

Cataloguing, Metadata, and Generative AI. Early Experiences and Future Perspectives

Gino Roncaglia^(a)

a) Roma Tre University, <https://orcid.org/0000-0003-4632-3342>

Contact: Gino Roncaglia, gino.roncaglia@uniroma3.it

Received: 30 July 2025; **Accepted:** 05 November 2025; **First Published:** 15 January 2026

ABSTRACT

The article deals with the intersection of generative artificial intelligence (AI) and bibliographic/metadata practices, assessing how large language models (LLMs) can support cataloguing and metadata creation while navigating the constraints of formal knowledge architectures. In the first section, the article discusses the evolution of cataloguing paradigms from MARC to Linked Open Data (LOD), emphasizing the shift from rigid records to semantic, entity-based models like FRBR, RDA, and BIBFRAME. The second section deals with the epistemological clash between deterministic, rule-based metadata standards (the “architect”) and probabilistic, generative AI systems (the “oracle”).

Three strategies are discussed for integrating AI into bibliographic workflows:

- 1) Specialized AI systems trained exclusively on controlled, high-quality datasets.
- 2) Retrieval-Augmented Generation (RAG), blending LLMs with authoritative knowledge bases.
- 3) Next-generation LLMs enhanced via reasoning models, multimodal inputs, expanded context windows, and small/medium-scale local models to align generative outputs with metadata standards.

Key challenges include hallucinations, data sparsity in bibliographic corpora, and the obsolescence of MARC-centric experiments. The article argues for caution against retrofitting AI onto outdated data models, urging alignment with LOD and IFLA’s Library Reference Model (LRM). Ethical considerations (bias, transparency, AI literacy) and the potential of local SLMs/MSLMs for privacy-sensitive applications are highlighted.

KEYWORDS

Artificial intelligence; Metadata creation/management; LLM; Bibliographic description; Cataloguing standards.

Catalogazione, metadati e IA generativa. Prime esperienze e prospettive future

ABSTRACT

L'articolo affronta l'intersezione tra intelligenza artificiale generativa (AI) e l'ambito catalografico e di gestione dei metadati, valutando in che modo i modelli linguistici di grandi dimensioni (LLM) possano supportare la catalogazione e la creazione di metadati, rispettando i vincoli posti dalle architetture formali della conoscenza.

Nella prima sezione, l'articolo analizza l'evoluzione dei paradigmi di catalogazione, dal formato MARC ai Linked Open Data (LOD), sottolineando il passaggio da record rigidi a modelli semantici basati su entità, come FRBR, RDA e BIBFRAME. La seconda sezione tratta del conflitto epistemologico tra gli standard di metadattazione formalizzati e basati su regole (l’“architetto”) e i sistemi di intelligenza artificiale generativa, per loro natura prevalentemente non deterministici e probabilistici (l’“oracolo”). Per risolvere tale conflitto, vengono prese in esame tre strategie per integrare l'AI nei flussi di lavoro bibliografici:

- 1) sistemi di AI specializzati, addestrati esclusivamente su dataset controllati e di alta qualità;
- 2) modelli di tipo Retrieval-Augmented Generation (RAG), che combinano LLM e basi di conoscenza autorevoli;
- 3) LLM di nuova generazione potenziati da modelli di ragionamento, input multimodali, finestre contestuali estese e modelli locali di piccola/media scala con fine tuning, per allineare gli output generativi agli standard di metadattazione.

Tra le principali criticità vengono evidenziati fenomeni di allucinazione, scarsità di dati nei corpora bibliografici e obsolescenza degli esperimenti basati su MARC. L'articolo invita alla cautela rispetto ai tentativi di adattare l'AI a modelli di dati superati, promuovendo invece l'allineamento con i LOD e con il Library Reference Model (LRM) dell'IFLA. Vengono infine discusse le implicazioni etiche (bias, trasparenza, alfabetizzazione all'AI) e il potenziale dei modelli SLM/MSLM locali per applicazioni che richiedono maggiore tutela della privacy.

PAROLE CHIAVE

Intelligenza artificiale; Metadattazione; LLM; Descrizioni bibliografiche; Standard di catalogazione.

0. Premessa

Questo articolo ha l'obiettivo di proporre alcune considerazioni sul ruolo che può avere l'intelligenza artificiale generativa in campo bibliotecario e documentale, in particolare per quanto riguarda le attività di catalogazione e metadattazione. Si tratterà di considerazioni non solo sintetiche ma anche inevitabilmente provvisorie, considerata la velocità di evoluzione del settore e in particolare dei grandi modelli linguistici, che rappresentano – almeno nel momento in cui scrivo – lo strumento più avanzato utilizzabile in questo contesto.

Prima di illustrare la struttura del testo, una premessa indispensabile riguarda l'uso dell'IA generativa nel lavoro di scrittura dell'articolo. Considerato il tema trattato, infatti, mi è sembrato non solo utile ma anche opportuno interagire in maniera abbastanza stretta con alcuni fra i grandi modelli linguistici (d'ora in poi LLM – *Large Language Model*) che rappresentano riferimenti obbligati per la trattazione. Nello specifico, gli strumenti utilizzati sono stati ChatGPT versioni 4o e o3 (di norma con Deep Research), Manus (una piattaforma basata su Claude Sonnet 4, con funzionalità avanzate di Deep Research) e il modello cinese Kimi K2 con funzionalità Researcher attiva, che nel momento in cui scrivo rappresenta una novità nel campo dei LLM e le cui prestazioni sembrano decisamente interessanti.¹

L'interazione con questi modelli ha riguardato tre fasi diverse del lavoro: 1. la richiesta al modello di fornire proposte su temi rilevanti da trattare, sulla base di un prompt che spiegava gli obiettivi dell'articolo e le sezioni principali in cui mi proponevo di suddividerlo; 2. la richiesta di approfondimenti specifici, riferimenti ulteriori e suggerimenti su casi rilevanti; 3. l'esame separato dei testi delle varie sezioni dell'articolo, nella mia stesura, con un prompt che chiedeva di individuare errori, punti deboli o riferimenti mancanti.

Le indicazioni fornite sono risultate in diversi casi utili (e in altri casi a mio avviso sbagliate o fuorvianti), ma tanto la stesura quanto la responsabilità del testo sono ovviamente del tutto mie, con una sola eccezione interessante che segnalerò in nota (v. oltre, nota 15). L'obiettivo principale nell'uso di questi sistemi era verificare se e in che misura il ricorso all'IA generativa in forma di *sparring partner* durante la stesura di un lavoro di ricerca potesse risultare utile. La valutazione del risultato spetta ovviamente a chi legge, ma la mia impressione come autore è che l'utilità ci sia (soprattutto nella segnalazione di alcune esperienze e risorse rilevanti e di alcune imprecisioni). Non

¹ Sulle differenze – in particolare in termini di allucinazioni e capacità di risalire a letteratura secondaria rilevante – fra modelli tradizionali e modelli Deep Research si veda (Roncaglia 2025a). Una sintetica spiegazione è fornita anche in chiusura di questo articolo.

vi è invece – né mi aspettavo ci fosse, considerato che non intendevo in alcun modo rinunciare al controllo autoriale – un risparmio di tempo nella stesura: direi semmai che il lavoro ha richiesto più tempo di quello che sarebbe stato necessario senza l'uso di questi strumenti².

Oltre a questa premessa, l'articolo comprende due sezioni principali:³ la prima è dedicata all'evoluzione dei paradigmi catalografici e di metadattazione, con l'obiettivo di rendere chiari i problemi legati alla loro standardizzazione ed evoluzione nel tempo, e il tipo di vincoli che la loro natura

² In sede di revisione anonima dell'articolo, pur apprezzando le indicazioni fornite nei paragrafi precedenti sulle modalità d'uso di LLM nel lavoro di stesura dell'articolo, è stato osservato che la disponibilità integrale delle interazioni avute con i LLM utilizzati – prompt usati e risposte ricevute – garantirebbe una trasparenza ancor maggiore. Si tratta di una osservazione mossa da una esigenza largamente condivisibile, quella di massima trasparenza nell'uso di strumenti di rete, e particolarmente di strumenti delicati e potenzialmente fuorvianti come indubbiamente sono i sistemi di intelligenza artificiale generativa, ma (non solo nel caso del mio articolo) anche assai problematica dal punto di vista pratico. Ho provato a fare una valutazione necessariamente parziale (dato che alcuni dei sistemi usati non conservano nel tempo tutte le interazioni avute, o le conservano solo per un periodo di tempo limitato o solo mantenendo attivo un abbonamento a pagamento), e posso dire che complessivamente nel lavorare sul tema i miei prompt e le risposte dei vari sistemi usati corrispondono ad almeno 230.000 (e probabilmente a più di 250.000) battute di testo: ovviamente non sarebbe possibile includerle tutte in questa sede (l'articolo, già in questa forma, supera largamente il limite di 50.000 battute che era stato fissato inizialmente dalla redazione). Inoltre, si tratterebbe di materiali potenzialmente fuorvianti, dato che includono alcune allucinazioni e diverse indicazioni a mio avviso sbagliate o problematiche, ciascuna delle quali richiederebbe una discussione specifica. Astrattamente, se fosse stato usato un singolo sistema capace di offrire una buona garanzia di permanenza nel tempo delle interazioni, con un link persistente e affidabile, ci si potrebbe limitare a fornire il link; ma i risultati andrebbero a quel punto comunque commentati, per spiegarne l'uso specifico; inoltre ho appunto usato più sistemi, e al momento direi che nessuno dei LLM disponibili offra garanzie sull'affidabilità e la persistenza dei link alle interazioni (che, al contrario, spesso spariscono). D'altro canto l'esigenza, certo condivisibile, della massima trasparenza relativamente a fonti e strumenti usati deve sempre – e non solo nel caso dei LLM – fare i conti con la ragionevolezza pratica: per fare solo qualche esempio, nel lavoro di ricerca tutte e tutti noi interagiamo sempre e comunque con colleghe e colleghi con cui discutiamo tesi o parti di un lavoro, ed è ottima pratica riconoscerlo nei ringraziamenti, ove possibile anche in maniera abbastanza specifica; ma – tranne in casi particolari – non è concretamente possibile (e appesantirebbe moltissimo il testo) citare *verbatim* tutte le interazioni avute, anche se queste hanno in alcuni casi contribuito a influenzare o indirizzare il lavoro svolto. Analogamente, tutte e tutti noi usiamo strumenti di ricerca e discovery, e sappiamo bene che ogni strumento di questo tipo utilizza algoritmi di ranking non sempre neutrali o trasparenti; ma di nuovo non sarebbe né possibile né probabilmente desiderabile citare ogni ricerca fatta su Google o Google Scholar (o su altri database bibliografici o discovery tools) nel lavorare su un articolo, o fornire l'insieme delle risposte di volta in volta ottenute. Infine, anche le indicazioni ottenute attraverso il meccanismo di peer review – come quella che sto discutendo in questa sede – indirizzano evidentemente l'elaborazione di un articolo (in questo caso, mi hanno spinto ad aggiungere questa nota e a modificare – spero utilmente – alcuni altri passi del testo): in linea di principio dovrebbero forse essere rese integralmente pubbliche e – dopo la fase di peer review – correttamente e individualmente attribuite. Essendo ormai gestite attraverso piattaforme di mediazione editoriale, la cosa è non solo possibile, ma anche facilmente realizzabile. Tuttavia, di nuovo, quanta parte di questo lavoro di elaborazione è sensato documentare integralmente, considerata l'esigenza di focalizzare l'attenzione del lettore su quel che è centrale, cioè l'articolo come prodotto finale dell'attività di ricerca?

Questo non significa, naturalmente, che il problema della trasparenza nell'uso dei LLM – come di qualunque altra fonte diretta o indiretta, ma ancor più in questo caso, per i caratteri specifici di questo tipo di fonti – non si ponga, e che non sia importante discuterlo, come ho cercato di fare sia nel testo sia in questa nota (anche se una discussione soddisfacente richiederebbe molto più spazio: il tema è assai rilevante, e meriterebbe una sede specifica). Resta fermo, ovviamente, il principio fondamentale di correttezza che impone di riportare ogni fonte diretta utilizzata, e in particolare tutti i testi citati: così, qui come in tutti i miei lavori, ogni citazione diretta o comunque sostanziale dalle risposte di un LLM è sempre chiaramente indicata e attribuita (in questo articolo, come già ricordato, l'unico caso di questo tipo è discusso nella nota 15), e impostazione, tesi generali e contenuto non virgolettato del testo sono miei, e non prodotti da sistemi di IA generativa.

³ Un'ulteriore tematica di grande interesse è quella della catalogazione e metadattazione di contenuti prodotti dall'IA generativa o risultato di una collaborazione con l'IA generativa: ma discuterla in questa sede richiederebbe ancora più spazio di quello, già notevole, richiesto dalla trattazione dei temi su cui ho scelto di concentrare il lavoro.

pone all'uso di sistemi di IA generativa. La seconda sezione propone la distinzione fra tre diverse strategie d'uso dell'IA generativa in ambito bibliotecario e documentale, con il riferimento ad alcune esperienze concrete in cui sistemi di IA generativa sono stati utilizzati in particolare in contesti di catalogazione e metadatazione.

1. Evoluzione dei paradigmi catalografici e di metadatazione: Da MARC ai nuovi framework nel contesto Linked Open Data

Le origini della standardizzazione in campo catalografico risalgono al XIX secolo, con la definizione di regole precise e uniformi per la realizzazione delle schede cartacee, che rappresentavano l'unità fondamentale di descrizione bibliografica. Regole catalografiche ragionevolmente standardizzate vengono progressivamente individuate nel corso dell'800 tanto nel contesto europeo quanto in quello nord-americano, a partire dalle famose 91 regole attribuite ad Antonio Panizzi⁴ e alle indicazioni contenute nei primi 'manuali' bibliotecari, come Namiur (1834). In Italia, il manuale di Giuseppe Fumagalli sottolinea il rapporto fra catalografia e bibliografia descrittiva (Fumagalli 1887, xiii) e nota già chiaramente il carattere semantico del rapporto fra scheda e libro catalogato: "...schede, che sono come i segni rappresentativi dei libri..." (xv).

Il lavoro di affinamento delle pratiche catalografiche su scheda prosegue fino ai Principi definiti nella Conferenza di Parigi del 1961 (Chaplin e Anderson 1963), ma già nel corso della prima metà degli anni '60 diviene chiaro che l'avvento dell'informatica comporta una rivoluzione nel campo della catalogazione: rivoluzione concretizzata nel 1966 con la nascita del formato MARC (Machine-Readable Cataloguing). Sviluppato dalla Library of Congress soprattutto ad opera di Henriette Avram (1975), il formato MARC permette di rappresentare informazioni bibliografiche in maniera non solo standardizzata ma direttamente leggibile e utilizzabile da parte dei sistemi informatici dell'epoca. MARC rappresenta dunque il passaggio dalla scheda cartacea al record digitale, pur mantenendo, almeno inizialmente, una stretta corrispondenza con la struttura e la logica della scheda catalografica tradizionale. I record MARC sono infatti composti da campi, sottocampi e indicatori che codificano i diversi elementi descrittivi (autore, titolo, editore, soggetto...), consentendo la creazione di cataloghi elettronici (OPAC – Online Public Access Catalog) e lo scambio di dati bibliografici tra biblioteche.

Nonostante l'enorme passo in avanti rappresentato da MARC e il suo successo internazionale (con la nascita di diversi formati nazionali, poi parzialmente unificati in UNIMARC), la sua struttura, ancora legata al concetto tradizionale di record catalografico, mostra ben presto i suoi limiti. La rigidità del formato, la difficoltà di estrarre e collegare informazioni specifiche presenti all'interno del record e la mancanza di una semantica esplicita, con la conseguente confusione fra componenti sintattiche e componenti semantiche nella definizione dei record (e a volte l'interferenza fra componenti diverse dello stesso record), rende progressivamente chiara la necessità di nuovi modelli e nuovi standard.⁵ È in primo luogo la geniale tesi di dottorato di Barbara Tillett ad avviare nel 1987

⁴ Le *Rules for the Compilation of the Catalogue* sono del 1839 e aprono il primo volume del *Catalogue of printed books in the British Museum* pubblicato nel 1841 (Panizzi 1841). Sulla genesi e sul contesto storico e teorico delle regole di Panizzi si vedano (Guerrini e Neri 2020; Guerrini e Gambari 2023).

⁵ Per una ricostruzione storica di questo processo, sintetizzato in (Tennant 2002) con lo slogan 'MARC must die', si veda

una riflessione di tipo nuovo, legata allo studio delle relazioni concettuali fra entità bibliografiche (Tillett 1987).⁶ Riflessione che nel decennio successivo si sviluppa soprattutto nell'ambito dell'IFLA Section on Cataloguing e porta all'elaborazione del modello concettuale FRBR, *Functional Requirements for Bibliographic Records*, approvato dalla 63^a Conferenza IFLA di Copenaghen nel 1997 e pubblicato, nella sua prima versione ufficiale, nel 1998 (IFLA 1998).⁷

Nel frattempo, la rapida diffusione di Internet e la nascita del web spingono verso una crescente richiesta di interoperabilità e riuso dei dati, mentre l'idea del semantic web, avanzata alla fine degli anni '90 da Tim Berners Lee, contribuisce a rafforzare la consapevolezza che la standardizzazione delle pratiche descrittive deve basarsi su sistemi classificatori (ontologie) a loro volta concettualmente rigorosi e ben definiti anche dal punto di vista formale. In particolare, dal lavoro sul semantic web nasce il *Resource Description Framework* (RDF), un modello adottato come raccomandazione dal W3C nel 1999 e pubblicato nella sua prima versione ufficiale nel 2004 (la versione 1.1 sarà pubblicata nel 2014).

RDF consente di formalizzare l'attribuzione di metadati descrittivi attraverso l'uso di triple soggetto-predicato-oggetto, in cui il soggetto rappresenta la risorsa da descrivere, il predicato il tipo di proprietà o relazione e l'oggetto il valore assunto da tale proprietà o relazione quando attribuita al soggetto. In un uso documentalmente solido di RDF, soggetto, predicato e oggetto devono essere espressi in maniera rigorosa: ad esempio, se il soggetto fosse l'autore di un libro, lo si dovrebbe identificare univocamente (di norma attraverso il ricorso a un authority file), la proprietà 'autore di' dovrebbe essere presente in una ontologia formale (che deve essere dichiarata), e l'oggetto dovrebbe essere a sua volta espresso in maniera univoca e formalmente corretta, cosa che viene fatta di norma attraverso un Uniform Resource Identifier (URI).

Come si vede, dunque, il framework RDF da solo non basta: si tratta solo di uno dei livelli necessari per costruire un complesso edificio di rappresentazioni formali della conoscenza. E in effetti il lavoro su RDF ha rappresentato la base per la nascita di ontologie formali come il *Web Ontology Language* (OWL), che permette di costruire in maniera rigorosa tassonomie e – in generale – sistemi di classificazione. Nel frattempo, il lavoro sul web semantico si è orientato in direzione dei *Linked Open Data* (LOD), che rappresentano una versione in parte semplificata ma più realisticamente gestibile dell'idea originaria di semantic web.⁸ Il nome scelto sottolinea da un lato l'elemento fortemente relazionale di qualunque attività di descrizione e metadattazione,⁹ dall'altro il requisito di apertura dei dati, che ne permette il riuso e quindi l'integrazione di ontologie e basi di conoscenze parziali e settoriali in costruzioni di più alto livello.

(Thomale 2010). Nonostante l'auspicio di Tennant, MARC sembra comunque ancora lontano dall'essere definitivamente abbandonato: è interessante notare come alcune delle sperimentazioni – anche recenti – di uso dell'IA di cui parleremo nella seconda sezione avvengano proprio su dati in uno dei molti formati derivati da MARC.

⁶ Si vedano (Norouzi 2012) e (Ghiringhelli e Guerrini 2019).

⁷ Non provo a sintetizzare la sterminata letteratura sul tema, ma per una breve introduzione al modello FRBR si può vedere (Galeffi e Sardo 2013), per una valutazione aggiornata del modello e del suo impatto è utile (Coyle 2022), mentre per il quadro concettuale complessivo dell'evoluzione qui delineata punti di riferimento utilissimi sono (Guerrini, Crupi, e Peruginelli 2013; Guerrini e Sardo 2018; Guerrini 2022a; 2022b).

⁸ Per una descrizione introduttiva di queste tematiche e la bibliografia essenziale al riguardo si veda (Roncaglia 2023, 57-66 e, molto più approfondito, (Tomasì 2022).

⁹ Riferimento obbligato in tema è (Zeng e Qin 2022).

Il lavoro di costruzione di livelli descrittivi man mano più rigorosi e formalizzati riguarda anche l'ambito strettamente bibliografico e documentale. Nonostante il deciso passo avanti comportato dall'adozione del modello entità-relazioni, infatti, il riferimento ai *bibliographic records* presente nello stesso acronimo FRBR lasciava ancora trasparire un'idea di dato bibliografico legata al concetto tradizionale di record. Occorreva dunque costruire anche per i dati (e metadati) bibliografici quel complesso edificio di livelli descrittivi formalizzati necessario alla piena adozione del paradigma degli Open Linked Data e di RDF, superando definitivamente la concezione del record come entità principale e arrivando a una rappresentazione basata su dati di livello più basso, interconnessi e semanticamente ricchi. E occorreva definire formalmente, per ciascuno dei diversi 'piani' di questo edificio, il passaggio dal generale modello entità-relazioni su cui è basato FRBR al più specifico modello delle triple RDF. Infatti, se è vero che una tripla RDF viene molto naturalmente concepita come la dichiarazione di una relazione fra due entità, è anche vero che la centralità delle triple – e dunque degli *statement* – RDF rappresenta in un certo senso una specificazione ulteriore. Questo percorso passa attraverso la definizione degli *International Cataloguing Principles* (ICP), pubblicati in una prima versione nel 2009 dopo sei anni di discussioni negli IFLA Meetings of Experts on an International Cataloguing Code, e dello standard *Resource Description and Access* (RDA), del 2010, che fornisce elementi, linee guida e indicazioni specifiche per la creazione di metadati relativi al patrimonio culturale, non solo (ma anche) in ambito bibliografico e documentale.¹⁰ Passa attraverso la progressiva creazione di una vera e propria 'famiglia' di standard FR, in cui a FRBR si affiancano i *Functional Requirements for Authority Data* (FRAD) e i *Functional Requirements for Subject Authority Data* (FRSAD), coordinati poi attraverso il *Library Reference Model* (LRM) sviluppato dall'IFLA FRBR Review Group e pubblicato nel 2017 (IFLA 2017).¹¹ Per quanto riguarda specificamente l'adozione in ambito bibliotecario del paradigma Linked Open data, fondamentale è anche il *Final Report* del W3C Library Linked Data Incubator Group, del 2011.¹²

È in questo contesto – che resta, come si vede, in continuo e progressivo affinamento concettuale – che si sviluppano oggi anche gli standard più strettamente bibliografici e catalografici. ISBD (*International Standard Bibliographic Description*), nato per promuovere il controllo bibliografico universale ma inizialmente assai condizionato dalla sua natura di 'mediazione' fra le diverse pratiche descrittive nazionali, ha visto così la sua portata ampliata per "migliorare la portabilità dei dati bibliografici nell'ambiente del Web Semantico e l'interoperabilità dell'ISBD con altri standard di contenuto" (IFLA 2011). Questo obiettivo è stato perseguito attraverso la pubblicazione di un insieme di elementi ISBD in RDF, inizialmente nell'Open Metadata Registry nel 2015 e successivamente, dal 2020, come Vocabolari ISBD nei Namespaces IFLA.

Fra gli standard catalografici prodotti dalla 'rivoluzione FRBR' un ruolo di primo piano spetta sicuramente a BIBFRAME (*Bibliographic Framework*), sviluppato dalla Library of Congress come potenziale sostituto di MARC. In Italia, sono influenzate da FRBR le *Regole italiane di catalogazione* (REICAT, ripensamento abbastanza radicale delle precedenti RICA). In entrambi i casi, la scel-

¹⁰ Su RDA una utile rassegna è fornita da (Bianchini e Guerrini 2016).

¹¹ Si vedano (Bianchini 2017; Guerrini e Sardo 2018). L'ultima versione consolidata è (IFLA 2024).

¹² W3C Library Linked Data Incubator Group 2011. Sull'uso dei linked data in ambito bibliotecario e documentale si vedano (Baker 2013; Guerrini, Crupi, e Peruginelli 2013; Guerrini e Possemato 2015).

ta di FRBR come punto di riferimento permette una progressiva evoluzione verso l'uso di RDF e LOD e (auspicabilmente) verso il contesto LRM. Il risultato è un cambiamento radicale nel modo di concepire e utilizzare gli standard catalografici: non più regole per la compilazione di record, ma veri e propri modelli concettuali espressi in linguaggi formali che definiscono classi, proprietà, vincoli e relazioni semantiche tra gli elementi descrittivi. Un dato, questo, che – come si vedrà – è estremamente rilevante volendo progettare e utilizzare sistemi di intelligenza artificiale utilizzabili in ambito bibliotecario e documentale.

Va notato peraltro come questa evoluzione renda sempre più angusta l'idea di standard catalografici nazionali o legati a singole istituzioni: strumenti come il *Virtual International Authority File* (VIAF) per gli autori, *GeoNames* per i luoghi, o le ultime versioni del *Library of Congress Subject Headings* (LCSH) e i thesauri LOD per i soggetti, indicano una strada di internazionalizzazione che sembra imposta dalla natura stessa dei sistemi adottati e – soprattutto – dal passaggio dalla catalogazione locale alla metadatazione. Passaggio che deve avvenire all'interno di uno spazio concettuale condiviso e collocato a un più alto livello di astrazione. Una evoluzione di cui dovranno evidentemente tenere conto gli sviluppi futuri di SBN e delle nostre pratiche catalografiche, in particolare nel processo verso l'Indice 3 SBN: ma queste tematiche sono sviluppate, assai meglio di quanto non potrei fare io, in altri contributi a questo fascicolo.

2. Metadati e IA generativa: un rapporto problematico ma promettente

La rapidissima sintesi proposta nella sezione precedente può sembrare forse superflua in un intervento centrato in primo luogo sull'uso di strumenti di IA generativa, ma risponde a una funzione specifica: quella di ricordare che nel mondo della metadatazione bibliotecaria si manifesta ormai da tempo una duplice tendenza. Da un lato, la tendenza alla standardizzazione e alla formalizzazione, nell'ambito di modelli basati su architetture della conoscenza a livello di astrazione sempre più alto. Dall'altro, una tendenza alla granularizzazione dei dati. È bene notare che, anche se queste tendenze sono associate, l'una non implica necessariamente l'altra. E che, nell'incontro con l'intelligenza artificiale generativa, possono avere ruoli almeno in parte diversi.

Per illustrare il rapporto fra strumenti di organizzazione strutturata delle informazioni e delle conoscenze e i nuovi paradigmi dell'IA generativa, ho usato altrove (Roncaglia 2023) la metafora dell'architetto e dell'oracolo. L'architetto corrisponde all'idea di creare sistemi descrittivi e organizzativi rigorosi, formalizzati, architettonicamente solidi. L'oracolo corrisponde invece alla generazione statistico-probabilistica di contenuti propria dei sistemi di intelligenza artificiale generativa basati su reti neurali profonde, in cui l'enorme ragnatela numerica di pesi, valori, funzioni di attivazione e rappresentazione vettoriale di unità minime informative (i token) produce risultati sorprendenti ma è almeno parzialmente oscura, sia per la sua complessità, sia per il suo carattere intrinsecamente non deterministico.

Semplificando in maniera estrema, anche a rischio di una certa imprecisione, nella storia dell'intelligenza artificiale si possono distinguere almeno due grandi fasi. Quella iniziale, fra gli anni '50 e gli anni '90 del secolo scorso, in cui si è lavorato soprattutto partendo da una concezione logico-simbolica e linguistica dell'intelligenza, e durante la quale si è cercato di costruire sistemi artificiali (inizialmente con ambizioni maggiori; man mano – dopo il fallimento del programma

‘forte’ iniziale, divenuto chiaro a metà degli anni ’70 – in ambiti più ristretti e limitati) guidati per lo più da regole deterministiche a loro volta di natura prevalentemente linguistica e logico-simbolica. Poi, negli ultimi decenni e con una qualche sovrapposizione dei due approcci negli anni ’80 e ’90, la fase legata all’uso di reti neurali profonde e alla loro trasformazione in senso statistico-probabilistico, che ha portato ai sistemi di intelligenza artificiale generativa di oggi.¹³ La metafora dell’architetto e dell’oracolo può essere applicata, almeno a grandi linee, anche a questa evoluzione e alle due fasi che la caratterizzano. Ovviamente, come ogni metafora, anche questa va presa *cum grano salis*: il riferimento all’oracolo vuole sottolineare il meccanismo predittivo e non deterministico di generazione, ma non implica affatto che i sistemi generativi siano infallibili, né che siano totalmente inesplicabili. D’altro canto, l’IA logico-simbolica, pur avendo deluso almeno in parte le aspettative iniziali, non è affatto morta, ed è abbastanza probabile che in futuro le due metodologie si intrecceranno sempre più spesso (la cosiddetta strada ‘neurosimbolica’, che molti osservatori vedono come la più probabile evoluzione futura del settore)

In ogni caso, la prima delle tendenze sopra ricordate relativamente all’organizzazione delle informazioni e delle conoscenze, quella alla standardizzazione e alla formalizzazione, è chiaramente architettonica. Potrebbe dunque sembrare, almeno in prima istanza, assai lontana dall’IA generativa e semmai più direttamente compatibile con l’IA logico-simbolica classica, che era basata su regole esplicite e su risultati prevedibili e sicuri. Come pensare, per fare solo l’esempio più ovvio, di affidare una qualsiasi forma di metadattazione standardizzata di contenuti a sistemi che possono produrre – e di fatto producono abbastanza spesso – risposte totalmente errate (allucinazioni), anche e particolarmente in campo bibliografico?¹⁴

D’altro canto, i sistemi di IA basati sul *machine learning* (ML), compresi quelli generativi, amano l’abbondanza di dati granulari, e ancor più di dati granulari affidabili, che risultano particolarmente adatti al loro addestramento. L’esplosione dei big data – di cui anche i LOD fanno a loro modo parte – costituisce anzi una delle precondizioni che spiegano i progressi realizzati negli ultimi anni in intelligenza artificiale. Il problema, però, è che i grandi modelli linguistici (che rappresentano la tipologia di IA generativa più avanzata e – almeno al momento – di più diretto interesse per il contesto discusso in questa sede) non sono addestrati solo o principalmente su dati granulari formalizzati, ma su enormi corpora testuali eterogenei e poco strutturati. Nell’universo di contenuti che alimenta l’IA generativa, i dati bibliografici e in generale i sistemi di dati formalizzati sono una piccola isola, per molti versi felice (si pensi al ruolo dei LOD in Wikipedia e al particolare valore di Wikipedia rispetto alla grande maggioranza delle risorse web), ma del tutto minoritaria.

In questa situazione, un incontro proficuo fra metadattazione bibliotecaria e IA generativa richiede l’adozione di strategie specifiche. Suggerisco di distinguerne tre: 1) creazione di sistemi IA specializzati, addestrati (spesso in forma di *machine learning* tradizionale, più che attraverso le nuove architetture proprie dell’IA generativa attuale) unicamente attraverso dati ben costruiti e standardizzati. 2) collaborazione fra sistemi architettonici tradizionali (ad esempio, una base di dati

¹³ Non è ovviamente possibile ripercorrere in questa sede la storia dell’evoluzione dell’intelligenza artificiale; mi limito a ricordare, fra i testi di riferimento al riguardo, il classico (Russell e Norvig 2020). Ricostruzioni di questo processo più dettagliate di quella possibile in questa sede sono in (Roncaglia 2023) e in (Roncaglia 2025b).

¹⁴ Sulle allucinazioni bibliografiche rimando a (Roncaglia 2025a).

strutturata e affidabile) e sistemi generativi. È la strada, in particolare, della *Retrieval Augmented Generation* (RAG); 3) costruzione di sistemi generativi che, pur conservando la flessibilità caratteristica dei grandi modelli linguistici, siano indirizzati, attraverso una o più fra le molte strade esplorate negli ultimissimi anni (e talvolta negli ultimi mesi: uso di tecniche avanzate di prompting e di ragionamento, selezione di un contesto adeguato, *deep research*, *fine tuning* specifico, capacità agentiche...), verso la produzione e/o la gestione di metadati standardizzati e affidabili.

In estrema sintesi, le tre strategie sopra delineate *rappresentano altrettanti tentativi di far dialogare i due approcci apparentemente antitetici dell'architetto e dell'oracolo. La prima cerca di 'addomesticare' l'IA confinandola in un ambito strutturato: per garantire l'obiettivo del rigore formale depotenzia o fa scomparire l'oracolo. La seconda cerca di integrare IA generativa e sistemi strutturati, e rappresenta una forma di collaborazione diretta fra l'architetto e l'oracolo, lasciando all'uno e all'altro le proprie caratteristiche. La terza cerca di guidare l'IA generativa verso la produzione di strutture: in sostanza cerca di spingere l'oracolo ad acquisire anche le competenze dell'architetto.*¹⁵

La distinzione di queste tre strategie va intesa, è bene chiarire subito, come assolutamente provvisoria e funzionale unicamente a fornire un primo schema espositivo a una tematica in rapidissima evoluzione. È del tutto possibile (e per certi versi anche probabile) che fra queste strategie siano possibili ibridazioni e contaminazioni, così come è possibile (e anzi probabile) la nascita di prospettive del tutto nuove. Inoltre, come si vedrà in conclusione, le prospettive aperte dai modelli locali di piccole e medie dimensioni possono contribuire a cambiare il quadro di riferimento e suggerire nuove e diverse sperimentazioni. Trattandosi di sviluppi assai recenti, è dunque probabilmente prematuro tentare un bilancio: vale la pena notare che un testo collettaneo e recentissimo che si propone come punto di riferimento sul tema (Balnaves et al. 2025) include contributi certo interessanti ma corrispondenti a strategie e casi di ricerca abbastanza eterogenei, dai quali non emergono ancora né un consenso su modelli e situazioni d'uso, né una chiara visione dell'IA bibliotecaria nella nuova era dell'IA generativa. Tuttavia, siccome un qualche ordine – pur se provvisorio e parzialmente arbitrario – è comunque preferibile al disordine totale, proverò a seguire la tripartizione sopra delineata, fornendo per ciascuna delle tipologie qualche indicazione aggiuntiva e qualche esempio.

2.1. IA specializzate e non-oracolari

La prima strategia si collega per certi versi all'eredità dell'intelligenza artificiale classica, e non a caso corrisponde in primo luogo a sperimentazioni che precedono l'introduzione dell'intelligenza artificiale generativa nel senso contemporaneo del termine, basata su architetture a Transformer, e la sua rapida diffusione: uno sviluppo avviato nel 2017 e accelerato dopo il rilascio pubblico di ChatGPT nel 2022.¹⁶

¹⁵ Il contenuto di questo specifico paragrafo di testo mi è stato suggerito il 2/8/2025 da Kimi K2 come possibile sintesi delle tre strategie, sulla base di un prompt in cui chiedevo al sistema di proporre critiche e controesempi alla tripartizione da me proposta. Mi è sembrata una sintesi particolarmente felice, e – anche se ho parzialmente riformulato o modificato il testo prodotto dall'IA (che per questo motivo ho marcato in corsivo anziché con l'uso delle virgolette) – si tratta dell'unico caso in cui un passo di questo articolo è stato effettivamente prodotto, nelle sue linee portanti, da un sistema di IA generativa.

¹⁶ Per una ricostruzione storica di questi sviluppi si veda (Roncaglia 2025b).

Ricorderò qui solo tre casi interessanti, che corrispondono a tre epoche diverse dello sviluppo dell'IA pre-Transformer. Il primo è rappresentato dall'idea di classificazione automatica dei documenti come naturale prosecuzione del lavoro puramente algoritmico di sviluppo di concordanze e indici: un lavoro avviato assai presto, a partire dalla pionieristica esperienza dell'*Index Thomisticus* di Padre Busa.¹⁷ Oggi sembra difficile considerare lavori di questo tipo come pertinenti all'intelligenza artificiale, ma all'epoca si riteneva che il passaggio dal semplice ordinamento di lemmi e occorrenze a forme di indicizzazione più astratta potesse avvenire attraverso algoritmi relativamente semplici di analisi statistica e fosse quasi a portata di mano. Lo mostra chiaramente una citazione del lontano 1963: "Raw data, i. e. , unedited natural language text, can be processed statistically so as to automatically assign index terms to each document and to classify the document into a subject category; this has been demonstrated" (Borko e Bernick 1963, 28).¹⁸ Nonostante il notevole lavoro svolto sul tema negli anni '60 e '70 del secolo scorso,¹⁹ anche in questo caso – come in quello più generale dell'AI logico-simbolica – l'ottimismo iniziale non sembra portare a strumenti effettivamente efficaci ed usabili: non a caso, anche questo indirizzo di ricerca sarà di fatto ripreso solo nel nuovo millennio attraverso le nuove metodologie offerte dal *machine learning*.²⁰

In sostanziale coincidenza con il periodo in cui l'IA logico-simbolica abbandona l'iniziale fiducia sulla possibilità di costruire sistemi 'generalisti' e si concentra su ambiti più limitati, anche il focus relativo alle sue applicazioni bibliotecarie si sposta su sistemi esperti specifici, in grado di facilitare le ricerche degli utenti: i primi sviluppi in questo campo sono degli anni '80 e '90 del secolo scorso.²¹ Un esempio emblematico è l'*Intelligent Reference Information System (IRIS) Project* sviluppato dalle biblioteche dell'Università di Houston tra il 1988 e il 1993. Il progetto IRIS ha portato a un prototipo e successivamente a un sistema di reference che assisteva gli utenti nella selezione degli indici e nell'accesso a reti LAN (ad esempio per l'accesso a CD-ROM),²² il cui codice aperto, in Prolog, è stato distribuito anche ad altre biblioteche.

Con il nuovo millennio l'attenzione si sposta nuovamente, passando dai sistemi esperti settoriali a soluzioni basate su basi di dati accessibili via web. In questo caso, un buon esempio è il sistema esperto per la selezione di database sviluppato nella biblioteca dell'università dell'Illinois da Wei Ma e Timothy Cole fra il 2000 e il 2002.²³ Il sistema si basava su un *selector* che, assumendo il ruolo di un bibliotecario di reference, proponeva all'utente i database più adeguati alla ricerca che si voleva svolgere. L'idea era quella di "analyze and then emulate in software the techniques most

¹⁷ Sulla figura di Padre Busa e sul progetto dell'*Index Thomisticus* si veda (Jones 2018).

¹⁸ Harold Borko, in particolare, ha fornito negli anni '60 e '70 del secolo scorso moltissimi contributi sul tema. Per rendersi conto di quanto poco questo lavoro pionieristico sia ricordato oggi, merita notare che nel momento in cui scrivo solo le versioni francese, spagnola e catalana di Wikipedia gli dedicano una voce, assente invece nella versione inglese (nonostante Borko sia statunitense, abbia lavorato tutta la vita negli Stati Uniti e sia stato presidente della American Society for Information Science).

¹⁹ Ne è buona testimonianza (Stevens 1965), assolutamente impressionante per la mole di riferimenti a linee di ricerca affascinanti e oggi quasi dimenticate, alcune delle quali basate peraltro su metodologie statistiche che conservano un notevole interesse. Per un'idea di queste metodologie, si può vedere ad es. (Heaps 1973).

²⁰ Un esempio che rappresenta un po' un ponte fra le esperienze precedenti e gli sviluppi successivi è fornito da (Sebastiani 2002).

²¹ Per una bibliografia in merito si veda (Bailey 2018).

²² Per una descrizione, si veda (Bailey 1990).

²³ Su questa esperienza cfr. (Ma e Cole 2001; Ma 2002).

commonly used by reference librarians when helping library users identify databases to search. A key component of this approach has been the inclusion in the system's index of database characteristics the controlled vocabulary terms used by most of the databases characterized" (Ma e Cole 2001, 282). Un sistema, dunque, basato sull'uso guidato di vocabolari controllati.

L'epoca successiva è quella del *machine learning*, che porta all'IA generativa attuale. La meta-ricerca di Islam et al. (2025) mostra come i riferimenti a progetti di IA bibliotecaria all'interno della letteratura specializzata crescano notevolmente nel periodo 2009-2018, prima dunque della rivoluzione dei Transformer, delineando un panorama in cui la presenza della Cina diventa maggioritaria rispetto alle pubblicazioni e tende progressivamente a crescere anche rispetto alle citazioni. Una buona rassegna, basata sulla situazione pre-Transformer, è fornita dal report *Machine Learning + Libraries* di Ryan Cordell (2020).

In questa fase si cominciano a usare intensivamente le reti neurali profonde, ed emergono per la prima volta i problemi legati alle allucinazioni: si affaccia dunque l'oracolo – anche se non ancora con le sue capacità attuali – e per limitarne gli effetti si lavora soprattutto con reti addestrate attraverso corpora controllati e con meccanismi di verifica umana (human-in-the-loop²⁴). Inizialmente, però, l'obiettivo non è tanto la produzione di metadati o la generazione di testi quanto il lavoro in ambiti in cui le reti neurali pre-Transformer (e dunque pre-LLM) sembrano eccellere, come il riconoscimento dei caratteri e delle grafie, inclusi il recupero di annotazioni manoscritte e l'uso di OCR su testi 'difficili' (molti programmi e progetti di trascrizione automatica o assistita da manoscritti – ad es. il software Transkribus²⁵ – si basano su sistemi di questo tipo).

A questa fase risale anche l'*IFLA Statement on Libraries and Artificial Intelligence* (IFLA 2020). Un documento di grande interesse, che per certi versi prefigura tematiche assolutamente attuali, a partire dal lavoro sull'AI Literacy, che costituisce oggi la parte probabilmente più importante e delicata dell'information literacy. Corretta pare anche l'identificazione delle possibilità di estrema personalizzazione dei contenuti come una delle conseguenze principali dei nuovi sistemi di IA, così come la sottolineatura del rilievo delle considerazioni etiche e della consapevolezza dei bias che possono essere presenti (e di norma sono presenti) nei corpora di addestramento e nell'output di questi sistemi. D'altro canto, diversi aspetti dell'analisi sono già datati, riferendosi alla situazione pre-LLM.²⁶

Sistemi più direttamente orientati all'ambito classificatorio e alla generazione automatizzata di metadati – ma non ancora basati su architetture generative a Transformer – si affacciano fra il 2020 e il 2022. Un buon esempio è Annif, toolkit realizzato dalla biblioteca nazionale finlandese e distribuito in accesso aperto.²⁷ Annif è un sistema di indicizzazione e classificazione bibliotecaria che riprende e affida al nuovo paradigma del *machine learning* la vecchia idea di indicizzazione automatica. Nel frattempo, si approfondisce l'uso del LOD come dataset per realizzare in *machine learning* applicazioni di generazione di metadati e annotazioni: uno sviluppo ben esemplificato dal

²⁴ L'espressione è usata in questo caso con riferimento alla partecipazione di operatori umani nella fase di verifica dei risultati forniti dal sistema. In un senso più generale, l'espressione 'human-in-the-loop' è molto usata anche nelle discussioni di etica dell'IA: al riguardo, si veda (De Caro e Giovanola 2025).

²⁵ Il sito di riferimento è <https://www.transkribus.org/>.

²⁶ Un aspetto giustamente sottolineato da (Lana 2023).

²⁷ Per maggiori informazioni si veda (Suominen et al. 2022).

progetto SAGE, che lavora sui dati di Europeana usando metodologie di *Natural Language Processing* (NLP) e *machine learning*, con l'obiettivo di "generating, enriching, validating, publishing, and searching RDF data" (Kaldeli et al. 2024, 358).

Progetti basati sul *machine learning* a partire da dati controllati sono anche quelli dell'OCLC (2025), focalizzato sulla scoperta di duplicati catalografici e addestrato anche attraverso segnalazioni degli utenti, standardizzate attraverso la maschera di segnalazione, o alcuni fra i progetti della British Library, come Languid, basato sull'uso di sistemi ML per l'identificazione automatica della lingua dei documenti, finalizzato all'inserimento del dato nei record catalografici MARC (Morris 2019).

Un altro esempio di iniziativa 'ponte' verso la situazione attuale, in cui si affaccia la generazione di metadati, è il progetto *Exploring Computational Description* (ECD) della Library of Congress, avviato nel 2022.²⁸ L'obiettivo è valutare l'efficacia dei modelli di *machine learning* nella generazione di record bibliografici di alta qualità su larga scala. Il progetto si concentra sull'utilizzo di reti neurali addestrate su un vasto corpus di dati bibliotecari, composto da alcune decine di migliaia di libri ed ebook e dai relativi record MARC, che fungono da *ground truth* per l'addestramento e la valutazione dei modelli (sono stati testati 5 diversi modelli, presumibilmente basati su architetture di rete neurale diverse, anche se la documentazione che ho potuto reperire non fornisce indicazioni specifiche al riguardo).

Le metodologie impiegate nel progetto ECD includono in primo luogo la classificazione del testo, utilizzata per suggerire i *Library of Congress Subject Headings* (LCSH). In questo caso, un sistema di *machine learning* analizza il testo degli ebook per identificare gli argomenti principali e li traduce in termini di soggettazione LCSH. Questo approccio consente di automatizzare una parte complessa e dispendiosa in termini di tempo del processo di catalogazione, fornendo suggerimenti basati su un vocabolario controllato e che i catalogatori possono poi rivedere e affinare. In secondo luogo, viene sperimentato un sistema di classificazione dei token che ha lo scopo di suggerire i nomi degli autori, come stabiliti nel *Library of Congress Name Authority File* (NAF). In questo caso, il modello di ML è addestrato per riconoscere e prevedere con precisione le stringhe di caratteri che costituiscono il nome dell'autore all'interno delle pagine del titolo, facilitando l'associazione corretta con le forme standardizzate dell'authority file.

I limiti di questa prima strategia, legata all'uso di AI classica o di *machine learning* basato su corpora controllati, sono fondamentalmente due. In primo luogo, anche se i corpora di partenza sono strettamente standardizzati, nei modelli più recenti basati su reti neurali profonde il rischio di allucinazioni non è totalmente evitato: si tratterà di allucinazioni che il sistema avrà cura di inserire nella struttura formale utilizzata – ad esempio, l'uso di un soggetto sbagliato, ma presente nel vocabolario controllato adottato e collocato nel luogo corretto di un record catalografico – e che però proprio per questo possono essere particolarmente difficili da individuare. In secondo luogo, in sistemi di questo tipo manca l'enorme flessibilità garantita dall'uso del linguaggio naturale: per quanto desiderabile e preziosa sia la formalizzazione, la costruzione di una sorta di recinto chiuso attorno a dati e metadati bibliografici potrebbe non essere particolarmente desiderabile, soprattutto in una prospettiva di integrazione dell'AI in uno spettro più vasto di funzionalità bibliotecarie.

²⁸ La descrizione che segue è basata su (Saccucci e Potter 2025) e su altri materiali – incluso il video di una conferenza di presentazione – disponibili in (Saccucci e Potter 2024a; 2024b).

Il primo problema è tipicamente affrontato attraverso l'enfasi sul controllo umano (*Human-In-The-Loop*), che però è spesso assai 'costoso' in termini di impegno delle risorse umane, e non è particolarmente gratificante: passare il proprio tempo a correggere errori nei metadati prodotti dall'IA sembra essere un compito assai più faticoso e alienante della produzione diretta degli stessi metadati da parte del bibliotecario, anche se alla fine i risultati sono quantitativamente di più e qualitativamente più ricchi. Il secondo problema non è superabile senza il ricorso a un LLM. Non sorprende, dunque, che negli ultimissimi anni – tendenzialmente, dopo il 2022 – l'attenzione si sia spostata sui risultati spettacolari ottenuti dalle architetture a Transformer e dai grandi modelli linguistici.

2.2. Tentativi di collaborazione: la RAG

Arriviamo così alla seconda fra le strategie sopra delineate, che prevede una collaborazione fra LLM (ma eventualmente anche sistemi generativi multimodali) e basi di conoscenze esterne affidabili e controllate. Tipicamente, questa collaborazione sfrutta la base di dati 'tradizionale' – e dunque l'architetto – per svolgere ricerche ed estrarre risultati, mentre usa l'IA generativa da un lato per trasformare l'input in linguaggio naturale in una query strutturata, e dall'altro per riformulare in linguaggio naturale e se necessario tradurre e personalizzare la risposta. Si tratta del modello corrispondente alla cosiddetta *Retrieval Augmented Generation*.²⁹

La RAG opera attraverso due fasi principali: il recupero (Retrieval) e la generazione (Generation). Nella fase di recupero, una query dell'utente viene utilizzata per interrogare una base di conoscenza esterna (ad esempio un database di documenti, un catalogo bibliotecario, un archivio digitale) al fine di identificare e recuperare i contenuti informativi più pertinenti. Questi contenuti divengono 'documenti contestuali' e vengono passati, insieme alla query originale, a un LLM. Nella fase di generazione, il LLM utilizza sia la query che i documenti recuperati per formulare una risposta coerente, informativa e basata sulle evidenze. Questo processo assicura che la risposta generata sia non solo fluida e naturale, ma anche accurata e supportata da fonti verificabili. Inoltre, l'uso di un LLM garantisce enormi possibilità di personalizzazione: il sistema può rispondere automaticamente nella lingua dell'utente, può adattarsi a registri linguistici e competenze diverse, può fornire sintesi, può usare – nel caso di sistemi multimodali – anche codici comunicativi diversi, ad esempio producendo una risposta in audio, una sintesi in forma di podcast e, in un futuro non troppo lontano, anche di video (ad esempio un documentario).

L'applicazione più naturale di sistemi RAG in ambito bibliotecario è evidentemente per il reference: un esempio tutto italiano è Alphy, il chatbot realizzato per il portale Alfabetica (IPaC 2024), ma le applicazioni di questo tipo, nonostante la novità dell'impostazione, sono già numerose, non solo in ambito bibliotecario: un buon esempio relativo all'ambito archivistico è ad es. Archibot 3.0, il chatbot per l'interrogazione in linguaggio naturale degli archivi storici e legislativi del parlamento europeo (Kimaïd 2024). Strumenti simili, basati su chatbot e RAG, sono stati realizzati o sono in via di realizzazione in diverse realtà bibliotecarie: fra i molti esempi che si potrebbero fare, il progetto "Enriching Collection Access and Use with Generative AI" delle Northwestern

²⁹ Una rassegna recente sulle metodologie RAG, con ulteriore bibliografia, è (Gan 2025).

University Libraries (Northwestern University Libraries 2025, con un video dimostrativo). Una rassegna utilissima, accompagnata da una discussione più approfondita degli usi della RAG in ambito bibliotecario, è in (La Gorga, Testa, e Verna 2025)³⁰.

Progressivamente, l'uso di LLM integrati da basi di conoscenze affidabili si è esteso negli ultimi anni anche alla generazione di metadati. I primi esperimenti affidati unicamente a LLM (in particolare nel biennio 2023-2024) si erano infatti scontrati subito con il problema delle allucinazioni,³¹ e il ricorso alla RAG – assieme all'enfasi sul ruolo del controllo umano, con il già ricordato paradigma Human-In-The-Loop – si è ben presto imposto come una delle strade migliori per limitare il problema. Così, ad esempio, il progetto “Improving Access to Digital Collections Using GenAI in Libraries” dell'Ontario Council of University Libraries (OCUL) e da Scholars Portal, che mira a migliorare la qualità dei metadati e dunque la reperibilità di quasi 50.000 documenti governativi distribuiti in forma aperta sull'Internet Archive (molti in forma di semplici scansioni), utilizzando OCR integrati da LLM per la trascrizione dei testi e la generazione dei metadati tratti da thesauri controllati.

Architetture RAG e/o sistemi di ML addestrati su dati specifici sono molto probabilmente alla base anche di molti fra i servizi AI-based offerti al momento dalle grandi aziende attive nel settore della mediazione informativa e dell'editoria accademica, anche se spesso i progetti di questo tipo non forniscono informazioni complete sui modelli utilizzati. Molte applicazioni di questo tipo sono legate alla fase di reference e discovery, ma non mancano quelle legate alla generazione automatica o assistita di metadati, come l'AI assistant presente nell'editor di metadati ALMA di Clarivate – Ex Libris.³² Una rassegna completa delle offerte di questo tipo esula però dai limiti di questo articolo e si scontrerebbe probabilmente con notevoli problemi di trasparenza, fattore su cui la comunità bibliotecaria deve naturalmente riflettere: la selezione dei servizi e degli strumenti di intelligenza artificiale da usare in questo campo non dovrebbe infatti prescindere da una informazione ragionevolmente corretta e completa sulle metodologie, sulle architetture e sui corpora di addestramenti utilizzati.

2.3. Nuovi sviluppi

Anche se l'architettura RAG sembra per molti versi unire il meglio dei due mondi, non è affatto scontato che rappresenti l'unica strada, o la strada necessariamente migliore, per l'incontro fra l'IA generativa e le esigenze di standardizzazione e validazione proprie del contesto bibliotecario. La terza fra le strategie qui delineate propone una strada diversa: costruire sistemi di IA generativa capaci di imparare a rispondere alle esigenze dell'architetto molto meglio di quelli attuali. Certo, i primi LLM presentavano (e continuano a presentare) enormi problemi di allucinazioni e di ragio-

³⁰ Ringrazio un revisore anonimo per la segnalazione di questo articolo, che non era ancora disponibile al momento della prima stesura di questo lavoro.

³¹ Diversi esempi curiosi – molti dei quali, anche in questo caso, legati al lavoro con record catalografici in formati MARC – sono in (Deng 2023).

³² La pagina di riferimento per le applicazioni AI offerte da Clarivate-Ex Libris è <https://exlibrisgroup.com/ai-we-can-use-and-trust/>, che però non fornisce informazioni specifiche sui diversi sistemi (così come non le offre, se non in minima parte, il report ExLibris 2023).

namento: problemi, come si è detto, legati ai meccanismi statistico-predittivi di generazione, che dipendono da una modellizzazione generale degli usi linguistici e non dal recupero e dalla verifica puntuale di informazioni validate. Ma gli sviluppi più recenti suggeriscono strade nuove per ‘agganciare’ alla realtà le predizioni dell’oracolo. Anche il progressivo affinamento delle capacità di programmazione (*coding*) – compito evidentemente ‘architetonico’ e basato sul rispetto rigoroso di regole – di molti modelli generativi attuali mostra che si tratta di una strada in linea di principio praticabile. Vediamo dunque sinteticamente alcune (per l’esattezza, sei) fra le innovazioni che possono risultare più rilevanti rispetto a questa prospettiva.

In primo luogo, va notato il progressivo allargamento della ‘finestra di contesto’ (*context window*) con cui possono lavorare i modelli. La finestra di contesto è la quantità di informazione che un modello può tenere presente contemporaneamente nel generare contenuti: in sostanza, lo spazio informativo al cui interno può distribuire la sua attenzione. Inizialmente, questa finestra era piuttosto ristretta: le poche righe di un prompt. Progressivamente, con il miglioramento dei modelli, la finestra si è allargata, e questo permette, oltre a prompt più lunghi e articolati, di ‘allegare’ al prompt materiali specifici che vogliamo siano tenuti presenti nel generare la risposta (ad esempio, testi da tradurre o sintetizzare). Nel momento in cui scrivo, Gemini 2.5 Pro – il modello di Google e uno dei sistemi con la finestra più ampia – permette ad esempio di utilizzare un contesto di oltre un milione di token, e dichiara che presto il contesto disponibile salirà a due milioni di token. Finestre di contesto ampie, accompagnate da metaprompt adeguati, permettono di usare il contesto stesso come strumento di ‘ancoraggio’ (*grounding*) ai fatti. È una strategia già praticabile (Roncaglia 2024), ma ha diversi limiti: per quanto ampi, i contesti sono comunque piuttosto angusti per istituzioni con un ampio patrimonio, la gestione di contesti con aggiornamento dinamico non è banale, e le allucinazioni – anche se assai più rare – restano comunque possibili soprattutto su interazioni non direttamente legate al contenuto del contesto (contenuto che l’utente di norma non ha modo di conoscere in dettaglio).

In secondo luogo, una prospettiva potenzialmente interessante è legata allo sviluppo dei sistemi multimodali. In questo testo mi sono concentrato soprattutto sui modelli linguistici, i più interessanti dal punto di vista bibliotecario, ma come sappiamo vi sono anche sistemi generativi che producono immagini, suoni, video... I sistemi multimodali di ultima generazione non si limitano a far lavorare insieme più sistemi monomodali, ma sono addestrati direttamente su corpora multimodali. Normalmente si pensa ai ‘modi’ in termini di codici comunicativi diversi, ma esistono già sistemi che lavorano su modi meno ovvi, ad esempio le temperature o il grado di inquinamento di un ambiente. È possibile considerare come ‘modi’ anche alcune tipologie di dati standardizzati, e usare le tecnologie proprie dei sistemi multimodali per integrare modi con maggior grado di libertà (come il linguaggio naturale) e modi con minor grado di libertà, come i dati bibliografici? Sullo sfondo, si parla anche di *world models*: modelli del mondo. In teoria, se riuscissimo ad addestrare un modello su molte tipologie diverse di dati, che corrispondono alle diverse tipologie di input percettivi umani (e non solo), potremmo pensare di produrre modelli con la capacità di comprendere il mondo in maniera simile alla nostra, o addirittura più ricca, e quindi con una capacità simile alla nostra nella distinzione fra contesti liberi e creativi e contesti vincolati e strutturati. Creare *world models* non è certo facile, e probabilmente non sarebbe banale neanche mettersi d’accordo su cosa esattamente dovrebbe conoscere e comprendere un modello per diventare ‘modello del mondo’. Ma anche questa linea di ricerca ha un ovvio, ed enorme, interesse.

Una terza categoria di innovazioni è legata allo sviluppo dei cosiddetti *reasoning models*. Questi modelli funzionano generando, prima della risposta vera e propria, una descrizione del processo di ragionamento che viene seguito per produrla. E sono sottoposti a un addestramento specifico che riguarda proprio la qualità di questo ragionamento: valutatori umani (e a volte reti generative avversarie) ‘premano’ i ragionamenti più funzionali. Il testo del ragionamento aiuta poi a sua volta il modello a produrre risposte di qualità migliore. Già oggi, quasi tutti i LLM di livello più alto permettono di usare – accanto alle risposte secche (modalità *one shot*) – anche la modalità *reasoning*. Anche in questo caso, si può senz’altro pensare di addestrare un modello all’uso di un *reasoning* orientato alla precisione, alla struttura, alla validazione... insomma all’insieme delle caratteristiche proprie di quello che potremmo definire ‘reasoning documentale’.

La quarta innovazione da ricordare è rappresentata dalla *deep research*, anche in questo caso una caratteristica già presente nei sistemi più avanzati. Si tratta di un passo ulteriore rispetto ai sistemi che cercano il *grounding* attraverso una semplice ricerca in rete. Sappiamo che le informazioni in rete non sono certo un database certificato e affidabile, anzi, possono contenere a loro volta errori marchiani; per questo la RAG usa basi di conoscenze affidabili e non una semplice ricerca in rete. Tuttavia, i sistemi più recenti non si limitano a ‘consultare’ la rete ma operano quella che in gergo è detta appunto ‘ricerca profonda’ (*deep research*): prendono in esame siti diversi, ‘ragionando’ sulla loro maggiore o minore affidabilità e confrontando fra loro i risultati ottenuti. I sistemi di *deep research* sono abbastanza costosi in termini computazionali, perché devono costruirsi un contesto composto da molte pagine di informazioni, selezionandole e valutandole man mano. Anziché rispondere immediatamente, richiedono di norma una decina o anche una ventina di minuti di lavoro. Ma i risultati che producono sono di qualità nettamente superiore, e in certe situazioni potrebbero avvicinarsi alla qualità della RAG, senza la necessità di costruire architetture esterne complesse. In particolare, ciò potrebbe avvenire (ed è questa la quinta delle sei innovazioni qui discusse) attraverso l’uso di sistemi con capacità agentiche, capaci cioè non solo di rispondere a prompt, ma di compiere autonomamente sequenze di azioni complesse. Sistemi di questo tipo possono svolgere ricerche su basi di conoscenze affidabili (ad esempio si un OPAC) senza bisogno di un’architettura RAG specifica.

Un sesto e ultimo sviluppo da ricordare è il miglioramento dei modelli (inclusi i modelli linguistici) di dimensioni piccole (*Small Language Models*: SLM) e medie (*Medium Size Language Models*: MSLM). Di per sé, non si tratta di una novità: la prima versione di Chat GPT, sviluppata da OpenAI nel 2018, l’anno dopo il già ricordato articolo sull’attenzione, aveva ‘solo’ 117 milioni di parametri. Con i criteri di oggi, si tratterebbe di un modello piccolissimo. In linea di massima, nel caso dei modelli linguistici oggi parliamo di modelli piccoli (SLM) sotto i 20 miliardi di parametri, di modelli di medie dimensioni fra i 20 e i 200 miliardi di parametri, e di modelli di grandi dimensioni sopra i 200 miliardi di parametri; ma sono criteri relativi, che potrebbero mutare in futuro. Abbiamo davvero bisogno di tutti i miliardi di parametri dei modelli più grandi? Poter usare modelli di piccole e medie dimensioni avrebbe il vantaggio di poterli installare su computer locali, senza doversi necessariamente affidare (e senza dover necessariamente affidare i nostri prompt, i nostri contenuti, le nostre ricerche) ai grandi fornitori su cloud. Tanto più che nel frattempo abbiamo cominciato a produrre computer più veloci, con processori ottimizzati per il tipo di calcoli richiesti dall’IA generativa. Negli ultimissimi anni, molto lavoro è andato proprio in questa direzione. Diverse aziende (fra i risultati migliori, quelli prodotti da alcune aziende cinesi) hanno rea-

lizzato e distribuito buoni modelli di medie e piccole dimensioni, con prestazioni a volte migliori dei modelli di grandi dimensioni delle generazioni precedenti. C'è però un elemento da tener presente: almeno finora, per avere un buon modello di piccole o medie dimensioni dobbiamo 'distillarlo' da un modello di grandi dimensioni, ad esempio riducendo il numero di cifre decimali dopo la virgola nei parametri, o con altre tecniche che permettono di 'potare' – *pruning* – il modello di partenza. I modelli di piccole e medie dimensioni rappresentano una strada molto interessante, ma per ora non permettono davvero di risparmiare le risorse – come si è visto, enormi – necessarie ad addestrare modelli di grandi dimensioni. Tuttavia, possono permettere di utilizzare in locale tecniche di *grounding* attraverso il cosiddetto *fine tuning*: al modello originale viene aggiunto una sorta di mini-modello locale costruito attraverso dati standardizzati e affidabili, che 'indirizza' le risposte complessive verso una maggiore precisione. Inoltre, l'uso di modelli locali avrebbe il notevole vantaggio di conservare in sede locale dati sensibili (ad esempio, dati sull'utenza o contenuti sotto copyright), aspetto sicuramente desiderabile in un contesto in cui i big player usano in maniera abbastanza disinvolta la loro posizione dominante per fare incetta di informazioni con politiche d'uso non sempre trasparenti (e, in prospettiva, non sempre rispondenti alle indicazioni dell'*AI Act* europeo).

Le innovazioni qui considerate sono troppo giovani per essere valutate attraverso concreti casi d'uso in ambito bibliotecario o documentale,³³ ma è probabile che nel prossimo futuro almeno alcune di queste possibilità (e magari altre) saranno esplorate anche nel nostro mondo.

Una considerazione conclusiva va fatta infine sulle tipologie di dati e metadati prodotti nelle prime sperimentazioni in ambito bibliotecario. In molti dei casi sopra ricordati, il lavoro riguarda ambiti specifici e, quando si lavora su dati catalografici, vengono spesso usati dati concettualmente 'vecchi': ho già sottolineato la frequente presenza nelle sperimentazioni fatte con IA generativa di dati catalografici in formati ricavati da MARC (UNIMARC. MARC 21 ecc.). Certo, MARC è indubbiamente un buon esempio di organizzazione strutturata delle informazioni, e certo, come si è detto, i formati MARC sono ancora assai diffusi; ma come si è visto nella prima parte dell'articolo, sono ormai da tempo concettualmente superati. Sarebbe bene che se ne tenesse conto, anche e soprattutto nel lavoro con l'IA generativa. Tanto più che questi sistemi, se utilizzati con il *reasoning*, sono probabilmente più adatti a gestire il livello di astrazione concettuale tipico dei modelli entità-relazioni e di RDF. Applicare modelli IA di punta su strutture di dati concettualmente obsolete potrebbe non essere – soprattutto nel medio e lungo periodo – una strategia particolarmente produttiva. A ulteriore conferma dello stretto legame che deve necessariamente esistere tra modelli concettuali di metadattazione e descrizione dell'universo informativo, e strumenti – anche generativi – per la produzione e gestione di tali dati.

³³ Includo però il riferimento a un esempio recentissimo che mi sembra particolarmente interessante: (D'Souza et al. 2025). In questo caso, un team di partecipanti ha lavorato presso il TIB Leibniz Information Centre for Science and Technology di Hannover su un progetto di classificazione automatica per il TIBKAT, il catalogo della Technische Informationsbibliothek, la biblioteca nazionale tedesca per i campi dell'ingegneria, tecnologia e scienze naturali. Il team si è diviso lo studio di possibili strategie diverse d'uso dei LLM nella generazione dei metadati, incluse diverse fra le strategie qui ricordate.

Riferimenti bibliografici*

Avram, Henriette D. 1975. *MARC. Its History and Implications*. Washington, DC: Library of Congress.

Bailey, Charles W., Jr. 1990. "Building Knowledge-Based Systems for Public Use: The Intelligent Reference Systems Project at the University of Houston Libraries." In *Convergence: Proceedings of the Second National Conference of the Library and Information Technology Association, October 2-6, 1988*, a cura di Michael Gorman, 190-4. Boston: American Library Association.

Bailey, Charles W. 2018. "Artificial Intelligence in Libraries in the Late 1980's and Early 1990's." *DigitalKoans*. <https://digital-scholarship.org/digitalkoans/2018/08/02/artificial-intelligence-in-libraries-in-the-late-1980s-and-early-1990s/>.

Baker, Thomas, 2013. "Designing data for the open world of the web." *JLIS.it* 4 (1): 63-6. <https://doi.org/10.4403/jlis.it-6308>.

Balnaves, Edmund, Leda Bultrini, Andrew Cox, e Raymond Uzwyshyn, a c. di. 2025. *New Horizons in Artificial Intelligence in Libraries*. Berlin, Boston: De Gruyter Saur.

Bianchini, Carlo. 2017. "Remarks about IFLA Library Reference Model." *JLIS.It* 8 (3): 86-99. <https://doi.org/10.4403/jlis.it-12416>.

Bianchini, Carlo, e Mauro Guerrini, a c. di. 2016. "RDA, Resource Description and Access: The metamorphosis of cataloguing." *JLIS.it* 7 (2). <http://jlis.it/index.php/jlis/issue/view/15>.

Borko Harold, e Myrna Bernick. 1963. "Automatic document classification." *J. Assoc. Comput. Mach* 10: 151-62.

Cordell, Ryan. 2020. *Machine Learning + Libraries: A Report on the State of the Field*. Commissioned by LC Labs Library of Congress. <https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf>.

Coyle, Karen. 2022. *FRBR, Before and After: A Look at Our Bibliographic Models*. Chicago: ALA.

De Caro, Mario, e Benedetta Giovanola. *Intelligenze. Etica e politica dell'IA*. Bologna: Il Mulino 2025.

Deng, Sui. 2023. "AI, Cataloging & Metadata." *STARS Faculty Scholarship and Creative Works*, University of Central Florida. <https://stars.library.ucf.edu/ucfscholar/1251/>.

D'Souza, Jennifer, Sameer Sadruddin, Holger Israel, Mathias Begoin, e Diana Slawig. 2025. *SemEval-2025 Task 5: LLMs4Subjects -- LLM-based Automated Subject Tagging for a National Technical Library's Open-Access Catalog*. <https://arxiv.org/abs/2504.07199v1>.

ExLibris. 2023. *The Impact of Generative AI on Libraries*. https://discover.clarivate.com/The_impact_of_Generative_AI_on_libraries.

Fumagalli, Giuseppe. 1887. *Cataloghi di biblioteche e indici bibliografici*. Firenze: G.C. Sansoni.

* Ultima data di consultazione dei siti web: luglio 2025.

- Galeffi, Agnese, e Lucia Sardo. 2013. *FRBR*. Roma: Associazione italiana biblioteche.
- Gan, Aoran, Hao Yu, Kai Zhang, Qi Liu, Wenyu Yan, Zhenya Huang, Shiwei Tong, e Guoping Hu. 2025. "Retrieval Augmented Generation Evaluation in the Era of Large Language Models: A Comprehensive Survey." <https://arxiv.org/abs/2504.14891>.
- Ghiringhelli, Lapo, e Mauro Guerrini. 2019. "Entità, attributi e relazioni bibliografiche: rileggendo la tesi PhD di Barbara B. Tillett trent'anni dopo." *AIB Studi* 58 (3): 417-25. <https://doi.org/10.2426/aibstudi-11868>.
- Guerrini, Mauro. 2022a. *Dalla catalogazione alla metadattazione. Tracce di un percorso*. Seconda edizione a cura di Denise Biagiotti e Laura Manzoni. Roma: Associazione italiana biblioteche.
- Guerrini, Mauro. 2022b. *Metadattazione*. Milano: Editrice Bibliografica.
- Guerrini, Mauro, Gianfranco Crupi, e Ginevra Peruginelli, a c. di. 2013. "Global Interoperability and Linked Data in Libraries: Special issue." *JLIS.it* 4 (1). <https://www.jlis.it/index.php/jlis/issue/view/23>.
- Guerrini, Mauro, e Tiziana Possemato. 2015. *Linked data per biblioteche, archivi e musei*. Milano: Editrice Bibliografica.
- Guerrini, Mauro, e Lucia Sardo. 2018. *IFLA Library Reference Model (LRM). Un modello concettuale per le biblioteche del XXI secolo*. Milano: Editrice Bibliografica.
- Guerrini, Mauro, e Franco Neri. 2020. "La tormentata formulazione delle Regole del British Museum del 1839." In *Scaffali come segmenti di storia: studi in onore di Vincenzo Trombetta*, a cura di Rosa Parlavecchia e Paola Zito, 153-65. Roma: Edizioni Quasar.
- Guerrini, Mauro, e Stefano Gambari. 2023. "'Definite cataloguing rules set down in writing': Antonio Panizzi's Rules and the catalogue's manifestations." *JLIS.it* 14 (2): 93-9. <https://doi.org/10.36253/jlis.it-528>.
- Heaps Harold Stanley. 1973. "A theory of relevance for automatic document classification." *Information & Computation* 22 (3): 268-78. [https://doi.org/10.1016/S0019-9958\(73\)90310-0](https://doi.org/10.1016/S0019-9958(73)90310-0).
- IFLA (International Federation of Library Associations and Institutions). 1963. *International Conference on cataloguing principles. Paris, october 9, 1961. Report*. A cura di Arthur Hugh Chaplin e Dorothy Anderson. London: IFLA.
- IFLA (International Federation of Library Associations and Institutions). 1998. *Functional Requirements for Bibliographic Records. Final Report*. Den Haag: IFLA.
- IFLA (International Federation of Library Associations and Institutions). 2011. *ISBD International Standard Bibliographic Description Consolidated Edition*. Den Haag: IFLA.
- IFLA (International Federation of Library Associations and Institutions). 2017. *IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*. A cura di Pat Riva, Patrick Le Bœuf, e Maja Žumer. Den Haag: IFLA.
- IFLA (International Federation of Library Associations and Institutions). 2020. *IFLA Statement on Libraries and Artificial Intelligence*. <https://repository.ifla.org/handle/20.500.14598/1646>.

IFLA (International Federation of Library Associations and Institutions). 2022. *IFLA Library Reference Model. 2017-2022 Consolidation*. A cura di Pat Riva, Patrick Le Bœuf, e Maja Žumer. Den Haag: IFLA.

IPaC (Infrastruttura e servizi digitali per il Patrimonio culturale). 2024. Aggiornamenti di progetto. Alphy 1.9.0. 9 agosto 2024. <https://ipac.cultura.gov.it/2024/08/09/alphy-1-9-0/>.

Islam, Md Nurul, Shakil Ahmad, Mohammad Aqil, Guangwei Hu, Murtaza Ashiq, Majed Mohammed Abusharhah, e Sheikh Abu Toha Md Saky. 2025. "Application of artificial intelligence in academic libraries: a bibliometric analysis and knowledge mapping." *Discov Artif Intell* 5, 59. <https://doi.org/10.1007/s44163-025-00295-9>.

Jones, Steven E. 2018. *Roberto Busa, S. J., and the Emergence of Humanities Computing: The Priest and the Punched Cards*. London: Routledge.

Kaldeli, Eirini, Alexandros Chortaras, Vassilis Lyberatos, Jason Liartis, Spyridon Kantarelis, e Giorgos Stamou. 2024. "Combining Automatic Annotation with Human Validation for the Semantic Enrichment of Cultural Heritage Metadata." In *CHR 2024 Computational Humanities. Proceedings of the Computational Humanities Research Conference 2024 Aarhus, Denmark, 4-6 dicembre, 2024*, a cura di Wouter Haverals, Marijn Koolen, e Laure Thompson, 353-68. <https://ceur-ws.org/Vol-3834/>.

Kimaid, Luís. 2024. "Artificial Intelligence-Driven Archibot: Transforming Access to European Union Parliament Archives." *LegisTech Library*. <https://library.bussola-tech.co/p/artificial-intelligence-archibot-eu-parliament>.

La Gorga, Angelo, Roberto Testa, e Lorenzo Verna. 2025. "LLMs and Retrieval Augmented Generation (RAG) for libraries." *DigitCult - Scientific Journal on Digital Cultures*, 10 (1): 59-73. <https://doi.org/10.36158/97912566920714>.

Lana, Maurizio. 2023. "Leggere l'IFLA Statement on Libraries and Artificial Intelligence al tempo di ChatGPT." *Biblioteche oggi Trends* 9 (1). <https://doi.org/10.3302/2421-3810-202301-004-1>.

Ma, Wei. 2002. "A Database Selection Expert System Based on Reference Librarian's Database Selection Strategy: A Usability and Empirical Evaluation." *Journal of the American Society for Information Science & Technology* 53 (7): 567-80.

Ma, Wei, e Timothy W. Cole. 2001. "Test and Evaluation of an Electronic Database Selection Expert System." In *ACRL 10th National Conference, "Crossing the Divide"*. American Library Association. <http://hdl.handle.net/11213/16793>.

Morris, Victoria. 2019. "Automated Language Identification of Bibliographic Resources." *Cataloging & Classification Quarterly* 58 (1): 1-27. <https://doi.org/10.1080/01639374.2019.1700201>.

Namur, Jean Pie. 1834. *Manuel du bibliothécaire, accompagné de notes critiques, historiques et littéraires*. Bruxelles: J.B. Tircher.

Noruzi, Alireza. 2012. "FRBR and Tillett's Taxonomy of Bibliographic Relationships." *Knowledge Organization* 39 (6): 409-16.

Northwestern University Libraries. 2025. *Enriching Collection Access and Use with Generative AI*. <https://collections-and-ai.library.northwestern.edu/>.

OCLC (Online Computer Library Center). 2025. *Implementing AI to further scale and accelerate WorldCat de-duplication*, 4 febbraio, 2025. <https://www.oclc.org/en/news/announcements/2025/ai-worldcat-deduplication.html>.

Panizzi, Antonio, a c. di. 1841. *Catalogue of Printed Books in the British Museum: Volume I*. London: Printed by order of the Trustees of the British Museum.

Rodriguez, Elena Escolano. 2013. "ISBD adaptation to semantic web of bibliographic data in linked data." *Jlis.it* 4 (1): 119-37. <https://www.jlis.it/index.php/jlis/article/view/259>.

Roncaglia, Gino. 2023. *L'architetto e l'oracolo. Modelli digitali di organizzazione delle conoscenze da Wikipedia a ChatGPT*. Roma-Bari: Laterza.

Roncaglia, Gino. 2024. "IA generativa, system prompt e biblioteche." In *Un incontro di sguardi: biblioteche, libri e lettura come nodi di un reticolo di possibilità creative e generative: scritti in onore di Maurizio Vivarelli*, a cura di Sara Dinotola e Anna Maria Marras, 343-54. Roma: Associazione italiana biblioteche.

Roncaglia, Gino. 2025a. "Verso il reference generativo?" *Biblioteche Oggi Trends* 11 (1). <https://www.bibliotecheoggi.it/it/articolo/4165/verso-il-reference-generativo>.

Roncaglia, Gino. 2025b. *Filosofia dell'intelligenza artificiale*. Milano: Edizioni Corriere della Sera.

Russell, Stuart, e Peter Norvig. 2020. *Artificial Intelligence: A Modern Approach* (4th Edition). Boston: Pearson. <https://aima.cs.berkeley.edu/>.

Saccucci, Caroline, e Abigail Potter. 2024a. "Exploring Computational Description: LC Labs Planning Framework in action" (presentazione PowerPoint tenuta presso il *Works in Progress Webinar: LC Labs AI Planning Framework in action – Understand, experiment, and implement AI tools that support catalogers*). <https://www.oclc.org/research/events/2024/ai-planning-framework-in-action.html>.

Saccucci, Caroline, e Abigail Potter. 2024b. "Could Artificial Intelligence Help Catalog Thousands of Digital Library Books? An Interview with Abigail Potter and Caroline Saccucci." *The Signal* (blog). 19 novembre 2024. <https://blogs.loc.gov/thesignal/2024/11/could-artificial-intelligence-help-catalog-thousands-of-digital-library-books-an-interview-with-abigail-potter-and-caroline-saccucci/>.

Saccucci, Caroline, e Abigail Potter. 2025. "16 Assessing Machine Learning for Cataloging at the Library of Congress." In *New Horizons in Artificial Intelligence in Libraries*, a cura di Edmund Balnaves, Leda Bultrini, Andrew Cox, e Raymond Uzwyshyn, 227-38. Berlin-Boston: De Gruyter Saur. <https://doi.org/10.1515/9783111336435-017>.

Sebastiani, Fabrizio. 2002. "Machine learning in automated text categorization." *ACM Computing Surveys (CSUR)* 34 (1): 1-47. <https://doi.org/10.1145/505282.505283>.

Stevens, Mary Elizabeth. 1965. *Automatic Indexing: A State-of-the-Art Report*. Washington, D.C.: United States Government Printing Office.

Suominen, Osmo, Juho Inkinen, e Mona Lehtinen. 2022. "Annif and Finto AI: Developing and Implementing Automated Subject Indexing." *JLIS.It* 13 (1): 265-82. <https://doi.org/10.4403/jlis.it-12740>.

Tennant, Roy. 2002. "MARC Must Die." *Library Journal* 127 (17): 26-7.

Thomale, Jason. 2010. "Interpreting MARC: Where's the Bibliographic Data?." *Code4Lib Journal* 11. <https://journal.code4lib.org/articles/3832>.

Tillett, Barbara B. "Bibliographic relationships: toward a conceptual structure of bibliographic information used in cataloging". Ph.D. dissertation, University of California, Los Angeles, 1987.

Tomasi, Francesca. 2022. *Organizzare la conoscenza. Digital Humanities e web semantico*. Milano: Editrice Bibliografica.

W3C Library Linked Data Incubator Group. 2011. *Final Report*. <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>.

Zeng, Marcia L., e Qin Jian. 2022. *Metadata*. London: Facet Publishing.