilizzo dei dati (detenuti da enti) pubblici, stabilendo un minimo comun denominatore per agevolarne la circolazione a livello intraeuropeo.

Obiettivo del DGA è stabilire a) le condizioni per il riutilizzo, all'interno dell'Unione Europea, di determinate categorie di dati (detenuti da enti) pubblici; b) una base giuridica per gli operatori digitali che facilitano la condivisione e lo scambio di questi stessi dati; c) un quadro per la registrazione volontaria dei soggetti che raccolgono e trattano i dati a fini altruistici; e d) una cornice per l'istituzione di un comitato europeo per l'innovazione in materia di dati (art. 1).

A rientrare nel perimetro del Regolamento sono le categorie di dati (detenuti da enti) pubblici, riconducibili alla normativa sui diritti di proprietà intellettuale, sul segreto commerciale o statistico, ovvero sulla protezione dei dati personali (art. 3). All'ambito di applicazione del *Data Governance Act* è perciò attratto ogni dato pubblico che possa essere considerato "sensibile" vuoi rispetto alla sfera personale vuoi rispetto alla dimensione imprenditoriale.

Ferme le prerogative nazionali nei settori strategici (sicurezza pubblica, difesa e sicurezza nazionale) il Regolamento per un verso allarga la base del mercato dei dati e per altro verso incide le libertà nazionali di regolazione dei rapporti, riportandole a una base comune, anzi uniforme.

La premessa di fondo è che gli enti pubblici (autorità statali, regionali o locali, organismi di diritto pubblico o associazioni formate da una o più di tali autorità oppure da uno o più di tali organismi di diritto pubblico) siano liberi di consentire o negare l'accesso, l'uso e il riutilizzo di dati pubblici.

Ove esercitino questa facoltà, gli enti pubblici "che detengono dati protetti per ragioni di a) riservatezza commerciale; b) riservatezza statistica; c) protezione dei diritti di proprietà intellettuale di terzi; d) protezione dei dati personali" sono gravati da alcuni oneri. In particolare, a livello transfrontaliero, devono indirizzare le

²⁷ Già in questo senso, V. FALCE, *Così l'Europa valorizza i contenuti digitali*, Milano Finanza, 23 settembre 2024, accessibile all'indirizzo internet <https://www.uer.it/jeanmonnetchair/digitrai/disseminations/>.

²⁸ Per una analisi, per tutti, Circolare Assonime, cit.

proprie condotte, nel senso di:

a) rinunciare, in linea con i principi del diritto *antitrust*, ad accordi o altre pratiche di esclusiva per il riutilizzo dei dati, a meno che l'esclusiva non sia necessaria alla fornitura di un servizio o di un prodotto di interesse generale che non sarebbe altrimenti possibile. Il periodo di esclusiva del diritto di riutilizzo dei dati non deve superare i tre anni (art. 4);

b) rendere pubbliche e trasparenti le condizioni di riutilizzo, che vengono ancorate a principi concorrenziali (dovendosi trattare di condizioni “*non discriminatorie, proporzionate e oggettivamente giustificate in relazione alle categorie di dati, alle finalità del riutilizzo e alla natura dei dati per i quali è consentito il riutilizzo*”) (art. 5, par. 1);

c) indirizzare i rapporti con i terzi, così da obbligarli a: 1) riutilizzare solo dati previamente anonimizzati/pseudonimizzati, 2) accedere e riutilizzare i dati solo all'interno di un ambiente anche fisico di trattamento sicuro, fornito e controllato dal settore pubblico, così riconducendo a regole comuni anche i dati ad apertura limitata e rispetto ai quali è richiesta l'attuazione di soluzioni tecniche;

d) garantire attraverso condizioni uniformi l'integrità dei sistemi tecnici e dell'ambiente di trattamento sicuro utilizzato, la riservatezza dei dati ed il rispetto dei diritti di proprietà intellettuale. A ciò si aggiunga che l'ente pubblico deve poter verificare i risultati del trattamento dei dati effettuato dal riutilizzatore e riservarsi il diritto di vietare l'uso dei risultati che contengono informazioni che compromettono i diritti e gli interessi di terzi (art. 5, par. 5).

La standardizzazione delle condizioni si estende ai profili economici, sia prevenendo che le tariffe (art. 6) del trasferimento del dato siano pubbliche, non discriminatorie, oggettive e proporzionate anche nella metodologia, sia aprendo il varco al cosiddetto uso altruistico, cioè alla condivisione volontaria dei dati per obiettivi di interesse generale, che il DGA caldeggia.

Vengono imposti, poi, obblighi rigorosi agli intermediari che si occupano di facilitare la condivisione e lo scambio dei dati. In proposito, per evitare situazioni di scorretto vantaggio competitivo e utilizzo improprio dei dati, il *Data Governance Act* richiede agli operatori digitali di assumere un comportamento puramente neutrale nell'attività di intermediazione, sia imponendo agli stessi di non “*utilizzare i dati per i quali fornisce servizi per scopi diversi dalla messa a disposizione di tali dati agli utenti dei dati*” sia di utilizzare “*i metadati raccolti nel corso della fornitura del servizio di condivisione dei dati solo per lo sviluppo di tale servizio*” (art. 11).

Il raggio d'azione del nuovo quadro normativo si estende ulteriormente anche alla condivisione di dati per scopi altruistici ovvero ad un utilizzo dei dati finalizzato a perseguire interessi generali, quali la ricerca scientifica o il miglioramento dei servizi pubblici. Si prevede, pertanto, la realizzazione di un “registro delle organizzazioni per l'altruismo dei dati riconosciute” (art. 15) all'interno del quale le organizzazioni interessate possono registrarsi solo in presenza di una serie di

requisiti (art. 16) e di specifici obblighi (art. 18) di trasparenza e di tutela dei diritti degli interessati²⁹.

Attraverso il Regolamento l'Europa puntella la correttezza e la trasparenza dei rapporti, il rispetto della concorrenza e delle libertà fondamentali sul mercato dei dati. Ma soprattutto conferma la *fairness* come principio generale e cartina di tornasole della *Data Strategy*³⁰.

Sullo slancio delle regole di cornice e di procedura, l'Europa è andata oltre sul fronte sostanziale. Il quadro abilitante si è arricchito di un ulteriore tassello, che intacca il paradigma proprietario, imponendo l'accesso e facilitando l'uso e il riuso dei dati (per il momento solo) dell'internet delle cose (IoT), stabilendo chi possa utilizzare i dati generati da prodotti smart, a quali condizioni e in quali contesti³¹.

Il Data Act disciplina accesso e uso ai dati generati dall'uso di prodotti e servizi connessi, diversificando tra rapporti B2C e B2B. In relazione ai primi è sia assicurata l'accessibilità gratuita, semplice, sicura e in formato strutturato ai dati che l'utilizzatore ha contribuito a creare, sia garantito il possibile uso e riuso commerciale di quegli stessi dati a condizioni di mercato. Dati e informazioni che assumono valore commerciale per il produttore e che, per via di quel valore, sono mantenuti in regime di riservatezza sfuggono al paradigma dell'apertura e rimangono coperti dalla tutela offerta dal segreto industriale.

Per i secondi (rapporti B2B) è apprestato un sistema efficace di protezione da squilibri contrattuali e pratiche abusive (*unfair contractual terms*). Il regolamento separa le clausole *black* che, unilateralmente imposte, sono nulle in quanto manifestamente abusive o scorrette, da quelle *grey*, la cui abusività è presunta e come tale superabile con prova contraria. Si sancisce la ragionevolezza di eventuali compensi per la messa a disposizione dei dati, con facoltà di ricorso ad organismi di composizione stragiudiziale delle controversie (per questioni relative al compenso o all'abusività delle clausole).

Sfugge a questa bipartizione il settore pubblico che, in caso di circostanze eccezionali ed emergenze pubbliche, può rivendicare l'accesso ai dati dei privati, anche per l'elaborazione statistiche e la creazione di standard di interoperabilità (contesto B2G).

Nuove regole intendono facilitare il passaggio tra servizi *cloud*, *edge* e altri

²⁹ Quanto all'*enforcement*, è prevista l'istituzione di un "Comitato Europeo per l'innovazione in materia di dati" a cui vengono affidati i compiti di consigliare e assistere la Commissione nel rafforzare l'interoperabilità dei servizi di intermediazione dei dati ed elaborare orientamenti sulle modalità con cui agevolare lo sviluppo degli spazi di dati (art. 27).

³⁰ N. FARAONE-V. FALCE, *Digital ecosystems in the wake of a legislative/regulatory turmoil: A first (tentative) antitrust assessment of the Italian (and European) experience in the AGCM case law*, cit., 37.

³¹ Per una prima analisi, Circolare Assonime, cit.

servizi di trattamento dati (*Data processing services* – DPS), e lo sviluppo di *standard* di interoperabilità per i dati da trasferire e riutilizzare tra i diversi settori industriali. Nel passaggio tra fornitori viene preservata la c.d. equivalenza funzionale, senza imporre interfacce tecniche e *standard* di interoperabilità, ma indirizzando il mercato verso *standard* aperti, se esistenti, conformi alla normativa europea. Vengono fornite garanzie contro il trasferimento illegale di dati non personali. Per gli spazi di dati, il Data Act stabilisce regole minime per l'interoperabilità e per i contratti intelligenti (*smart contracts*) in linea con la tendenza a ridurre gli spazi di libera negoziazione, che nella direttiva Copyright ha trovato un primo rilevantisimo approdo³².

2.2. Il perno del bilanciamento tra interessi

La spinta verso l'apertura, l'uso e il riuso dei dati sembra frenata dalla seconda traiettoria, la Strategia sulla proprietà intellettuale, che, marcando le tappe di un processo di modernizzazione degli istituti proprietari a favore dei beni immateriali, sul fronte sia genetico, che funzionale, pare indirizzare l'Unione europea verso una deriva protezionistica.

Approdi del primo fronte sono, come noto, le nuove regole in materia autoriale (con le quali si è inteso adeguare la materia alle nuove tecnologie abilitanti, contenere in maniera dinamica ed uniforme il sistema delle eccezioni e favorire le forme di circolazione delle opere dell'ingegno nell'economia digitale), anticipate da quelle in materia di segreto industriale (con le quali si è inteso uniformare la nozione e armonizzarne le regole per la tutela civilistica in caso di infrazione) e quelle relative all'istituzione di un nuovo titolo proprietario, il brevetto unitario (che, superando il principio della territorialità, nasce come titolo a curvatura europea direttamente azionabile e proteggibile nei 27 Stati Membri). Significativa espressione del secondo fronte sono, tra l'altro, le regole europee per la protezione dei diritti di proprietà intellettuale, che, per un verso, responsabilizzano piattaforme e fornitori di intermediazione e, per altro verso, impongono oneri di intervento tempestivo per la rimozione di contenuti illegali o contraffatti.

A guardare meglio, tuttavia, ciascuna delle azioni che declina la Strategia dei dati e quella IP muove da un punto fermo: che il titolare dei beni immateriali (dati o contenuti che siano) ne disponga, decidendo a che uso destinarli e con quale remunerazione.

Limitandoci ai dati commerciali, il *Data Governance Act* dispone che soggetti

³² Quanto alla governance, il regolamento rimanda ad ogni Stato membro l'approntamento delle garanzie necessarie al rispetto delle norme e alla trattazione di reclami e denunce. Alla Commissione, invece, è riservata la definizione di clausole contrattuali tipo per l'accesso e il riutilizzo dei dati, l'adozione di atti delegati per il controllo dei costi di switching nei DPS, nonché di prescrizioni per l'interoperabilità.

pubblici e privati possano concedere – a condizioni di mercato o anche gratuitamente e per le finalità che autorizzino – l’accesso ai dati e contenuti tutelati da diritti di proprietà intellettuale, di cui possano disporre.

Nella stessa direzione si muove il *Data Act*. Che introduce a livello unionale un nuovo diritto a favore degli utenti dei dati generati da prodotti ovvero nell’ambito di servizi intelligenti. Si tratta del diritto ad ottenerli gratuitamente, per poi sfruttarli o direttamente o tramite terzi, per usi da concedere a condizioni gratuite o di mercato.

L’AI Act³³ si pone nella stessa scia ed, anzi, la prosegue specialmente in relazione ad una specifica tecnologia, quella dei modelli generali, che sono progettati, impostati e sviluppati per essere addestrati con dati, immagini e suoni. In particolare, per il nuovo Regolamento, titolari di dati e contenuti devono mantenerne il controllo attraverso i diritti di proprietà intellettuale anche rispetto ai fornitori di sistemi e modelli IA che, come si sa, non possono essere progettati, impostati e sviluppati se non a fronte di un costante addestramento con dati di qualità (anche testi, immagini, suoni).

Fermo il rispetto del diritto unionale per tutti i sistemi e modelli, l’AI Act grava questi fornitori di un doppio obbligo di risultato: i) attuare una politica volta ad adempiere alla normativa dell’Unione in materia di diritto d’autore e, in particolare, a individuare e rispettare, anche attraverso tecnologie all’avanguardia, la riserva dei diritti disciplinata dalla direttiva Copyright; ii) redigere e mettere a disposizione del pubblico una sintesi sufficientemente dettagliata dei contenuti utilizzati per l’addestramento.

Si tratta di **obblighi organizzativi e di processo** che non interferiscono sulla sostanza e sulla tutela dei diritti, né esauriscono le intersezioni ed interferenze con la tutela autoriale, che rimane intatta, limitandosi a favorire il rispetto di regole che già esistono sull’estrazione di dati, richiedendo che vengano applicate

³³ Al fine di comprendere la portata è utile ricordare che il nuovo Regolamento definisce la cornice regolatoria per l’immissione in commercio di sistemi e modelli IA in Europa. Definisce i caratteri minimi dell’intelligenza artificiale (software con capacità inferenziale), riconosce nella trasparenza e nella alfabetizzazione le regole minime cui devono conformarsi fornitori e utilizzatori, gradua gli obblighi sulla base della loro pericolosità sino al punto di vietare alcuni sistemi e modelli (se hanno ad esempio come oggetto o effetto il *social scoring* o le pratiche manipolative). Proprio perché non introduce il codice europeo dell’IA né si propone come testo unico delle tecnologie abilitanti, non sorprende che il regolamento richiami il marchio al solo fine di qualificare chi lo appone come il soggetto responsabile degli obblighi che ricadono sul fornitore del sistema o modello e per responsabilizzare i soggetti della catena del valore, incluso il *deployer*, cioè l’utilizzatore professionale, che, modificando il sistema per funzioni o finalità, subentra nello status di fornitore. Insomma, all’uso del marchio, l’AI Act fa corrispondere una specifica qualità soggettiva e da questa fa discendere uno specifico statuto, lo statuto del fornitore di IA. Né sorprende che il Regolamento ribadisca e confermi il segreto industriale, perché gli obblighi di trasparenza che gravano sui fornitori, ridimensionati nel caso in cui i sistemi vengano rilasciati in modalità aperta e open source, non devono interferire e pregiudicare la proprietà industriale.

indipendentemente dal luogo di addestramento dei modelli e dalla relativa giurisdizione, dalle modalità di diffusione (licenze aperte e chiuse) e dalla dimensione dei fornitori.

Perché introdurli allora? Perché **l'Europa spinge per la parità di trattamento tra i diversi fornitori e pretende che le regole che già esistono sull'estrazione di dati, testi, contenuti, suoni e immagini, vengano applicate indipendentemente dal luogo di addestramento dei modelli e dalla relativa giurisdizione**, dalle modalità di diffusione (licenze aperte e chiuse) ed indipendentemente dalla dimensione dei fornitori. Nessun fornitore, ricorda il considerando 106, dovrebbe essere in grado di ottenere un vantaggio competitivo nel mercato dell'Unione applicando norme in materia di diritto d'autore meno rigorose di quelle previste nell'Unione.

A tal fine, **l'eccezione di estrazione di testo**, dati, suoni e immagini, i.e. *Text and Data Mining*, già disciplinata a livello europeo sia per scopi di ricerca (che deve essere svolta senza finalità lucrative e non deve contrastare con gli interessi dei titolari dei diritti), sia per finalità commerciali e lucrative (come per l'assunzione di "decisioni commerciali complesse e lo sviluppo di nuove applicazioni o tecnologie") è **ribadita**. Come pure è confermato, in linea con il diritto dell'Unione ma anche con i principi internazionali, che si tratta di **un'eccezione non assoluta**³⁴, né incondizionata, ma subordinata al rispetto di presupposti specifici e generali. Tra i primi: che ai contenuti si abbia legalmente accesso (grazie a condizioni contrattuali o abbonamenti), che i contenuti non siano riservati e che l'accesso per la finalità di *text e data mining* sia concesso. Tra i secondi: che l'uso di quei materiali non contrasti con il normale sfruttamento delle opere o altri materiali e non arrechi indebitamente pregiudizio ai legittimi interessi dei titolari dei diritti³⁵.

³⁴ "Tutte le eccezioni, infatti, sono sottoposte a limitazioni" prima fra tutte quella di essere condizionate al rispetto del *test* a tre fasi (nel senso che devono essere definite con precisione e interpretate in modo restrittivo, non devono interferire con il "normale sfruttamento dell'opera" e non devono "pregiudicare in modo irragionevole il legittimo interesse del titolare dei diritti". Cfr. R. DUCATO-A. STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, 2021, disponibile su https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3829858, ultimo accesso 15 dicembre 2024. Si veda inoltre J.P. QUINTAIS, *The New Copyright in the Digital Single Market Directive: A Critical Look*, 2019, disponibile su https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3424770, ultimo accesso 15 dicembre 2024.

³⁵ Cfr. T. SCHONWETTE, *The three-step test within the copyright system*, available at <http://pcf4.dec.uwi.edu/viewpaper.php?id=58&print=1>; L. GUIBAULT, *Copyright Limitations and Contracts – An Analysis of the Contractual Overridability of Limitations on Copyright*, The Hague, 2002; M. FICSOR, *The Law of Copyright and the Internet – The 1996 WIPO Treaties, their Interpretation and Implementation*, Oxford, 2002; D. GERVAIS, *Towards a New Core International Copyright Norm: The Reverse Three-Step Test*, in *Marquette Intellectual Property Law Review*, vol. 9, no. 1, 2005, 1-37; J.C. GINSBURG, *Toward Supranational Copyright Law? The WTO Panel Decision and the "Three-Step Test" for Copyright Exceptions*, Working Paper No. 181 of the Columbia Law School – *For Revue Internationale du Droit d'Auteur*, 2001, 1-16; Implications of the TRIPS Agreement on Treaties Administered by WIPO / World Intellectual Property Organisation, WIPO publication No.

Il Regolamento, pertanto, non innova la sostanza, ma aiuta a presidiarla. Per agevolare i terzi e soprattutto i titolari dei diritti d'autore, impone ai fornitori di introdurre, rispettare e far misure di progettazione e sviluppo, tra cui quelle per l'identificazione automatica delle eventuali riserve dei diritti³⁶, nonché di pubblicare una sintesi delle banche dati e degli archivi di dati privati o pubblici usati per l'addestramento e una descrizione delle altre principali fonti³⁷.

3. Tensioni tra accesso e controllo dei dati per i sistemi e modelli di intelligenza artificiale

È nell'ambito della Strategia europea sui dati che si inquadrano, dunque, le intersezioni ed interferenze della legislazione dell'Unione in materia di proprietà intellettuale (PI) con la stagione delle innovazioni tecnologiche, a partire dalle tecnologie abilitanti e dall'IA, di cui si avvalgono piattaforme ed ecosistemi *in primis*.

Si tratta di interferenze solo apparentemente nuove, perché rimandano alla tensione mai sopita tra spinte all'apertura dei dati, in quanto volano di concorrenza dinamica, accesso sociale all'informazione, alla cultura e all'innovazione, e le resistenze dei titolari di diritti di proprietà intellettuale, che rivendicano uno *ius excludendi alios* su dati e contenuti, pena l'appropriazione indebita e l'esercizio di un ingiustificato comportamento di *free riding* a fronte di investimenti e sforzi creativi ed inventivi.

Un nuovo capitolo di questa storia, all'insegna di un bilanciamento moderno, è

464 (E) (1997), Geneva; R. KNIGHTS, *Limitations and Exceptions Under the "Three-Step Test" and in National Legislation – Differences Between the Analog and Digital Environments*, WIPO document WIPO/DA/MVD/00/4, Geneva, 2020; J. OLIVER, PANEL DISCUSSION: Copyright in the WTO: The Panel Decision on the Three-Step Test, 25 *Columbia Journal of Law & the Arts*, 2002, 119-170; S. RICKETSON, *The Berne Convention for the Protection of Literary and Artistic Works: 1886-1986*, The Hague, 1987; S. RICKETSON, *WIPO Study on Limitations and Exceptions of Copyright and Related Rights in the Digital Environment*, WIPO publication SCCR/9/7, Geneva, 2003; M. SENFTLEBEN, *Copyright, Limitations and the Three-Step Test – An Analysis of the Three-Step Test in International and EC Copyright Law*, The Hague, 2004; WTO Copyright Panel decision (2000), case WT/DS160, United States – Section 110 (5) of the US Copyright Act.

³⁶ A. PEUKERT, *Copyright in the Artificial Intelligence Act – A Primer*, 2024, disponibile su https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4771976, ultimo accesso 15 dicembre 2024.

³⁷ Per chiudere il cerchio, l'AI Act chiede che i sistemi IA siano progettati in modo da informare gli utilizzatori sia nel caso in cui stiamo interagendo direttamente con un sistema di intelligenza artificiale, sia in relazione agli output del sistema di IA, che devono essere marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente. Di più, gli stessi utilizzatori professionali devono chiarire che stanno usando contenuti generati o manipolati artificialmente per informare il pubblico. Per sigillare la conformità con la disciplina europea, il Regolamento prevede, infine, che quest'ultimo obbligo non si applichi se il contenuto sintetico sia stato sottoposto a processo di revisione umana o a controllo editoriale e se pertanto una persona fisica o giuridica ne assume la responsabilità editoriale.

d'altra parte in corso e si può apprezzare guardando ai sistemi di intelligenza artificiale, oggi definiti in sede europea nel Regolamento sull'IA, che ne costituiscono la più recente cartina di tornasole. Se, infatti, tali sistemi nella prospettiva IP attraggono una combinazione di soggetti e diritti IP (in termini di *know how*, segreto industriale, brevetti, tutela autoriale e sulle banche dati), d'altra parte sono le loro stesse modalità di funzionamento a sollecitare una nuova riflessione rispetto sia alle opere/risultati prodotti dall'IA – l'*output* – sia rispetto ai dati, ai contenuti, ai suoni e alle immagini di cui questi stessi sistemi devono nutrirsi per addestrarsi e per addivenire a un certo risultato – l'*input*.

In questa specifica prospettiva, le intersezioni e le interferenze tra IA e dati possono essere apprezzate almeno a due livelli e in due fasi distinte: 1. “a monte”, quando dati, informazioni e contenuti vengono utilizzati per addestrare sistemi e modelli di IA. In questo senso, i sistemi sono alimentati da opere protette da *copyright*, la cui versione digitale – collegata all'originale, sia esso analogico o digitale³⁸ – deve essere acquisita dal sistema nel rispetto delle prerogative autoriali; 2. “a valle”, quando il sistema di IA, sulla base dei dati, contenuti e informazioni estratti da un'opera originale, elabora una sintesi o fornisce una nuova elaborazione di dati, che non può essere ricondotta direttamente e specificamente a nessuno dei contenuti utilizzati se non come fonte di ispirazione, generando così un'opera nuova o recante (almeno) un contenuto “trasformativo”.

Limitandoci per il momento al primo livello di intersezione (opere protette da *copyright* come materiale di addestramento), il bilanciamento si trova nel raccordo tra le previsioni la direttiva (UE) 2019/790 sul diritto d'autore e sui diritti connessi nel mercato unico digitale (“Direttiva Copyright”)³⁹, con particolare riferimento all'eccezione di *text and data mining* (“TDM”)⁴⁰ con quelle del Regolamento

³⁸ Cfr. P.B. HUGENHOLTZ-J.P. QUINTAIS, *Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?*, in *I.I.C. – International Review of Intellectual Property and Competition Law*, 2021, vol. 52, 1190; D. GERVAIS, *The Machine as Author*, in *Iowa Law Review*, 2020, vol. 105, 2053; J.C. GINSBURG, *People Not Machines: Authorship and What It Means in the Berne Convention*, in *IIC*, 2018, vol. 49, 131; J.C. GINSBURG-L.A. BUDIARDJO, *Authors and Machines*, in *Berkeley Technology Law Journal*, 2019, vol. 34, Issue 2, 343; A. GUADAMUZ, *Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works*, in R.M. HILTY-J.-A. LEE AND K.-C. LIU (eds.), *Artificial Intelligence and Intellectual Property*, Oxford, 2021, 147.

³⁹ Si rimanda al Capitolo V *infra* per l'analisi puntuale della direttiva Copyright.

⁴⁰ Si rimanda a J. DREXL-R.M. HILTY *et al.*, *Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective*, Max Planck Institute for Innovation and Competition Research Paper No. 19-13 (2019), available on SSRN: <https://ssrn.com/abstract=3465577>; A. GUADAMUZ, *A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs*, in *GRUR International*, 73, 2024, 2, 111, disponibile al link: <https://academic.oup.com/grurint/advance-article/doi/10.1093/grurint/ikad140/7529098?login=false>; N. LUCCHI, *ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems*, in *European Journal of Risk Regulation*, 2023, 1, disponibile al link: [//doi.org/10.1017/err.2023.59](https://doi.org/10.1017/err.2023.59).

sull'IA.

Minimo comune denominatore tra le une e le altre è, infatti, che ogni eccezione al diritto d'autore, inclusa quella di *Text and Data Mining* – tecnica di estrazione che si articola in tre principali fasi: i) accesso a testi e/o dati; ii) estrazione e/o copia di contenuti; iii) estrazione di testi e/o dati e scoperta di conoscenze) – deve essere applicata nel rispetto delle condizioni alla base del *Three-step Test*, pena la violazione del diritto nazionale, europeo ed internazionale⁴¹.

Ma procediamo con ordine. Punto di partenza per inquadrare il sistema delle eccezioni in generale e di qui quella di *text and data mining* sono le disposizioni introdotte dal Capo II della direttiva InfoSoc, adottata nel 2001 con l'obiettivo di adattare le leggi sul diritto d'autore dell'UE alle sfide poste dagli ambienti digitali e *on-line*. Con la finalità, infatti, di armonizzare le discipline autoriali tra gli Stati membri dell'UE, ridurre la frammentazione giuridica e garantire che le norme sul diritto d'autore siano applicate in modo uniforme in tutta l'UE, la direttiva InfoSoc rafforza la tutela del diritto d'autore e dei diritti connessi nell'ambito del mercato interno, contemperandoli con i diritti fondamentali e l'interesse pubblico.

Il considerando 31 richiede specificamente che venga raggiunto un “giusto equilibrio” tra i diritti e gli interessi delle varie categorie di titolari nonché tra quelli dei vari titolari e quelli degli utenti dei materiali protetti, con l'aspettativa che “*le eccezioni e limitazioni alla protezione esistenti nelle legislazioni degli Stati membri devono essere riesaminate alla luce del nuovo ambiente elettronico*”. Nel perseguire tale obiettivo, la direttiva InfoSoc interviene non unicamente nel garantire un livello di protezione elevato ai titolari dei diritti attraverso la protezione di misure tecnologiche (come i sistemi DRM che impediscono l'accesso non autorizzato o la copia di opere protette da *copyright*) ma anche rispetto alla definizione di un quadro di regole armonizzato che disciplini lo sfruttamento delle opere e altri materiali protetti tenendo conto anche di interessi altri e di pari livello.

Seguendo questa traiettoria, l'art. 5, par. 5, della direttiva InfoSoc⁴² introduce una serie di E&L definite sulla base delle tre condizioni stabilite dall'art. 9, par. 2, della Convenzione di Berna (1886) e note in maniera cumulativa come *Three-step Test*, secondo cui le limitazioni al diritto d'autore sono permesse: a) in casi particolari; b) quando non sono in conflitto con il normale sfruttamento dell'opera; e c) quando non pregiudicano irragionevolmente i legittimi interessi dell'autore.

Ai sensi dell'art. 5, par. 2, della direttiva InfoSoc, gli Stati membri dell'UE godono della discrezionalità di decidere se incorporare una E&L dall'elenco nelle proprie leggi. Tale libertà è tuttavia limitata e condizionata sotto due distinti e

⁴¹ Cfr. *supra*, nota 34.

⁴² Art. 5, par. 5, direttiva InfoSoc: “Le eccezioni e limitazioni di cui ai paragrafi 1, 2, 3 e 4 sono applicate esclusivamente in determinati casi speciali che non siano in contrasto con lo sfruttamento normale dell'opera o degli altri materiali e non arrechino ingiustificato pregiudizio agli interessi legittimi del titolare”.

correlati profili: i) le E&L devono essere coerenti con l'art. 5, par. 5, e dunque conformarsi alle condizioni stabilite dal Three-step test⁴³; ii) nel conformarsi al triplice test, le eccezioni devono essere interpretate restrittivamente⁴⁴. Nella sentenza Infopaq⁴⁵, la CGUE ha, infatti, stabilito che ogni eccezione al diritto d'autore, in quanto derogatorie rispetto alla regola generale per la quale le creazioni originali degli autori meritano adeguata protezione, deve conformarsi ad un criterio di stretta necessità, nel senso che gli usi non autorizzati sono ammissibili nella misura strettamente necessaria, adeguata e proporzionata per raggiungere lo scopo di interesse superiore (sentenza Painer)⁴⁶.

La direttiva Copyright, proseguendo lungo la medesima traiettoria: i) rende obbligatorio il sistema delle eccezioni e subordina la legittimità di ciascuna, inclusa quella di *Text and Data mining*, al rispetto delle condizioni di Berna, ribadendo all'art. 7, par. 2⁴⁷ e al relativo considerando 6 che le E&L, mirando al raggiungimento di un giusto equilibrio tra i diritti e gli interessi degli autori e degli altri titolari di diritti, da un lato, e degli utilizzatori, dall'altro, “[s]ono applicabili solo in taluni casi specifici che non siano in contrasto con il normale sfruttamento delle opere o altri materiali e non arrechino indebitamente pregiudizio ai legittimi interessi dei titolari dei diritti”; ii) rinnova il paradigma del “giusto equilibrio” come parametro per bilanciare diritti e interessi attraverso l'art. 17, par. 4⁴⁸, che stabilisce

⁴³ Così facendo, il legislatore europeo ha, tra l'altro, recepito nell'ordinamento europeo i trattati dell'Organizzazione Mondiale della Proprietà Intellettuale (WIPO) che, a sua volta, impongono ai Paesi aderenti il rispetto degli Artt. 1-21 della Convenzione di Berna e dunque anche del *Three-Step Test*. Inoltre, aderendo agli accordi TRIPs (Accordo sugli aspetti commerciali dei diritti di proprietà intellettuale) del 1995, l'UE ha riaffermato la sua aderenza al Three-Step Test, incorporato nell'art. 13, in forza del quale: “*I membri limiteranno le limitazioni o le eccezioni ai diritti esclusivi a determinati casi speciali che non siano in conflitto con il normale sfruttamento dell'opera e non pregiudichino irragionevolmente i legittimi interessi del titolare del diritto*”.

⁴⁴ Kapper, C-476/01, par. 72, e Commissione/Spagna, C-36/05, par. 31. In riferimento allo *standard* di stretta interpretazione delle E&L, si veda: Spiegel Online, C-516/17, EU:C:2019:625, par. 53; Funke Medien, C-469/17, EU:C:2019:623, par. 69; VCAST, C-265/16, EU:C:2017:913, par. 32; AKM, C-138/16, EU:C:2017:218, par. 37; Stichting Brein, C-527/15, EU:C:2017:300, par. 62; Vereniging Openbare Bibliotheken, C-174/15, EU:C: 2016:856, par. 50; Ranks e Vasiļevičs, C-166/15, EU:C:2016:762, par. 42; Public Relations Consultants Association, C-360/13, EU:C:2014:1195, par. 23; Deckmyn e Vrijheidsfonds, C-201/13, EU:C:2014:2132, par. 22; ACI Adam e altri, C-435/12, EU:C:2014:254, par. 22.

⁴⁵ Infopaq International, C-5/08, EU:C:2009:465, 58; ACI Adam e a., C-435/12, EU:C:2014:254, 26.

⁴⁶ Painer contro Standard Verlags GmbH e altri, C-145/10, par. 105 e 106.

⁴⁷ Art. 7, par. 2, direttiva Copyright: “L'articolo 5, paragrafo 5, della Direttiva 2001/29/CE si applica alle eccezioni e alle limitazioni di cui al presente titolo. L'art. 6, paragrafo 4, primo, terzo e quinto comma, della Direttiva 2001/29/CE si applica agli articoli da 3 a 6 della presente direttiva”.

⁴⁸ Art. 17, par. 4, direttiva Copyright: “Qualora non sia concessa alcuna autorizzazione, i prestatori di servizi di condivisione di contenuti online sono responsabili per atti non autorizzati di

la responsabilità dei prestatori di servizi di condivisione di contenuti *on-line* (“OCSSP”) per l’utilizzo di contenuti protetti dal diritto d’autore, prevedendo anche delle esclusioni. Secondo tale disposizione, gli OCSSP (ad esempio, piattaforme di *social media*, piattaforme di condivisione di video e altri servizi di condivisione di contenuti *on-line*) possono escludere la loro responsabilità per violazione del diritto d’autore solo se dimostrano di: a) aver compiuto massimi sforzi per ottenere un’autorizzazione; b) aver compiuto massimi sforzi, secondo elevati standard di diligenza professionale di settore, per assicurare che non siano disponibili opere e materiali specifici per i quali abbiano ricevuto le informazioni necessarie e pertinenti dai titolari dei diritti; e c) aver agito tempestivamente, dopo aver ricevuto una segnalazione sufficientemente motivata dai titolari dei diritti, per disabilitare l’accesso o rimuovere dai loro siti *web* i contenuti ed essersi adoperati al meglio per impedire il futuro caricamento di opere oggetto di segnalazione⁴⁹.

Per l’effetto, l’accesso e le estrazioni da opere o altri materiali⁵⁰, quand’anche

comunicazione al pubblico, compresa la messa a disposizione del pubblico, di opere e altri materiali protetti dal diritto d’autore, a meno che non dimostrino di:

- a) aver compiuto i massimi sforzi per ottenere un’autorizzazione;
- b) aver compiuto, secondo elevati standard di diligenza professionale di settore, i massimi sforzi per assicurare che non siano disponibili opere e altri materiali specifici per i quali abbiano ricevuto le informazioni pertinenti e necessarie dai titolari dei diritti; e in ogni caso;
- c) aver agito tempestivamente, dopo aver ricevuto una segnalazione sufficientemente motivata dai titolari dei diritti, per disabilitare l’accesso o rimuovere dai loro siti *web* le opere o altri materiali oggetto di segnalazione e aver compiuto i massimi sforzi per impedirne il caricamento in futuro conformemente alla lettera b)”.

⁴⁹ Le linee guida della Commissione del 2021 precisano che l’art. 17, par. 4, della direttiva Copyright introduce un regime specifico che consente agli OCSSP, a determinate condizioni, di evitare responsabilità derivanti dall’atto di comunicazione al pubblico ai sensi del par. 1 dello stesso art. 16. Come evidenziato nel considerando 66, tale regime specifico tiene conto del fatto che gli OCSSP forniscono accesso a materiali caricati non da loro stessi, bensì dagli utenti e che tali condizioni devono essere esplicitamente recepite nel diritto nazionale, poiché, la nozione di “adoperarsi al meglio non è definita e non vi è alcun riferimento al diritto nazionale”. La nozione di “massimi sforzi” costituisce pertanto “una nozione autonoma di diritto dell’UE che dovrebbe essere recepita dagli Stati membri” conformemente agli orientamenti forniti dalla Commissione e interpretata alla luce della finalità e degli obiettivi dell’art. 17. Le linee guida della Commissione specificano inoltre che: Gli obblighi derivanti dal paradigma dei “massimi sforzi” sono subordinati al rispetto delle libertà degli utenti, come previsto dall’art. 17, par. 7. In particolare, stabiliscono che le misure adottate dagli OCSSP, quali la rimozione dei contenuti oggetto di segnalazione, non devono comportare una violazione della libertà di espressione e del diritto di accesso alle informazioni. Ciò implica evitare forme di censura indebita o restrizioni eccessive che possano limitare l’accesso a contenuti legittimi (c.d. “*overblocking*”), garantendo un equilibrio tra la tutela dei diritti d’autore e i diritti fondamentali degli utenti. La proporzionalità (art. 17, par. 5) è centrale nella valutazione del rispetto degli obblighi, specialmente per distinguere tra grandi piattaforme e operatori più piccoli.

⁵⁰ Art. 4 direttiva Copyright: “1. Gli Stati membri dispongono un’eccezione o una limitazione ai diritti di cui all’articolo 5, lettera a), e all’articolo 7, paragrafo 1, della direttiva 96/9/CE, all’articolo 2 della direttiva 2001/29/CE, all’articolo 4, paragrafo 1, lettere a) e b), della direttiva 2009/24/CE e

effettuate per attività commerciali (tenuto conto che il considerando 18 chiarisce che, oltre alla rilevanza in ambito scientifico, le tecniche di TDM “*sono ampiamente utilizzate da soggetti sia pubblici che privati per analizzare grandi quantità di dati in diversi settori della vita e per vari scopi, tra cui i servizi della pubblica amministrazione, decisioni commerciali complesse e lo sviluppo di nuove applicazioni o tecnologie*”) sono legittime nella misura in cui rispettino condizioni specifiche, disposte dalla direttiva, e generali, disposte dal *test* tripartito.

Quanto alle condizioni generali: A) Al fine di bilanciare i diritti esclusivi e l’innovazione digitale in quanto interesse pubblico tutelato dal diritto dell’UE⁵¹, l’art. 4, par. 3, della direttiva Copyright impone agli Stati membri di disporre di E&L al diritto d’autore per le riproduzioni e le estrazioni effettuate da opere o altri materiali cui si abbia legalmente accesso ai fini dell’attività di TDM, a condizione che l’utilizzo delle opere e di altri materiali non sia stato espressamente riservato dai titolari dei diritti in modo appropriato (“*opt-out*”), ad esempio attraverso strumenti che consentano la lettura automatizzata di contenuti resi pubblicamente disponibili *online*, rispetto ai quali il titolare abbia deciso di rinunciare alle facoltà escludenti. B) rispetto alla ricerca, il considerando 14 della direttiva Copyright specifica che la nozione di “legittimo accesso” dovrebbe essere intesa come 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. C) rispetto alle attività commerciali, il considerando 18 precisa che l’eccezione deve applicarsi: i) “*solo in caso di accesso legale all’opera o altri materiali da parte del beneficiario, anche quando l’opera o altri materiali sono stati messi a disposizione del pubblico online, e ii) nella misura in cui i titolari dei diritti non abbiano riservato in maniera appropriata i diritti di effettuare riproduzioni ed estrazioni ai fini dell’estrazione di testo e di dati. Nel caso dei contenuti resi disponibili al pubblico online, dovrebbe essere ritenuto appropriato riservare tali diritti solo attraverso l’uso di strumenti che consentano una lettura automatizzata, inclusi i metadati e i termini e le condizioni di un sito web o di*

all’articolo 15, paragrafo 1, della presente direttiva per le riproduzioni e le estrazioni effettuate da opere o altri materiali cui si abbia legalmente accesso ai fini dell’estrazione di testo e di dati.

2. Le riproduzioni e le estrazioni effettuate a norma del paragrafo 1 possono essere conservate per il tempo necessario ai fini dell’estrazione di testo e di dati.

3. L’eccezione o la limitazione di cui al paragrafo 1 si applica a condizione che l’utilizzo delle opere e di altri materiali di cui a tale paragrafo non sia stato espressamente riservato dai titolari dei diritti in modo appropriato, ad esempio attraverso strumenti che consentano lettura automatizzata in caso di contenuti resi pubblicamente disponibili *online*.

4. Il presente articolo non pregiudica l’applicazione dell’articolo 3 della presente direttiva”.

⁵¹ L’art. 3 TUE stabilisce dell’UE mira a promuovere lo sviluppo sostenibile, la crescita economica e il progresso tecnologico. Questi obiettivi si basano intrinsecamente sulla promozione dell’innovazione. Inoltre, l’art. 179 del TFUE affronta specificamente il ruolo dell’UE nel rafforzamento della sua base scientifica e tecnologica e nella realizzazione di un mercato interno che favorisca la ricerca, lo sviluppo tecnologico e l’innovazione.

un servizio”; iii) consentendo agli utilizzatori di conservare le copie realizzate per il tempo necessario ai fini dell’attività.

Come ampiamente riconosciuto in dottrina, “*lo schema è chiaro nella sua apparente semplicità. Il TDM può essere effettuato per qualsiasi scopo, commerciale o non commerciale (ad esempio, giornalismo investigativo, software, ecc.), purché si limiti ad atti di riproduzione (e adattamento nel caso di software)*”⁵². Tuttavia, i titolari dei diritti hanno la possibilità di “disinnescare” l’operatività dell’eccezione riservando espressamente l’uso di tali opere in modo appropriato⁵³, ad esempio con mezzi leggibili a macchina nel caso di contenuti resi pubblicamente disponibili *on-line* (art. 4, par. 3, direttiva Copyright).

Sull’eccezione di TDM si innesta l’AI Act, che all’art. 53, disciplinando gli obblighi dei fornitori di sistemi di IA generativa, impone agli stessi di attuare “*una politica volta 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 [Copyright]*” (par. 1, lett. c))⁵⁴.

Premesso che tanto l’“immagazzinamento” di dati, quanto lo sviluppo e l’implementazione dei sistemi di IA a valle⁵⁵ può incidere le prerogative autoriali, l’art. 53, par. 1, rappresenta, in sintesi, “il modo per rendere operativa, almeno all’interno della sottocategoria dell’IA generativa, la possibilità per i titolari dei diritti di monetizzare l’uso delle loro opere una volta che abbiano optato per l’esclusione dall’addestramento (o presumibilmente negato l’accesso alle stesse)”⁵⁶.

Il considerando 105 specifica a questo proposito 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*”. Ciò implica che, qualora i fornitori di sistemi di IA generativa intendano utilizzare il TDM, questi debbano verificare se i titolari dei diritti abbiano riservato l’uso delle loro opere per il TDM utilizzando “*mezzi leggibili a macchina, compresi i metadati e i termini e le condizioni di un sito web o di un servizio*” o attraverso “*accordi contrattuali o una dichiarazione unilaterale*” (art. 4, par. 3 e considerando 18 della direttiva Copyright) ed in ogni caso che ricorrano le condizioni del test tripartito.

Si aggiunga che l’obbligo di trasparenza di cui all’art. 53, par. 1, lett. d), dell’AI

⁵² T. MARGONI, *TDM and Generative AI: Lawful Access and Opt-Outs*, 2024, disponibile su https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5036164, ultimo accesso 15 dicembre 2024.

⁵³ *Ibidem*.

⁵⁴ A. PEUKERT, *Copyright in the Artificial Intelligence Act*, cit., *passim*.

⁵⁵ *Ibidem*.

⁵⁶ T. MARGONI, *TDM and Generative AI*, cit.

Act⁵⁷ consente ai titolari dei diritti di determinare se le loro opere sono state utilizzate o meno negli insiemi di dati per l'addestramento e, se necessario, escluderle, grazie all'obbligo per i fornitori di redigere e mettere a disposizione del pubblico una sintesi sufficientemente dettagliata dei contenuti utilizzati per l'addestramento del modello di IA per finalità generali, secondo un modello fornito dall'AI Office.

4. La terza bozza del Codice di condotta generale per l'IA⁵⁸

L'analisi del quadro normativo deve, infine, considerare la terza bozza del Codice di condotta generale per l'IA (pubblicata l'11 marzo scorso), il cui obiettivo principale è stabilire principi etici e linee guida vincolanti per assicurare che lo sviluppo e l'utilizzo dei modelli di intelligenza artificiale avvengano in conformità con i principi di sicurezza, responsabilità e rispetto dei diritti fondamentali.

A tal proposito, l'art. 53, par. 4, dell'AI Act, fa esplicito riferimento al Codice che può essere utilizzato dai fornitori di sistemi di IA generativa *“per dimostrare la conformità agli obblighi relativi al diritto d'autore, fino alla pubblicazione di uno standard armonizzato”*.

A sua volta, il Codice richiama l'AI Act ed in particolare l'art. 53, par. 1, lett. c), che impone ai fornitori di sistemi di IA generici immessi sul mercato dell'Unione l'obbligo di implementare una politica di conformità al diritto d'autore dell'Unione e ai diritti connessi. In particolare, i fornitori devono: i) rilevare ed eventualmente rispettare, attraverso tecnologie avanzate, una riserva di diritti espressa ai sensi dell'art. 4, par. 3, della direttiva (UE) 2019/790; ii) garantire tale conformità a tale disposizione indipendentemente dalla giurisdizione in cui si svolgono gli atti rilevanti per il diritto d'autore alla base della formazione di tali sistemi di IA generici.

La bozza del Codice contiene disposizioni specifiche sui sistemi di IA per finalità generali, sottolineando che la sezione dedicata al copyright (*“Commitment I.2. Copyright policy”*) non pregiudica *“l'applicazione e l'attuazione della normativa dell'Unione in materia di diritto d'autore e diritti connessi”*, né *“gli accordi commerciali tra i firmatari e i titolari dei diritti che autorizzano l'uso di opere e altri materiali protetti”*. In particolare, il Codice evidenzia la necessità di interpretare le sue disposizioni nel rispetto delle norme dell'Unione sul diritto d'autore, con particolare riferimento alla riserva di diritti espressa ai sensi dell'articolo 4, paragrafo

⁵⁷ Art. 53, par. 1, lett. d), AI Act: *“1. I fornitori di modelli di IA per finalità generali: d) redigono e mettono a disposizione del pubblico una sintesi sufficientemente dettagliata dei contenuti utilizzati per l'addestramento del modello di IA per finalità generali, secondo un modello fornito dall'ufficio per l'IA”*.

⁵⁸ I contenuti del paragrafo sono stati elaborati in collaborazione con la Prof. Federica Pascucci nell'ambito del Progetto PRIN E-Agorà.

3, della Direttiva (UE) 2019/790.

In tale contesto, riveste particolare importanza il processo di acquisizione e formazione dei dataset, finalizzato all'addestramento dei Large Language Models (LLM), che costituiscono la base dell'intelligenza artificiale generativa, attraverso il processo automatizzato di *web-scraping*⁵⁹. La terza bozza del Codice (Misura I.2.3., paragrafo 2) specifica, a questo proposito, che l'impegno dei firmatari a "identificare e rispettare, anche attraverso tecnologie all'avanguardia, le riserve di diritti leggibili a macchina espresse ai sensi dell'articolo 4, paragrafo 3, della direttiva (UE) 2019/790 se utilizzano web-crawler o fanno utilizzare tali web-crawler per loro conto per effettuare il crawling⁶⁰, il *web-scraping*⁶¹ e/o la compilazione di

59 La navigazione effettuata ai fini di scraping si svolge attraverso modalità sostanzialmente analoghe a quelle impiegate da un utente umano nella fruizione ordinaria della rete. Gli algoritmi preposti allo scraping possono, infatti, accedere direttamente a un sito web digitandone il nome a dominio, oppure procedere alla consultazione sequenziale delle pagine interne al medesimo sito, ovvero di siti differenti, mediante l'utilizzo di collegamenti ipertestuali (link). In entrambi i casi, l'algoritmo fa uso di specifiche stringhe alfanumeriche denominate URL (Uniform Resource Locator), le quali sono strutturate secondo una sintassi codificata. Al fine di permettere l'accesso effettivo alle risorse web, tali URL devono essere tradotti in indirizzi numerici compatibili con il protocollo di rete IP (Internet Protocol) mediante appositi protocolli di risoluzione denominati DNS (Domain Name System). Una volta ottenuto l'indirizzo IP corrispondente, è possibile stabilire una connessione con il server web ospitante il sito richiesto e ricevere da quest'ultimo le pagine richieste, generalmente espresse in linguaggio HTML (HyperText Markup Language), idoneo alla rappresentazione ipertestuale dei contenuti. Una volta stabilita la connessione e ottenute le risorse web richieste, l'algoritmo di scraping dà avvio all'attività di individuazione, estrazione e raccolta dei dati. Tale operazione può essere eseguita direttamente sul codice sorgente delle pagine, generalmente strutturato in linguaggio HTML, oppure sulla rappresentazione visuale generata a schermo dai browser web, i quali interpretano e combinano elementi HTML, immagini, video, documenti in formato PDF e altri tipi di contenuti multimediali o testuali. Nel primo caso, in cui l'interazione avviene a livello di codice, si fa riferimento alle tecniche note come web scraping o web crawling. Nel secondo caso, invece, in cui l'estrazione avviene a partire dall'interfaccia grafica risultante, si parla di screen scraping.

60 La distinzione tra scraping e crawling si fonda sulla tipologia di dati oggetto di raccolta nonché sulle finalità e modalità di aggregazione degli stessi. In entrambi i casi, i dati vengono estratti direttamente dal codice sorgente delle pagine web, prevalentemente redatto in linguaggio HTML; tuttavia, le rispettive tecniche perseguono obiettivi differenti. Nel caso del web-crawling, l'attività si concentra prevalentemente sull'acquisizione di parole chiave (keywords) e di collegamenti ipertestuali (hyperlinks) verso altre pagine. Le parole chiave possono essere identificate in base a specifici tag inseriti dagli sviluppatori all'interno del codice HTML (ad esempio, meta tag), oppure determinate attraverso l'analisi lessicale del contenuto, con particolare riferimento alla frequenza e ricorrenza di determinati termini o concetti. Tale forma di raccolta dati è funzionale alla successiva indicizzazione dei contenuti, così da consentire l'efficace reperimento delle pagine web tramite motori di ricerca.

61 Nel caso del web-scraping, l'attività non si limita alla raccolta di parole chiave o collegamenti ipertestuali, ma si estende all'acquisizione dell'intero contenuto informativo delle pagine web. Tali contenuti vengono estratti al fine di ricavare informazioni, intese come dati elaborati e strutturati, suscettibili di impieghi ulteriori. Tra le applicazioni più rilevanti si annovera la generazione di dataset finalizzati all'addestramento di modelli di intelligenza artificiale. Le informazioni così ottenute vengono successivamente organizzate e conservate all'interno di basi di dati dedicate, per consentirne l'utilizzo in ambiti specifici.

dati in altro modo ai fini del TDM ai sensi dell'articolo 2, paragrafo 2, della direttiva (UE) 2019/790 e dell'addestramento dei loro modelli di IA per scopi generali” non pregiudica “il diritto dei titolari dei diritti di riservare espressamente l'uso di opere legittimamente accessibili e di altri materiali protetti ai fini dell'estrazione di testi e dati ai sensi dell'articolo 4, paragrafo 3, della direttiva (UE) 2019/790 in qualsiasi modo appropriato, ad esempio con mezzi leggibili a macchina nel caso di contenuti resi pubblicamente disponibili online”.

Tra le ulteriori misure predisposte al fine di far rispettare i limiti dell'eccezione TDM, il par. 1, lett. a, della misura I.2.3. della terza bozza del Codice⁶² prevede l'obbligo per i firmatari di utilizzare esclusivamente *web-crawler* “che leggano e seguano le istruzioni espresse in conformità al Protocollo di Esclusione dei Robot (*robots.txt*)”⁶³. La misura I.2.3, par. 1, lett. b, espande ulteriormente la tutela delle riserve di diritti in quanto, oltre a far impegnare i firmatari a “fare il massimo degli sforzi per identificare e rispettare altri protocolli leggibili dalle macchine per esprimere riserve sui diritti, ai sensi dell'articolo 4(3) della Direttiva (UE) 2019/790”, fornisce esempi di espressioni di protocolli leggibili dalle macchine per esprimere riserve sui diritti, inclusi “metadati basati su asset o sulla posizione, che derivano da un processo di standardizzazione intersettoriale”, come previsto nel paragrafo 3

62 Misura I.2.3., par. 1, lett. a: “Al fine di garantire che i firmatari identifichino e rispettino, anche attraverso tecnologie all'avanguardia, le riserve di diritti leggibili meccanicamente espresse ai sensi dell'articolo 4, paragrafo 3, della direttiva (UE) 2019/790, qualora utilizzino *web-crawler* o facciano utilizzare tali *web-crawler* per loro conto per effettuare crawling, scrape e/o compilare in altro modo dati ai fini dell'estrazione di testo e di dati ai sensi dell'articolo 2, paragrafo 2, della direttiva (UE) 2019/790 e dell'addestramento dei loro modelli di intelligenza artificiale generica, i firmatari impiegheranno *web-crawler* che leggano e seguano le istruzioni espresse in conformità al Protocollo di Esclusione dei Robot (*robots.txt*), come specificato nella richiesta di commenti n. 9309 dell'Internet Engineering Task Force (IETF) e in qualsiasi versione successiva di questo standard IETF”.

63 In considerazione dell'elevata accessibilità delle risorse web e della relativa semplicità con cui gli algoritmi di scraping possono interagire con i contenuti online, risulta particolarmente complesso anticipare e prevenire in modo esaustivo le molteplici modalità attraverso cui tali pratiche possono essere implementate. A tal fine, nel corso del tempo sono stati sviluppati alcuni strumenti tecnici e protocolli volti a disciplinare, sebbene su base volontaria e non vincolante, l'attività di accesso automatizzato ai siti web. Tra questi si annoverano i meccanismi riconducibili al cosiddetto Protocollo di Esclusione dei Robot, un insieme di specifiche che consentono agli amministratori di un sito web di indicare agli user-agent automatizzati quali aree del sito debbano essere escluse dalla scansione o raccolta. I Protocolli di Esclusione dei Robot maggiormente usati sono (1) File *robots.txt*; (2) Tag HTML meta *robots*; (3) Intestazione HTTP X-Robots-Tag. I Protocolli di Esclusione dei Robot contribuiscono a delineare il perimetro di legittima operatività degli algoritmi di scraping, attraverso l'indicazione esplicita di quali risorse, all'interno di un sito web, possano essere oggetto di scansione e indicizzazione, e quali, invece, debbano essere escluse da tali attività. Tali indicazioni vengono solitamente fornite mediante l'implementazione di file di configurazione o tramite tag HTML che stabiliscono istruzioni rivolte agli user-agent automatizzati. Pur non avendo natura vincolante in senso giuridico, questi strumenti costituiscono una prassi tecnica riconosciuta e, in determinati contesti, possono essere valorizzati quale espressione della volontà del titolare del sito in merito all'accesso e all'utilizzo dei propri contenuti digitali.

di questa misura⁶⁴, o “sono all’avanguardia e ampiamente adottati dai titolari dei diritti, considerando i diversi settori culturali, e generalmente concordati attraverso un processo inclusivo basato su discussioni di buona fede da facilitare a livello UE con il coinvolgimento dei titolari dei diritti, fornitori di IA e altre parti interessate come una soluzione più immediata, anticipando lo sviluppo degli standard intersettoriali di cui al paragrafo 3”.

Relativamente alla Misura I.2.3., appare opportuno precisare che il *web-crawling* consiste in una fase del *web-scraping* che include 1) analisi del sito web, 2) *web-crawling* e 3) organizzazione dei dati⁶⁵. Come noto, in dottrina, dato che “la legge sul copyright non impedisce di raccogliere i dati in sé, né un sito *web* possiede necessariamente i dati presenti sul proprio sito, in particolare quando sono generati dagli utenti (ad esempio, contenuti provenienti da siti di *social media*)”⁶⁶, il *web-scraping* può comportare violazioni del diritto d’autore. Per contenere tali rischi, sono stati sviluppati degli strumenti che regolano l’accesso dei crawler/robot alle pagine web, primo fra tutti il *robots.txt*, file di configurazione che informa il *web-scrafer* circa le restrizioni di accesso al sito *web*, con un impatto significativo sulla tutela del diritto d’autore⁶⁷. Più precisamente, la presenza o assenza di tali file può essere usata per determinare se un proprietario di un sito *web* consenta o meno l’accesso ai *web-crawler* (ad esempio nel caso *Field v. Google*⁶⁸).

È fondamentale sottolineare che i Protocolli di Esclusione dei Robot presentano delle limitazioni sotto il profilo tecnologico, in particolare per quanto concerne la granularità del controllo, che può risultare circoscritta alle singole pagine web direttamente accessibili dallo scraper, senza offrire un controllo completo e dettagliato su tutte le risorse protette o non protette da diritto d’autore all’interno della stessa. Un limite significativo dello strumento *robots.txt* è dato proprio dall’accesso diretto ad un URL.

Nel caso in cui una sottosezione ad accesso limitato di un sito (“sito A”) venga raggiunta tramite un *link* diretto da un altro sito (“sito B”), il *bot* non sarà in grado

64 Misura I.2.3., par. 3: “3) Tenendo in debita considerazione i processi di standardizzazione internazionali ed europei pertinenti, i firmatari sono incoraggiati a sostenere gli sforzi di standardizzazione pertinenti e a impegnarsi su base volontaria in discussioni in buona fede con altre parti interessate, compresi i titolari dei diritti, al fine di sviluppare standard leggibili a macchina appropriati per esprimere una riserva di diritti ai sensi dell’articolo 4, paragrafo 3, della direttiva (UE) 2019/790”.

65 V. Krotov-L. Silva, *Legality and Ethics of Web Scraping*, 2018.

66 V. Krotov-L. Johnson-L. Silva, *Tutorial: Legality and Ethics of Web Scraping*, 2020, *Communications of the Association for Information Systems*, Volume 47, art. 22.

67 Collocato sulla directory iniziale di un sito web, il *robot.txt* indica quali sezioni del sito sono visitabili e quali no, limitando l’accesso a file, singole pagine e intere directory. Il file *robots.txt* è opzionale: qualora il bot non trovi tale file, assume implicitamente che può esplorare e recuperare dati da tutte le pagine che costituiscono un sito.

68 *Field v. Google Inc.*, caso 2:2004cv00413 (2006), Corte Distrettuale degli Stati Uniti per il Distretto del Nevada.

di recuperare il file robots.txt⁶⁹ del sito A, riuscendo dunque ad accedervi⁷⁰. Per ovviare alle suddette limitazioni, sono attualmente in fase di sviluppo ulteriori strumenti, sebbene essi presentino significative difficoltà di armonizzazione e, di conseguenza, non siano ancora stati formalmente standardizzati nei protocolli ufficiali di Internet⁷¹. A questo proposito, pur riconoscendo che lo strumento robots.txt rappresenta una difesa iniziale contro il *web-scraping*, il Codice di condotta generale per l'IA, introduce l'ulteriore misura I.2.3, par. 1, lett. b, evidenziando così che robot.txt necessita ancora di perfezionamenti per tutelare efficacemente il diritto d'autore.

Proprio a tal fine, tale disposizione offre un ulteriore livello di protezione, richiamando esplicitamente l'art. 17, par. 4, della direttiva Copyright che, come visto sopra, integra nella normativa comunitaria il paradigma dei “massimi sforzi”, stabilendo gli standard di comportamento delle piattaforme al fine di prevenire le violazioni del diritto d'autore e garantire un bilanciamento tra gli interessi dei titolari dei diritti e quelli degli utenti. In particolare, i firmatari sono incoraggiati a implementare strumenti conformi con gli standard industriali ampiamente utilizzati che consentano di “identificare e rispettare [...] altri mezzi leggibili a macchina per esprimere in modo appropriato una riserva di diritti a livello di fonte o di opera ai sensi dell'articolo 4, paragrafo 3, della direttiva (UE) 2019/790 nel caso di contenuti resi pubblicamente disponibili *online*”.

69 Il file robots.txt è un file di testo introdotto nel 2002 dalla Internet Engineering Task Force (IETF) nell'ambito degli standard del Robot Exclusion Protocol (REP). Tale file può essere richiesto da un algoritmo di scraping al momento dell'accesso a un sito web, a partire dalla sua pagina principale (ossia la directory radice). All'interno di questo file vengono fornite istruzioni destinate agli user-agent automatizzati, tra cui l'indicazione delle pagine o delle directory che possono essere esplorate, nonché eventuali vincoli tecnici, quali i tempi di attesa (crawl-delay) da rispettare tra una richiesta e l'altra. Il file robots.txt ha carattere meramente facoltativo: in sua assenza, l'algoritmo può presumere, secondo la prassi tecnica consolidata, che l'intero contenuto del sito sia liberamente accessibile e indicizzabile, salvo l'adozione di ulteriori misure di protezione da parte del titolare del sito. In linea generale, il file robots.txt stabilisce le direttive di accesso per gli algoritmi di scraping, specificando le pagine del sito che possono essere visitate a partire dalla directory radice. Tuttavia, qualora lo scraper acceda al sito attraverso una URL specifica, è necessario ricorrere a strumenti aggiuntivi per regolamentare l'accesso alle singole pagine. Tra questi strumenti figurano i meta tag HTML, in particolare il tag robots, e le intestazioni HTTP X-Robots-Tag, che consentono di definire a livello di pagina specifico le restrizioni di accesso.

70 La sentenza sul caso C-466/12 (noto come caso Svensson) della Corte di Giustizia ha stabilito che è consentito condividere link a siti web con contenuti gratuiti senza la previa autorizzazione del titolare dei diritti, rendendo di fatto difficile limitare la diffusione dei contenuti tramite linking. Tuttavia, questa possibilità è esclusa se i contenuti sono accessibili solo tramite paywall o previa registrazione al sito web.

71 Tali strumenti sono comunemente indicati come Rights Expression Language (REL). I REL possono essere espresse tramite linguaggi formali di rappresentazione dati, quali XML, RDF, RDF Schema e JSON, e possono essere integrate direttamente nei siti web o incorporati come metadati in vari tipi di documenti, inclusi eBook, immagini, file audio e video.

L'associazione di queste due misure (misura I.2.3, par. 1, lett. a e b) riconosce che il protocollo robots.txt rappresenta una prima linea di difesa contro il *web-scraping*, ma di per sé non sia sufficiente a garantire una tutela efficace del diritto d'autore. In tal senso, l'obbligo per i provider di compiere "i migliori sforzi proporzionati alle loro dimensioni e capacità, in conformità con gli standard industriali ampiamente diffusi, per identificare e rispettare, anche attraverso tecnologie all'avanguardia, altri mezzi leggibili a macchina per esprimere in modo appropriato una riserva di diritti a livello di fonte o di opera ai sensi dell'articolo 4, paragrafo 3, della direttiva (UE) 2019/790 nel caso di contenuti resi pubblicamente disponibili online" rappresenta un'integrazione necessaria per la tutela del diritto d'autore e dei diritti connessi nell'ambito del mercato interno, a fronte di soluzioni tecnologiche che ancora in via di sviluppo.

5. Conclusioni

Mentre **l'innovazione** galoppa al di qua e al di là dell'oceano, l'Europa, anche attraverso il Codice di condotta in fase di avanzata elaborazione, **la addomestica per salvaguardare il diritto di ciascuno all'innovazione e per questa via sancisce un obbligo di fare, cioè introdurre misure attive** (tecniche, di progettazione e di sviluppo dei sistemi), che assicurino il rispetto della disciplina autoriale **sia a monte (in fase di addestramento) sia a valle**, quando dei contenuti sintetici vengono rilasciati.

Nell'attesa che venga "messo a terra" il nuovo quadro abilitante, consentendo la circolazione dei dati attraverso misure tecniche ed organizzative adeguate che, nel facilitare la condivisione, garantiscano la protezione dei diritti, la giurisprudenza europea⁷² e quella statunitense⁷³ confermano anche da ultimo un **punto fermo per il diritto internazionale, europeo e nazionale**. Vale a dire che per valutare la compatibilità di una condotta anche di addestramento di un modello o di un sistema di IA occorre dimostrare che lo stesso non abbia violato le regole in materia di text and data mining⁷⁴.

72 Corte regionale di Amburgo, sentenza 310 O 227/23 (2024).

73 *New York Times vs OpenAI*, Southern District of New York, n. civ. 1:23-cv-11195

74 Il caso affronta la questione dell'uso di dati protetti da diritto d'autore per l'addestramento di sistemi di IA, esaminando la compatibilità con le eccezioni di TDM previste dalla normativa europea e stabilendo che l'utilizzo di opere protette per il training di sistemi di IA può rientrare nelle eccezioni previste dall'art. 3 della direttiva Copyright e dall'art. 60d della legge tedesca sul diritto d'autore (TDM per scopi di ricerca scientifica). Nel caso specifico, il ricorrente è un fotografo che ha reso una delle sue foto liberamente disponibili al pubblico tramite il sito web di Bigstockphoto, subordinatamente ai seguenti termini di servizio: "L'utente non può utilizzare programmi automatici, applet, bot o simili per accedere al sito Web di Bigstock.com o a qualsiasi contenuto in esso presente per qualsiasi scopo, inclusi, a titolo puramente esemplificativo, il download di contenuti, l'indicizzazione, lo

L'Europa, insomma, attraverso le linee di indirizzo sin qui declinate, nel favorire la condivisione e lo scambio dei dati, non rinuncia al proprio modello di bilanciamento di interessi, chiarendo nelle azioni della Strategia europea che **il prezzo della corsa all'innovazione non può consistere nell'espropriazione dei diritti né nell'appropriazione indebita degli investimenti sottesi alla creazione dei dati che costituiscono il *quid proprium* dell'attività imprenditoriale.**

Il Regolamento sull'IA ne è la più recente conferma, ribadendo che **per rafforzare la competitività europea nel digitale non si tratta di privare le imprese del proprio *know how*, di sottrarre arbitrariamente e senza compenso dati e contenuti a favore di nuove tecnologie abilitanti, né tantomeno di consentire l'addestramento gratuito dei nuovi modelli.**

Si tratta, piuttosto, di identificare e riconoscere prima e massimizzare poi il valore dei dati, estraendolo ogni volta e per ogni uso alle migliori condizioni possibili nell'interesse strategico dell'impresa e a beneficio della collettività.

scraping o la memorizzazione nella cache di qualsiasi contenuto del sito Web". La convenuta, un'organizzazione senza scopo di lucro dedicata alla promozione di attività di ricerca in materia di IA, crea e fornisce set di dati aperti costituiti da coppie testo-immagine pronte per l'uso per addestrare l'IA generativa. Tramite l'abbinamento di un gran numero di testi e immagini, l'IA è in grado di apprendere l'aspetto di persone, animali o oggetti, in base al quale l'IA può essere autorizzata a distinguere e creare artificialmente persone, animali o oggetti da sola. Ai fini delle attività di formazione IA, ha scaricato e archiviato una copia di una serie di immagini da risorse disponibili al pubblico. In caso di mancata corrispondenza, il software elimina la particolare immagine dal set di dati poiché quest'ultimo deve essere affidabile e coerente per garantire una corretta formazione AI. Il set di dati dell'imputato comprende quasi sei miliardi di coppie testo-immagine, tra cui una foto creata dal querelante. Robert Kneschke ha quindi chiesto a LAION di interrompere qualsiasi attività di riproduzione della foto in questione. La Corte ha ritenuto che il tipo di TDM effettuato da LAION possa essere considerato TDM ai sensi dell'art. 60d della legge tedesca sul diritto d'autore che si applica anche quando i diritti sono espressamente riservati. L'analisi di LAION del file di immagine per abbinarlo ad una descrizione preesistente dell'immagine equivarrebbe ad un'analisi finalizzata a ottenere informazioni sulle "correlazioni" (vale a dire, la questione della non concordanza tra immagini e descrizioni di immagini) che, in quanto resa disponibile gratuitamente, non perseguirebbe alcun fine commerciale. Se è vero che l'insieme di dati così creato può essere successivamente utilizzato per addestrare reti neurali artificiali e che il contenuto generato dall'IA può entrare in concorrenza con le opere di autori (umani), ciò non giustifica che la creazione di insiemi di dati sia considerata una violazione dei diritti di sfruttamento delle opere ai sensi dell'art. 5(5) della direttiva InfoSoc. L'applicazione della limitazione TDM sarebbe inoltre compatibile con il Three-step test in quanto non è evidente, né è stato affermato dai ricorrenti, che le possibilità di sfruttamento delle opere in questione da parte di Robert Kneschke sarebbero state pregiudicate dalla creazione di dataset da parte di LAION. Questo appare in linea con la giurisprudenza consolidata della Corte di Giustizia, secondo la quale nel valutare se una disposizione nazionale sul TDM sia applicabile in un determinato caso, un giudice non dovrebbe esclusivamente soffermarsi sulle condizioni specifiche per la sua applicazione ma dovrebbe inoltre considerare se i requisiti del Three-Step Test siano rispettati.

