

Life-on-life: l'umanesimo
di fronte alle sfide
dell'intelligenza artificiale

a cura di Mario Cosenza e Luca Lo Sapia

Federico II University Press



fedOA Press

Life-on-life: l'umanesimo di fronte alle sfide dell'intelligenza artificiale / a cura di Mario Cosenza e Luca Lo Sapiro. – Napoli : FedOAPress, 2025. – XVI+267 p. ; 21 cm. – (Human & Future ; 1).

Accesso alla versione elettronica: <http://www.fedoabooks.unina.it>

ISBN: 978-88-6887-355-4

DOI: 10.6093/978-88-6887-355-4

Volume pubblicato con il contributo dell'Università degli Studi di Napoli Federico II, Fondi Dipartimentali del Dipartimento di Studi Umanistici (2024)

© 2025 FedOAPress – Federico II University Press

Università degli Studi di Napoli Federico II
Centro di Ateneo per le Biblioteche “Roberto Pettorino”
Piazza Bellini 59-60
80138 Napoli, Italy
<http://www.fedoapress.unina.it/>
Published in Italy
Prima edizione: luglio 2025

Gli E-Book di FedOAPress sono pubblicati con licenza Creative Commons Attribution 4.0 International

Indice

Mario Cosenza, Luca Lo Sapia, <i>Introduzione</i>	VII
<i>Ringraziamenti</i>	XV
Mario Cosenza, <i>Esistenze artificiali. Strutture, politica e filosofia nel digitale</i>	1
Roberto Delle Donne, <i>L'editoria accademica, i dati della ricerca e l'intelligenza artificiale</i>	31
Fabiana Gambardella, <i>Words matter (?): su dilemmi e mistificazioni del Secolo post-nichilista</i>	55
Gianluca Giannini, <i>L'insostenibile leggerezza dell'InfoSapiens (del Digitocene e dintorni)</i>	77
Emiliano Grimaldi, <i>L'intelligenza Artificiale e l'Educazione. Appunti per una sociologia pubblica</i>	101
Luca Lo Sapia, <i>Creatività umana e creatività computazionale: una visione pluralistica dell'atto creativo</i>	123
Simon Pietro Romano, <i>Guerra cibernetica e frontiere etiche. L'evoluzione dei conflitti nell'arena digitale delle sfide globali</i>	157
Francesco Romeo, <i>Intelligenza Artificiale, povertà dell'etica, inutilità del giurista</i>	197
Steven Umbrello, <i>L'intersezione tra il realismo critico di Bernard Lonergan, il bene comune e l'intelligenza artificiale nelle pratiche religiose moderne</i>	219

ROBERTO DELLE DONNE

*L'editoria accademica, i dati della ricerca
e l'intelligenza artificiale*

Abstract: The article examines the transformative impact of artificial intelligence (AI) and research data management on academic publishing, emphasizing global trends and the conservative stance of Italian publishers. It highlights initiatives such as Open Science, FAIR principles (Findable, Accessible, Interoperable, Reusable), and Text and Data Mining (TDM), discussing their challenges and potential for innovation. The role of Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) in automating academic workflows, addressing ethical concerns, and promoting inclusivity is explored. The article concludes by advocating for equitable access to AI tools and data, aligning technology with the principles of cultural heritage preservation and knowledge democratization.

Keywords: Academic Publishing, Artificial Intelligence, Open Science, FAIR Data, Text and Data Mining, Large Language Models, Cultural Heritage, Knowledge Democratization.

Introduzione

L'editoria accademica sta attraversando una profonda trasformazione, guidata dall'integrazione di tecnologie avanzate come l'intelligenza artificiale (IA) e da un'attenzione crescente verso la gestione dei dati della ricerca. Tuttavia, questa evoluzione non è uniforme a livello globale. In Europa e su scala internazionale prevalgono approcci innovativi e collaborativi, mentre in Italia, dove predominano piccole e medie imprese spesso con risorse limitate per investire in nuove tecnologie, si registrano orientamenti più conservativi, anche

se non mancano eccezioni¹. Ne è un esempio la lettera inviata di recente dall'Associazione Italiana Editori (AIE) ai rettori delle università italiane², in cui emerge un approccio prevalentemente difensivo rispetto ai cambiamenti indotti dalla digitalizzazione e dal rapido affermarsi dell'intelligenza artificiale. Questo orientamento si discosta dalle tendenze europee e internazionali, che adottano un approccio proattivo orientato a promuovere modelli di Open Science, a garantire la sostenibilità tramite l'accesso aperto ai dati e a integrare tecnologie digitali avanzate per ottimizzare e automatizzare i processi editoriali e di gestione dei dati, favorendo così una condivisione più equa e inclusiva della conoscenza scientifica. Un aspetto cruciale di queste tendenze è il *Text and Data Mining* (TDM)³, una tecnica utilizzata da tempo per alimentare motori di ricerca, social network, piattaforme di vario genere e strumenti di ricerca bibliografica, come i *discovery tools* usati nelle biblioteche. Tuttavia, il TDM è oggi al centro di criticità diverse dal passato perché viene impiegato anche

¹ Per il panorama internazionale di veda Ruediger Wischenbart Content and Consulting, *Global 50 Publishing Ranking 2023*: <<https://www.wischenbart.com/product/global-50-publishing-ranking-2024/>>. Per quello italiano AIE. Ufficio Studi, *Rapporto sullo stato dell'editoria in Italia 2024*.

² Lettera circolare inviata dall'AIE, a firma di I. Cipolletta e M. Messina, ai rettori delle università italiane il 23 dicembre 2024: «Scriviamo oggi a tutti i Rettori delle università italiane perché sentiamo l'esigenza di avviare un dialogo sul tema della tutela e della valorizzazione del diritto d'autore in ambito universitario. La digitalizzazione e, di recente, il rapido affermarsi di strumenti basati sull'intelligenza artificiale rendono ogni giorno più complessa la gestione di entrambi gli aspetti. In relazione alla tutela, se nel mondo analogico era sufficiente combattere il fenomeno delle fotocopie pirata, oggi le violazioni con cui confrontarsi sono più complesse: dalla messa a disposizione in rete di intere opere fino al commercio di riassunti o slide realizzate senza il consenso degli aventi diritto, che non di rado – ci è stato segnalato – travisano il pensiero degli autori».

³ Per una panoramica: F. L. Kleinkopf, *Text- und Data-Mining*.

per allenare l'intelligenza artificiale, in particolare quella generativa, da parte di aziende private che sfruttano dati pubblici e privati, dando luogo a un ampio ventaglio di questioni etiche e regolamentari⁴. Inoltre, il TDM per l'intelligenza artificiale è divenuto un segmento di mercato sempre più rilevante per molte imprese.

1. *La Scienza Aperta, i dati della ricerca e l'editoria*

1.1. Principali iniziative europee e internazionali per la Scienza Aperta

La Commissione Europea e i principali enti finanziatori internazionali hanno promosso negli ultimi anni numerose iniziative per incentivare la scienza aperta e l'utilizzo dei dati condivisi⁵. È impossibile ricordarle tutte, ma tra esse spiccano:

- *Horizon Europe e il Programma per l'Open Science*. Horizon Europe pone un forte accento sull'accesso aperto alle pubblicazioni e sui principi FAIR (*Findable, Accessible, Interoperable, Reusable*) per i dati della ricerca. Questo include l'obbligo di depositare i dati in repository aperti e di utilizzare licenze che ne consentano il riutilizzo⁶.

⁴ Per queste ultime si veda il *Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio* del 13 giugno 2024: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L_202401689>.

⁵ Per una panoramica sia consentito rimandare a R. Delle Donne, *Open Access, Biblioteche digitali e Big Data*.

⁶ Commissione Europea, *Your Guide to Open Science in Horizon Europe*: <https://intellectual-property-helpdesk.ec.europa.eu/publications/your-guide-open-science-horizon-europe_en>.

- *EOSC (European Open Science Cloud)*. Una piattaforma digitale che mira a fornire agli scienziati europei un’infrastruttura per archiviare, condividere e accedere ai dati della ricerca in modo interoperabile e sicuro⁷.
- *Plan S*. Lanciato da cOAlition S, un consorzio di enti finanziatori della ricerca, Plan S richiede che tutti i risultati della ricerca finanziata siano pubblicati in modalità open access o resi immediatamente disponibili attraverso repository aperti⁸.
- *Iniziative Nazionali e Transnazionali*. Oltre alla Commissione Europea, enti come il Wellcome Trust e il National Institute of Health (NIH) negli Stati Uniti stanno spingendo per politiche simili, evidenziando l’importanza dei dati aperti per accelerare l’innovazione scientifica⁹.

Queste iniziative non solo promuovono una maggiore trasparenza e collaborazione nella ricerca, ma rappresentano anche un modello per gli stati membri e altre organizzazioni internazionali interessate a favorire l’accesso equo e il riutilizzo dei dati scientifici¹⁰.

⁷ Commissione Europea, *European Open Science Cloud (EOSC)*: <https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en>.

⁸ *Plan S. Making full and immediate Open Access a reality*: <https://www.coalition-s.org/about/>. Lanciato nel 2018 ha subito nel tempo numerose messe a punto: cOAlitionS Office, *Annual Review 2023. Accelerating Open Access, 2023*: <<https://www.coalition-s.org/wp-content/uploads/2024/02/PlanS-2023-annual-review.pdf>>.

⁹ Wellcome Trust, *Open Access Policy*: <<https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy>>; National Institute of Health (NIH), *What is the NIH Public Access Policy?*: <<https://www.nih.gov/health-information/nih-clinical-research-trials-you/what-is-nih-public-access-policy>>.

¹⁰ Va ancora ricordato almeno il sostegno dato da *United Nations Educational, Scientific and Cultural Organization (UNESCO)* e da *European University Association (EUS)* all’Open Science: UNESCO, *UNESCO Recommendation on Open*

1.2. La grande editoria internazionale e i dati della ricerca

I grandi editori internazionali, come Elsevier, Springer Nature e Wiley, hanno sviluppato iniziative innovative per adattarsi alle nuove esigenze dell'editoria accademica e alla crescente attenzione per la gestione dei dati della ricerca:

- *Elsevier*. Attraverso piattaforme come Mendeley Data, Elsevier offre soluzioni per l'archiviazione, la condivisione e la gestione dei dati della ricerca¹¹. L'integrazione con strumenti come DataSeer aiuta i ricercatori a identificare e preparare i dati per l'accesso aperto¹².
- *Springer Nature*. L'editore ha lanciato guide per i ricercatori sull'adozione dei principi FAIR e offre servizi per il deposito dei dati attraverso partnership con repository riconosciuti, come figshare¹³.
- *Wiley*. Ha sviluppato una serie di strumenti per il supporto alla gestione dei dati e collabora con enti come Dryad per garantire la disponibilità e la conservazione a lungo termine dei dataset¹⁴.

Science, <<https://www.unesco.org/en/open-science/about?hub=686>>; EUA, *The EUA Open Science Agenda 2025*: <<https://www.eua.eu/publications/positions/the-eua-open-science-agenda-2025.html>>.

¹¹ *Mendeley Data*: <<https://data.mendeley.com/>>.

¹² *About DataSeer. Uncover. Advise. Monitor*: <<https://dataseer.ai/about/>>.

¹³ Springer Nature, *Research Data Policy*: <<https://www.springernature.com/gp/authors/research-data-policy>>; Springer Nature, *Figshare Integration. Guidance for Authors*: <<https://www.springernature.com/gp/authors/research-data/figshare-integration>>.

¹⁴ Wiley, *Open Data. Maximizing the value of research*: <<https://authorservices.wiley.com/author-resources/Journal-Authors/open-research-policies/open-data/index.html>>; Wiley, *Data Sharing with Dryad*: <<https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-service.html>>.

- *Taylor & Francis*. Fornisce risorse dettagliate per i ricercatori su come rendere i dati FAIR e ha collaborato con piattaforme come Code Ocean per integrare codice e dati all'interno delle pubblicazioni¹⁵.
- *Collaborazioni trasversali*. Molti editori partecipano a iniziative globali come Research Data Alliance (RDA) e ORCID per promuovere standard comuni e interoperabilità tra sistemi¹⁶.

Questi sforzi dimostrano come i grandi editori stiano cercando di consolidare la propria posizione non solo come distributori di contenuti, ma anche come facilitatori di pratiche innovative nella gestione e nella condivisione dei dati.

1.3. I dati della ricerca in Italia

In Italia, la gestione dei dati della ricerca è un tema cruciale, alimentato dalle politiche europee come Horizon Europe, dalle iniziative del MUR e, soprattutto, della CRUI (Conferenza dei Rettori delle Università Italiane).

Tra le principali iniziative italiane per promuovere il libero accesso ai dati della ricerca, vanno menzionati i bandi PRIN (Progetti di Rilevante Interesse Nazionale) e SIR (Scientific Independence of young Researchers) del Ministero dell'Università e della Ricerca (MUR). Questi

¹⁵ Taylor & Francis, *Open Data Sharing Policy*: <<https://authorservices.taylorandfrancis.com/data-sharing-policies/open-data/>>; la lista dei data repository raccomandati, sia disciplinari come Code Ocean sia generalisti come FAIRsharing and re3data.org, è alla pagina: <<https://authorservices.taylorandfrancis.com/data-sharing/share-your-data/repositories/>>.

¹⁶ L'elenco delle diverse collaborazioni di Research Data Alliance con editori privati ed enti è alla pagina: <<https://www.rd-alliance.org/organisation-directory/>>; per le collaborazioni con Orcid: <<https://orcid.org/members>>.

bandi prevedono dal 2017 linee guida specifiche per l'adozione di politiche di Open Science e per la gestione dei dati secondo i principi FAIR. Attraverso il finanziamento di progetti che incentivano la condivisione dei dati, il MUR promuove una maggiore trasparenza e accessibilità della ricerca, allineandosi alle *best practice* europee e internazionali.

A supporto di queste iniziative, il *Piano Nazionale per la Scienza Aperta*, recentemente introdotto dal MUR, rappresenta un quadro strategico fondamentale per la promozione dell'Open Science in Italia¹⁷. Il piano definisce obiettivi e strumenti per favorire la condivisione dei dati della ricerca, la trasparenza dei processi scientifici e la valorizzazione dei risultati attraverso l'adozione di infrastrutture digitali e politiche coordinate a livello nazionale. Questo approccio rafforza il ruolo dell'Italia nella comunità scientifica globale, creando sinergie tra enti di ricerca, università e industrie.

La CRUI promuove l'accesso aperto ai risultati della ricerca scientifica e l'Open Science da oltre venti anni, in modo molto incisivo. Più specificamente, il documento sulla gestione dei dati della ricerca prodotto dalla CRUI nel 2024 stabilisce linee guida fondamentali per supportare le università italiane nella transizione verso pratiche di Open Science e adesione ai principi FAIR¹⁸.

¹⁷ MUR, *PNR - Programma Nazionale per la Ricerca*.

¹⁸ La CRUI è dal 2004 il principale attore in Italia per la promozione dell'accesso aperto e della scienza aperta, attraverso le attività della Commissione Biblioteche, al cui interno operano il gruppo di contrattazione con gli editori CARE, coordinato da Nino Grizzuti, e il gruppo Open Access, poi divenuto nel 2020 Osservatorio sulla scienza aperta: R. Delle Donne, *L'accesso aperto, le università e le SSH*. L'Osservatorio sulla scienza aperta, da me coordinato dal 2020, ha pubblicato a ottobre 2024 il documento: *Policy sulla gestione dei dati della ricerca e Data Management Plan (2024)*, <<http://osa.crui.it/wp-content/uploads/2024/12/policy-gestionedati.pdf>>. In diversi contratti con i grandi editori commerciali sottoscritti da CRUI attraverso il gruppo CARE (Elsevier, Springer, Wiley,

I punti principali evidenziati nel documento sono:

- *Pianificazione e Policy Istituzionali*. Le università sono incoraggiate a sviluppare policy interne per la gestione dei dati, definendo ruoli, responsabilità e strumenti per garantire la sostenibilità dei dati stessi.
- *Data Management Plan (DMP)*. L'adozione di piani di gestione dei dati è centrale per strutturare il ciclo di vita dei dati della ricerca, dalla raccolta alla conservazione a lungo termine.
- *Infrastrutture Digitali*. Sebbene il documento non prescriva l'uso di repository specifici, evidenzia l'importanza di infrastrutture affidabili e interoperabili, raccomandando l'utilizzo di piattaforme riconosciute a livello nazionale e internazionale.
- *Formazione e Supporto*. Si sottolinea la necessità di formare i ricercatori e il personale tecnico per implementare le pratiche di gestione dei dati in linea con gli standard internazionali.

Queste indicazioni riflettono un tentativo di colmare il divario tra le pratiche italiane e quelle più avanzate a livello internazionale, promuovendo una cultura della condivisione e della trasparenza nella ricerca.

1.4. Il ruolo del Text and Data Mining

Il *Text and Data Mining (TDM)* rappresenta una delle tecnologie più promettenti per l'analisi avanzata dei dati della ricerca¹⁹. Questa tecnica consente di:

Taylor & Francis) sono previste clausole che consentono il TDM, anche se con alcune limitazioni per Elsevier e Springer.

¹⁹ C. Zong - R. Xia - J. Zhang, *Text Data Mining*. Spunti di riflessione interessanti in R. Morriello, *IA ed editoria scientifica*.

- *Estrarre conoscenza complessa.* Il TDM è in grado di analizzare grandi volumi di dati testuali e numerici per identificare schemi, correlazioni e *trend* che altrimenti sfuggirebbero all'osservazione umana.
- *Supportare la scoperta scientifica.* Attraverso il TDM, è possibile accelerare il progresso scientifico: ad esempio, individuando nuove connessioni tra articoli scientifici, migliorando la diagnosi medica basata su grandi *dataset* clinici e ancora ottimizzando i processi industriali grazie all'analisi di brevetti.
- *Promuovere l'interdisciplinarietà.* L'integrazione di dati provenienti da discipline diverse permette di sviluppare nuove prospettive e soluzioni innovative.
- *Ottimizzare i processi editoriali.* Le redazioni e gli editori possono utilizzare il TDM per individuare lacune nella letteratura disciplinare già esistente e migliorare la qualità delle pubblicazioni.

Nonostante il suo potenziale, l'accesso ai dati per il TDM è spesso regolato da politiche restrittive imposte dai grandi editori. Alcuni punti critici includono:

- *Licenze specifiche.* Molti editori richiedono licenze aggiuntive per consentire il TDM sui contenuti da loro pubblicati, limitando l'accesso ai dati anche agli abbonati delle loro piattaforme²⁰.
- *Barriere tecniche.* Gli editori implementano spesso misure tecniche, come CAPTCHA o particolari limiti alla possibilità di scaricare dati, che ostacolano l'automazione necessaria per il TDM.
- *Costi elevati.* Alcuni editori impongono costi significativi per consentire il TDM, rendendolo inaccessibile per ricercatori con budget limitati.

²⁰ Si veda la rassegna: Cranfield University. Library Services, *Publisher policies for TDM*.

Negli ultimi anni, alcune iniziative globali hanno cercato di rendere il TDM più accessibile, non senza alcune ombre:

- *Politiche dell'Unione Europea*. La direttiva sul diritto d'autore del 2019 ha introdotto eccezioni che consentono il TDM per scopi di ricerca scientifica, riducendo le restrizioni legali per istituzioni di ricerca e università. Questo rappresenta un passo significativo verso la democratizzazione dell'accesso ai dati per finalità scientifiche. In Italia, diversamente da altri paesi, le due eccezioni relative al TDM previste dalla Direttiva sono state correttamente recepite dal legislatore, che ne ha garantito l'applicazione anche in ambiti non strettamente legati alla ricerca accademica, posizionandosi come un esempio positivo nel panorama europeo²¹.
- *Partnership pubblico-privato*. Alcuni editori, come Elsevier e Springer Nature, hanno sviluppato politiche specifiche per il TDM, che sembrano venire incontro alle esigenze dei ricercatori, ma che in realtà suscitano preoccupazioni significative riguardo al controllo e alla gestione delle risorse di dati. Elsevier richiede che le istituzioni aggiornino i loro accordi di licenza per consentire l'accesso ai contenuti a fini di TDM, limitando tale accesso attraverso un'*Application Programming Interface* (API) proprietaria. Questo approccio è stato criticato perché impone condizioni restrittive e limita l'uso di tecniche comuni di *mining*, come l'uti-

²¹ Sul recepimento in Europa: G. Vrakas, *A literature review of "lawful" text and data mining*. Per il recepimento in Italia: B. Calabrese, *Scientific TDM Exception and Communication to the Public*; D. De Angelis, *The Italian Implementation of the New EU Text and Data Mining Exceptions*. Più in generale, sulla Direttiva, si legga A. Musso, *Eccezioni e limitazioni ai diritti d'autore*; sul recepimento in Italia di un'altra eccezione fondamentale per il mondo della ricerca, relativa alla conservazione del patrimonio culturale: R. Maiello-G. Mazziotti-C. Sappa, *Qualche spunto sull'eccezione per la conservazione del patrimonio culturale*.

lizzo di robot o crawler. Inoltre, Elsevier richiede ai ricercatori di registrarsi come sviluppatori e ottenere un'APIKey, imponendo ulteriori vincoli operativi²². Springer Nature, invece, offre una varietà di API per consentire il TDM su contenuti come riviste, libri e altre risorse. Mentre i ricercatori possono scaricare materiali direttamente dalle piattaforme di Springer Nature, l'accesso tramite API dedicate comporta costi aggiuntivi. Per le istituzioni accademiche e governative che hanno sottoscritto abbonamenti, Springer Nature include generalmente i diritti di TDM senza ulteriori costi, prevedendoli nelle licenze di abbonamento. Tuttavia, queste condizioni di utilizzo, che si applicano in caso di nuovi contratti o di rinnovi di abbonamenti, comportano costi aggiuntivi per quegli utenti che richiedono invece l'accesso prima del rinnovo²³.

In altri termini, le condizioni poste dagli editori finiscono con lo sbilanciare a loro vantaggio i rapporti con la comunità accademica, creando barriere economiche e operative che finiscono con il limitare le possibilità di molti ricercatori di utilizzo dei dati. È quindi essenziale che, in fase di contrattazione degli abbonamenti con gli editori, siano negoziati anche i diritti relativi all'uso del TDM; a maggior ragione, perché la direttiva EU 2019/790 e le norme di recepimento nazionale prevedono la nullità di patti contrari alle eccezioni previste sul TDM. A tal fine è necessario un dialogo con gli editori più equo e trasparente, per assicurare che il TDM resti uno strumento accessibile e inclusivo, evitando che si levino nuove barriere che ostacolino la democratizzazione della conoscenza. Inoltre, è importante sottolineare che l'uso di

²² R. Van Noorden, *Elsevier opens its papers to text-mining*.

²³ Springer Nature, *Text and Data Mining*.

API proprietarie può complicare il processo di *mining*, rendendo più difficile per i ricercatori accedere ai dati in modo efficiente. Sarebbe auspicabile l'adozione di tecnologie open source che consentano di lavorare direttamente sui PDF, evitando le restrizioni imposte dalle soluzioni proprietarie impiegate dagli editori. In sintesi, mentre le partnership pubblico-privato nel campo del TDM possono offrire opportunità per l'innovazione, è fondamentale garantire che le politiche adottate non limitino l'accesso equo ai dati e non impongano restrizioni che ostacolino la ricerca scientifica.

- *Open Access e TDM*. Le pubblicazioni in open access rappresentano una risorsa cruciale per il TDM, dal momento che i contenuti sono liberamente accessibili e non soggetti alle limitazioni imposte dagli editori commerciali. L'adozione di politiche open access da parte di molte istituzioni accademiche e degli enti finanziatori della ricerca contribuisce a rendere il TDM uno strumento realmente inclusivo, riducendo le barriere economiche e promuovendo la diversità delle fonti disponibili per l'analisi. Tuttavia, persistono sfide legate alla qualità e all'interoperabilità dei dati, che richiedono un impegno costante nella definizione di standard comuni.

In conclusione, l'adozione del TDM continua a essere un terreno di confronto tra editori, ricercatori e *policy maker*, ma il suo potenziale per la scienza aperta e l'innovazione lo rende una componente cruciale per il futuro dell'editoria accademica. Nonostante il suo potenziale, il TDM continua a scontrarsi con ostacoli normativi e tecnologici. Anche in Italia, l'accesso ai dati per il TDM è spesso limitato dalle condizioni imposte dagli editori nazionali e internazionali.

2. *LLM e RAG: le nuove frontiere della conoscenza*

I “modelli linguistici di grandi dimensioni” (Large Language Models, LLM) e le tecniche di “generazione aumentata del recupero” dei dati (Retrieval-Augmented Generation, RAG) rappresentano innovazioni significative nell’ambito dell’intelligenza artificiale applicata all’editoria accademica e alla gestione della conoscenza. Questi strumenti combinano la potenza della comprensione linguistica avanzata con la capacità di accedere a dati esterni, creando risposte puntuali e altamente personalizzate. Gli LLM, come GPT di OpenAI o Llama di Meta, sono modelli di *machine learning* pre-addestrati su vasti *dataset* per comprendere e generare linguaggio naturale. Le tecniche di RAG combinano questi modelli con sistemi di recupero dei dati per generare risposte contestualizzate, migliorando la precisione e riducendo errori²⁴.

I loro ambiti di applicazione nell’editoria sono:

- *L’automazione della produzione editoriale.* Gli LLM generano abstract, traduzioni e riassunti con notevole rapidità e precisione. I RAG li integrano con dati aggiornati, rendendo le risposte più pertinenti e accurate.
- *L’accesso personalizzato alla conoscenza.* Le tecniche di RAG consentono di filtrare e personalizzare i contenuti accademici in base alle esigenze degli utenti.
- *La peer review assistita.* Gli strumenti basati su LLM e RAG possono automatizzare il controllo della qualità e dell’originalità dei manoscritti.

²⁴ Per gli LLM: W. Xin Zhao-K. Zhou-J. Li et al., *A Survey of Large Language Models*. Per RAG: Y. Gao-Y. Xiong-X. Gao et al., *Retrieval-Augmented Generation for Large Language Models*.

Gli LLM rappresentano una delle innovazioni più potenti e controverse nell'ambito dell'intelligenza artificiale applicata alla conoscenza. Il loro utilizzo solleva infatti notevoli interrogativi, riconducibili ad alcune macrotematiche: a) epistemologia e verità, *bias* e inclusività; b) ruolo dell'umano. Inoltre, completano il quadro complessivo delle loro implicazioni le opportunità e le sfide operative, come l'efficienza (c) e la governance dei dati (d).

- *Bias, verità e inclusività nei modelli linguistici avanzati.* Gli LLM generano risposte probabilistiche, coerenti con i dati di addestramento, ma non sempre allineate ai principi tradizionali della verità scientifica, quali la riproducibilità e la revisione paritaria. Questo introduce rischi significativi, come la perpetuazione di *bias* e di disuguaglianze sistemiche. I *dataset* di addestramento, spesso influenzati da pregiudizi di genere, razza o classe sociale, amplificano queste distorsioni, marginalizzando contributi di minoranze o aree meno rappresentate. Per contenere tali rischi, è essenziale ampliare la diversità dei dati, sviluppare algoritmi correttivi e coinvolgere esperti di diverse discipline. In definitiva, sono sfide che sollecitano una riflessione critica sul ruolo della verità e della neutralità all'interno dei modelli linguistici avanzati. Ad esempio, un modello linguistico potrebbe generare risposte apparentemente corrette basate su dati distorti, e suggerire che una minoranza culturale abbia contribuito meno di altre alla produzione scientifica globale semplicemente perché i *dataset* di addestramento riflettevano una sua sottorappresentazione storica. Questo esempio illustra la necessità di progettare *dataset* sempre più inclusivi e algoritmi che sappiano riconoscere e compensare tali squilibri.
- *Collaborazione uomo-macchina: etica e responsabilità nell'innovazione.* L'efficienza e la precisione degli LLM, che rappresentano un prezioso supporto per analisi preliminari e automazione di

- processi ripetitivi, devono essere bilanciate da un controllo umano attento e rigoroso, soprattutto in ambiti che richiedono una comprensione critica e contestuale. Ad esempio, la supervisione umana potrebbe essere decisiva nell'identificare *bias* nascosti o risposte inesatte generate dai modelli, specialmente quando i *dataset* riflettono pregiudizi sistemici. Questa sinergia tra uomo e macchina non solo mitiga i rischi epistemologici, ma rafforza anche la fiducia nel loro utilizzo in contesti accademici e operativi.
- *Efficienza, scalabilità e sostenibilità.* Gli LLM possono migliorare significativamente l'efficienza di diverse attività accademiche ed editoriali, automatizzando processi come la revisione dei testi, la generazione di abstract e la traduzione dei contenuti. Tuttavia, la loro implementazione richiede infrastrutture avanzate, competenze specifiche e una gestione sostenibile dei costi energetici e dell'impatto ambientale.
 - *Governance dei dati e collaborazione globale.* Le sfide epistemologiche e sociali associate all'uso degli LLM evidenziano rischi significativi, rendendo necessaria una gestione dei dati trasparente ed equa, fondamentale per generare fiducia nelle comunità scientifiche e indurle all'adozione di queste tecnologie avanzate per l'analisi e la rielaborazione dei dati. La protezione della privacy e la conformità alle normative vigenti, come il GDPR in Europa, sono elementi chiave per costruire e mantenere tale fiducia. La governance dei dati deve garantire un accesso equo, evitando che il loro controllo si traduca in concentrazioni di conoscenza e di potere nelle mani di grandi attori tecnologici. Allo stesso tempo, l'intelligenza artificiale può favorire la condivisione della conoscenza e promuovere collaborazioni inclusive su scala internazionale, purché si affrontino le disuguaglianze e le disparità tecnologiche che persistono tra i diversi paesi.

3. *La conoscenza come bene comune e la memoria culturale*

Nell'affrontare il ruolo delle nuove tecnologie nella gestione e diffusione del sapere, diviene essenziale considerare il valore della conoscenza come bene comune e il suo legame indissolubile con la memoria culturale. Questa prospettiva invita a riflettere sul rapporto tra innovazione tecnologica e conservazione del patrimonio immateriale.

- *La conoscenza come bene comune e il paradosso della proprietà intellettuale.* L'adozione di tecnologie come LLM e RAG rinnova una tensione storica: la conoscenza scientifica intesa come bene comune opposta al controllo proprietario da parte di aziende e istituzioni²⁵. Questo dibattito, che affonda le sue radici nell'Illuminismo, richiama la necessità di bilanciare l'apertura del sapere con la protezione dei diritti. Mentre i principi dell'*Open Science* promuovono la condivisione della conoscenza per il bene collettivo, la realtà tecnologica odierna è spesso dominata da brevetti e licenze restrittive che limitano l'accesso. Questo paradosso evidenzia la necessità di politiche globali che favoriscano la democratizzazione della conoscenza senza penalizzare la ricerca dell'innovazione.
- *Tecnologia e asimmetrie di potere.* Le tecnologie avanzate, come LLM e RAG, amplificano le disuguaglianze già esistenti nell'editoria accademica. La capacità di sfruttare appieno queste tecnologie dipende dall'accesso a risorse avanzate, come infrastrutture computazionali e *dataset* di alta qualità, che sono spesso monopolizzate da grandi aziende tecnologiche. Questa dinamica concentra ulteriormente il potere nelle mani di pochi attori, raf-

²⁵ Si rimanda al volume collettaneo *Understanding Knowledge as a Commons*.

forzando quella che Michel Foucault definirebbe una biopolitica del sapere²⁶. Per contrastare tali asimmetrie, è necessario promuovere infrastrutture aperte e investimenti pubblici che garantiscano un accesso equo alle risorse tecnologiche.

- *Rischi e opportunità della mediaticità digitale.* Come ogni medium innovativo, LLM e RAG stanno ridefinendo il rapporto tra umano e sapere²⁷. La velocità nella generazione di contenuti rappresenta un'opportunità per democratizzare l'accesso all'informazione, ma rischia di sacrificare la profondità analitica e la riflessione critica. Inoltre, la mediaticità digitale può favorire un consumo passivo della conoscenza, sostituendo la ricerca attiva con risposte sintetiche generate automaticamente. Per sfruttare appieno le potenzialità di queste tecnologie, è fondamentale integrarle con metodologie tradizionali che valorizzino il pensiero critico e l'interpretazione contestuale.
- *Il ruolo dell'editoria nella conservazione della memoria culturale.* L'editoria non è solo produzione di contenuti, ma anche conservazione della memoria collettiva. Le tecnologie come LLM e RAG possono rafforzare questa funzione facilitando la catalogazione, la ricerca e il recupero di informazioni, ma introducono anche nuovi rischi. L'automazione nella gestione dei contenuti potrebbe favorire la perdita di materiali ritenuti marginali o di

²⁶ M. Foucault, *Securité, territoire, population*.

²⁷ Già Marshall McLuhan, in *Understanding Media: The Extensions of Man*, aveva richiamato l'attenzione sul fatto che i media influenzino il nostro modo di percepire e interagire con il mondo, anticipando molte riflessioni sull'impatto che i nuovi strumenti tecnologici hanno sul rapporto tra umano e sapere. Più recentemente Shoshana Zuboff, in *The Age of Surveillance Capitalism*, ha analizzato l'impatto delle tecnologie digitali sulla società, mettendo in luce i rischi connessi alla mediaticità digitale di concentrazione del potere e di riduzione dell'individuo a consumatore passivo di dati e informazioni.

nicchia e considerati meno rilevanti da criteri algoritmici come rilevanza dei contenuti, popolarità, valutazioni di qualità, *bias* incorporati nei dati di addestramento, ordine di priorità, filtraggio di “rumori” ecc. Perché l’editoria possa continuare a contribuire alla trasmissione della memoria culturale, è fondamentale riconoscere il suo legame con i depositi della memoria, come biblioteche e archivi, che conservano e custodiscono il patrimonio culturale collettivo, come sottolineato da Aleida Assmann²⁸. Per svolgere efficacemente questa funzione, l’editoria dovrebbe integrare strumenti tecnologici avanzati con una curatela umana capace di valorizzare le diverse prospettive storiche e culturali, evitando che i criteri algoritmici riducano la complessità del sapere a schemi standardizzati o limitino la rappresentazione delle voci marginali.

Conclusione

In un contesto di rapida innovazione tecnologica, è fondamentale trovare un equilibrio tra automazione e supervisione umana. Gli LLM e le tecnologie di RAG offrono straordinarie opportunità per aumentare l’efficienza e democratizzare l’accesso alla conoscenza, ma il loro impiego deve essere guidato da una visione etica e sostenibile. Progetti infrastrutturali come European Open Science Cloud (EOSC), pur non utilizzando direttamente LLM o RAG, mostrano come la collaborazione tra editori, ricercatori e *policy maker* possa creare ecosistemi di dati che abilitano nuove applicazioni basate sull’intelligenza artificiale. EOSC, ad esempio, supporta la scienza

²⁸ A. Assmann, *Erinnerungsräume*.

aperta garantendo l'accesso a dati FAIR, una risorsa essenziale per addestrare tecnologie avanzate in modo trasparente e responsabile.

Le riflessioni sviluppate da Gino Roncaglia sull'intelligenza artificiale in un suo recente volume inducono a considerare il ruolo di questi strumenti non solo come tecnologie operative, ma anche come componenti essenziali di una trasformazione epistemologica²⁹. Roncaglia evidenzia come l'IA generativa, con il suo approccio oracolare-predittivo, rappresenti sia una sfida sia un'opportunità per strutturare e aggregare conoscenza in un'"età della frammentazione". Il compito degli "architetti del sapere", di studiosi, educatori, bibliotecari, archivisti, editori e progettisti di sistemi informativi impegnati nella costruzione di infrastrutture e metodologie per organizzare e trasmettere conoscenza, diventa quindi cruciale per integrare strumenti di IA in modelli collaborativi e inclusivi, come quelli rappresentati da Wikipedia e dai Linked data, superando i limiti di un'informazione dispersa e frammentata.

In questa prospettiva, anche gli atenei possono e devono svolgere un ruolo chiave, promuovendo l'adozione di tecnologie digitali in linea con i principi della scienza aperta e favorendo un approccio critico al loro utilizzo. La capacità di coniugare l'efficienza degli strumenti di IA con metodologie che valorizzino il pensiero critico e l'interpretazione contestuale rappresenta una sfida centrale per evitare che la velocità della generazione automatica di contenuti sacrifichi la profondità analitica e la riflessione. È necessario progettare un futuro in cui i valori fondamentali della conoscenza – apertura, inclusione e sostenibilità – siano rispettati e rafforzati.

Solo attraverso iniziative che promuovano interoperabilità, equità e responsabilità sarà possibile costruire un futuro in cui l'intel-

²⁹ G. Roncaglia, *L'architetto e l'oracolo*.

ligenza artificiale diventi uno strumento davvero al servizio della conoscenza, favorendo non solo l'innovazione tecnologica, ma anche una comprensione più profonda e consapevole del sapere umano.

Bibliografia

Tutti gli URL sono stati verificati il 13 gennaio 2025

- AIE. Ufficio Studi, *Rapporto sullo stato dell'editoria in Italia 2024 su consolidato 2023*, Milano, Associazione Italiana Editori, 2024.
- A. Assmann, *Erinnerungsräume: Formen und Wandlungen des kulturellen Gedächtnisses*, München, Verlag C.H. Beck, 1999; trad. it. *Ricordare. Forme e mutamenti della memoria culturale*, Bologna, il Mulino, 2002.
- B. Calabrese, *Scientific TDM Exception and Communication to the Public: Did Italians Do It Better... or at Least Not Worse?*, in «Journal of Intellectual Property Law & Practice», 17, 5, May 2022, pp. 399-406, <<https://doi.org/10.1093/jiplp/jpac026>>.
- I. Cipolletta - M. Messina, *Lettera circolare dell'AIE ai rettori delle università italiane*, 23 dicembre 2024.
- cOAlitionS Office, *Annual Review 2023. Accelerating Open Access*, 2023, <<https://www.coalition-s.org/wp-content/uploads/2024/02/PlanS-2023-annual-review.pdf>>.
- cOAlitionS, *Plan S. Making full and immediate Open Access a reality*, <<https://www.coalition-s.org/about/>>.
- Commissione Europea, *European Open Science Cloud (EOSC)*, Bruxelles, 2023, <https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en>.
- Commissione Europea, *Your Guide to Open Science in Horizon Europe*, Bruxelles, 2024, <https://intellectual-property-helpdesk.ec.europa.eu/publications/your-guide-open-science-horizon-europe_en>.
- Cranfield University. Library Services, *Publisher policies for TDM*, <<https://library.cranfield.ac.uk/text-and-data-mining/publisher-policies>>.
- CRUI. Osservatorio sulla scienza aperta, *Policy sulla gestione dei dati della ricerca e Data Management Plan (2024)*, Roma, 2024, <<http://osa.crui.it/wp-content/uploads/2024/12/policy-gestionedati.pdf>>.
- DataSeer Research Data Service, *About DataSeer. Uncover. Advise. Monitor*, <<https://dataseer.ai/about/>>.

- D. De Angelis, *The Italian Implementation of the New EU Text and Data Mining Exceptions*, in «Communia», Dec 14, 2022, <<https://communia-association.org/2022/12/14/italian-implementation-of-the-new-eu-tdm-exceptions/>>.
- R. Delle Donne, *L'accesso aperto, le università e le SSH / Open Access, Universities and SSH*, in «Il capitale culturale», 17, 2018, pp. 17-45, <<https://doi.org/10.13138/2039-2362/1944>>.
- R. Delle Donne, *Open Access, Biblioteche digitali e Big Data*, in *I magnifici incontri 2018. Piano nazionale università digitale*, a cura di G. Bernardi - A. Tercig, Forum, Udine, pp. 87-96.
- Elsevier, *Mendeley Data*, <<https://data.mendeley.com/>>.
- European University Association, *The EUA Open Science Agenda 2025*, Bruxelles, 2023, <<https://www.eua.eu/publications/positions/the-eua-open-science-agenda-2025.html>>.
- M. Foucault, *Securité, territoire, population*, Paris, Seuil-Gallimard, 2004; trad. it. *Sicurezza, territorio, popolazione*, Milano, Feltrinelli, 2007.
- Y. Gao - Y. Xiong - X. Gao - K. Jia - J. Pan - Y. Bi - Y. Dai - J. Sun - M. Wang - H. Wang, *Retrieval-Augmented Generation for Large Language Models: A Survey*, in «ArXiv», 27 Mar 2024, <<https://doi.org/10.48550/arXiv.2312.10997>>.
- F. L. Kleinkopf, *Text- und Data-Mining. Die Anforderungen digitaler Forschungsmethoden an ein innovations- und wissenschaftsfreundliches Urheberrecht*, Baden-Baden, Nomos Verlagsgesellschaft, 2022.
- R. Maiello - G. Mazziotti - C. Sappa, *Qualche spunto sull'eccezione per la conservazione del patrimonio culturale*, in *Crisi e resilienza del diritto d'autore. Il recepimento italiano della direttiva 2019/790*, a cura di M.L. Bixio, Milano, Giappichelli, 2024, pp. 51-62.
- M. McLuhan, *Understanding Media: The Extensions of Man*, New York, McGraw-Hill, 1964; trad. it. *Gli strumenti del comunicare*, Milano, Il Saggiatore, 1967.
- R. Morriello, *IA ed editoria scientifica: come la Direttiva UE sta cambiando le regole del gioco*, in «Informatica Umanistica e Cultura Digitale: il blog dell'AIUCD», 31 Dicembre 2024, <<https://doi.org/10.58079/1301b>>.
- MUR, *PNR - Programma Nazionale per la Ricerca 2021-2027. Piano Nazionale per la Scienza Aperta*, <https://www.mur.gov.it/sites/default/files/2022-06/Piano_Nazionale_per_la_Scienza_Aperta.pdf>.
- A. Musso, *Eccezioni e limitazioni ai diritti d'autore nella Direttiva UE n. 790/2019*, in «Il diritto dell'informazione e dell'informatica», 36, 3, 2020, pp. 411-464.
- National Institute of Health (NIH), *What is the NIH Public Access Policy?*, <<https://www.nih.gov/health-information/nih-clinical-research-trials-you/what-is-nih-public-access-policy>>.

- Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio*, 13 giugno 2024: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L_202401689>.
- G. Roncaglia, *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*, Roma-Bari, Laterza, 2023.
- Ruediger Wischenbart Content and Consulting, *Global 50 Publishing Ranking 2023*, Wien, 2024, <<https://www.wischenbart.com/product/global-50-publishing-ranking-2024/>>.
- Springer Nature, *Figshare Integration. Guidance for Authors*, <<https://www.springernature.com/gp/authors/research-data/figshare-integration>>.
- Springer Nature, *Research Data Policy*, <<https://www.springernature.com/gp/authors/research-data-policy>>.
- Springer Nature, *Text and Data Mining/TDM*, in «Nature Portfolio», Nov. 6, 2024, <https://support.nature.com/en/support/solutions/articles/6000251800-text-and-data-mining-tdm?utm_source=chatgpt.com>.
- Taylor & Francis, *Open Data Sharing Policy*, <<https://authorservices.taylorandfrancis.com/data-sharing-policies/open-data/>>.
- Understanding Knowledge as a Common. From Theory to Practice*, edited by Ch. Hess - E. Ostrom, Massachusetts - London, England, Cambridge, The MIT Press, 2007; trad. it. *La conoscenza come bene comune. Dalla Teoria alla pratica*, a cura di P. Ferri, Milano, Mondadori, 2009.
- UNESCO, *UNESCO Recommendation on Open Science*, <<https://www.unesco.org/en/open-science/about?hub=686>>.
- R. Van Noorden, *Elsevier opens its papers to text-mining*, in «Nature», 506, 17, 2014, <<https://doi.org/10.1038/506017a>>.
- Wellcome Trust, *Open Access Policy*, <<https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy>>.
- Wiley, *Open Data. Maximizing the value of research*, <<https://authorservices.wiley.com/author-resources/Journal-Authors/open-research-policies/open-data/index.html>>.
- Wiley, *Data Sharing with Dryad*, <<https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-service.html>>.
- G. Vrakas, *A literature review of "lawful" text and data mining*, in «Open Research Europe», 30 Oct 2024, <<https://doi.org/10.12688/openreseurope.18013.2>>.
- W. Xin Zhao - K. Zhou - J. Li - T. Tang - X. Wang - Y. Hou - Y. Min - B. Zhang - J. Zhang - Z. Dong - Y. Du - C. Yang - Y. Chen - Z. Chen - J. Jiang - R. Ren - Y. Li - X. Tang - Z. Liu - P. Liu - J.-Y. Nie - J.-R. Wen, *A Survey of Large Language Models*, in «ArXiv», last revised 13 Oct 2024 (v15), <<https://doi.org/10.48550/arXiv.2303.18223>>.

L'editoria accademica, i dati della ricerca e l'intelligenza artificiale

- C. Zong - R. Xia - J. Zhang, *Text Data Mining*, Singapore, Springer, 2021.
S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London, Profile Books Ltd, 2019; trad. it. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, LUISS University Press, 2023.

Roberto Delle Donne
Università degli Studi di Napoli Federico II
roberto.delledonne@unina.it
<https://orcid.org/0000-0001-8331-9436>

