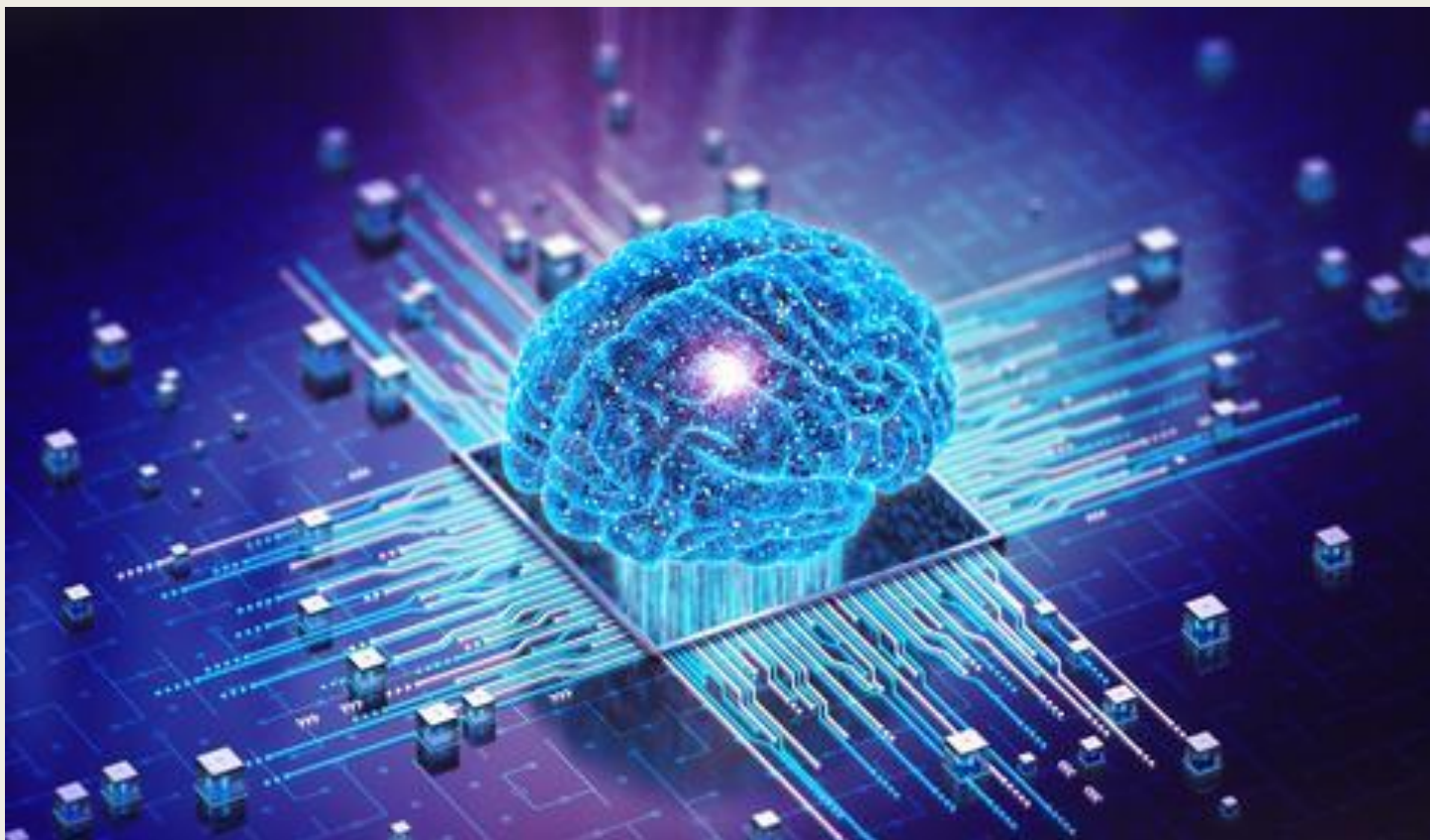




L'addestramento dei modelli linguistici, l'appetibilità dei dati bibliografici e il ruolo delle biblioteche

Simona Turbanti
Università degli studi di Milano

Giornata OCLC Italia
17 marzo 2026
Biblioteca nazionale centrale di Firenze



Intelligenza artificiale e biblioteconomia

L'intelligenza artificiale
può interessare vari
ambiti della
biblioteconomia e delle
biblioteche

Aree di applicazione delle potenzialità dell'IA in biblioteconomia e nelle biblioteche

- Information retrieval (discovery tool)
- Reference di 'prima linea'
- Sistemi di raccomandazione per un'organizzazione del lavoro più efficiente e per la 'scoperta' delle collezioni
- Indicizzazione e, in parte, catalogazione descrittiva
- Report statistici e gestionali
- Attività sociali in biblioteca

...

L'architettura dei modelli linguistici di grandi dimensioni (LLM)

La **dimensione**, la **scalabilità** e la **capacità di generalizzazione** sono caratteristiche dei modelli di machine learning addestrati su vasti insiemi di testi con tecniche di apprendimento supervisionato

L'architettura trasformativa, descritta per la prima volta in un articolo del 2017 dal titolo *Attention is all you need* di Vaswani e altri collaboratori, ha significato un cambio di rotta nel deep learning

Limiti:

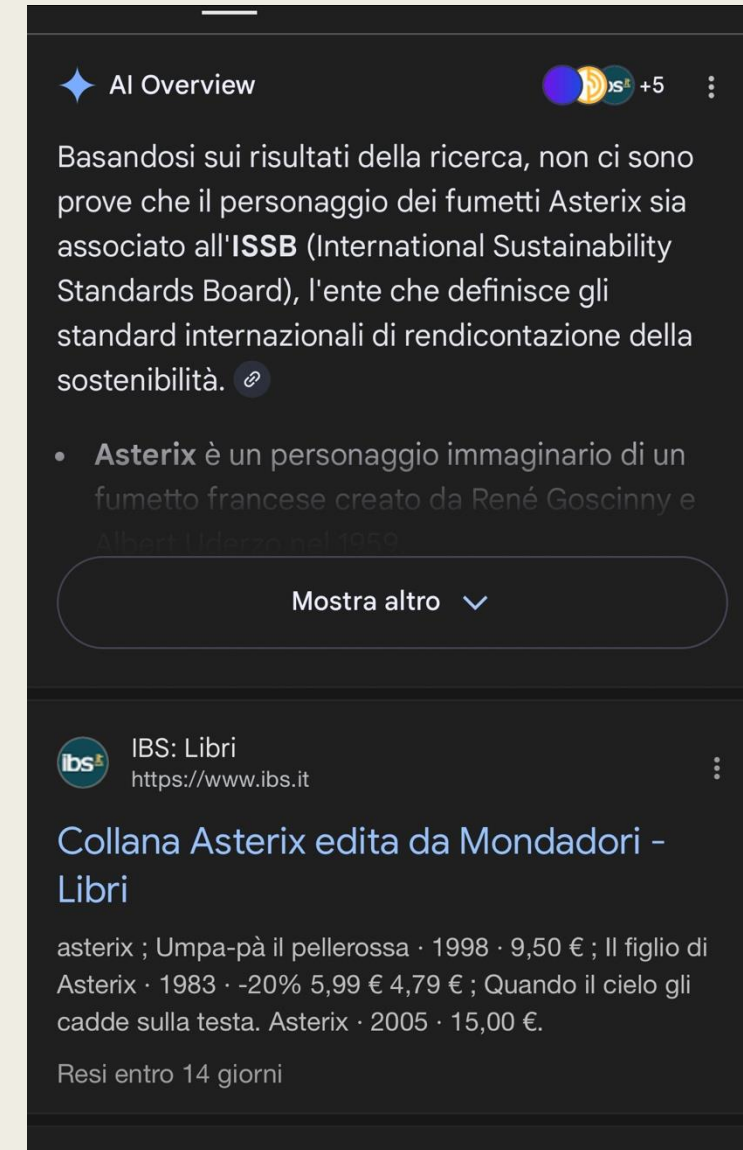
- non dinamismo dell'aggiornamento della base di conoscenza degli LLM
- produzione di risposte plausibili ma non sempre corrette

L'architettura retrieval-augmented generation (RAG)

Tecnica basata su una fase di ricerca (“retrieval”) e una di generazione di contenuto (“generation”) che attinge a diverse fonti di conoscenza sia strutturate sia non, anche in tempo reale →

- riduzione della possibilità di risposte con allucinazioni
- maggiore trasparenza nell'indicare le fonti
- possibilità di personalizzazione in settori specifici e nello stile comunicativo

... ma con vari limiti, almeno per ora

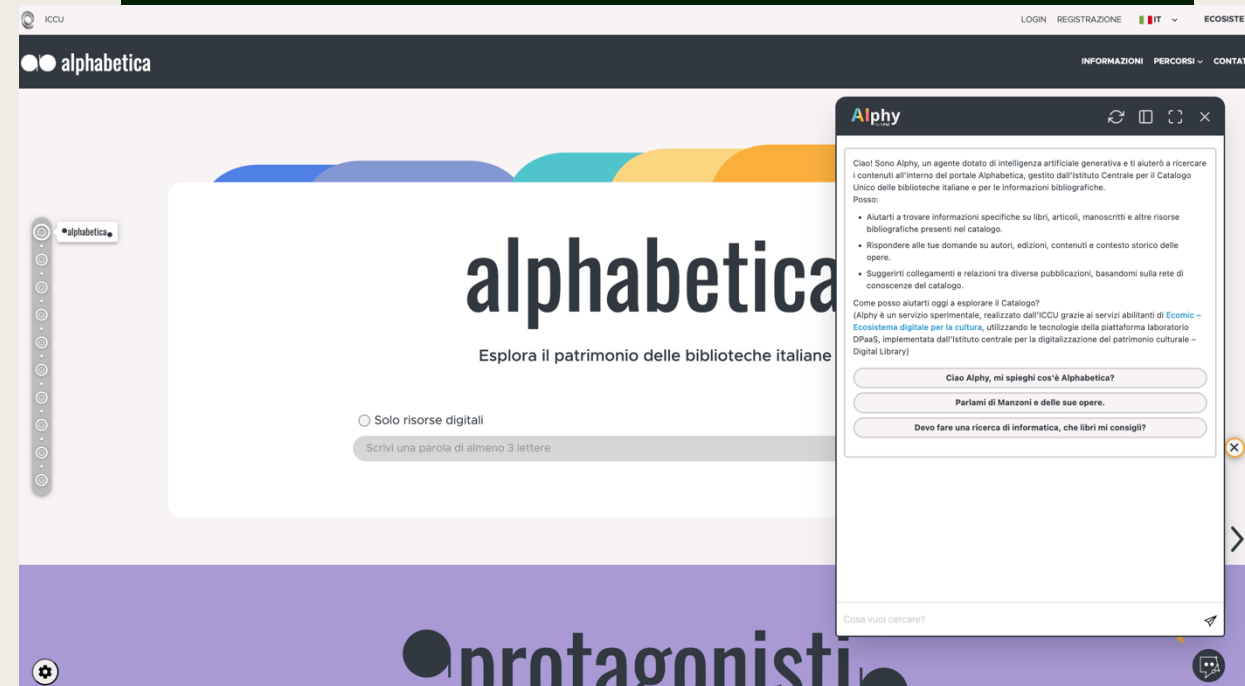
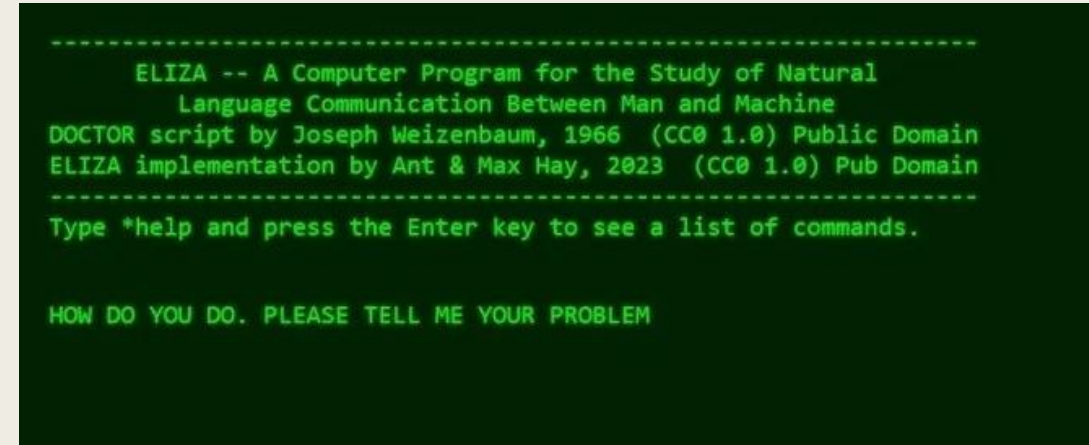


Chatbot e cataloghi/discovery

I **chatbot**, sistemi che hanno lo scopo di conversare con utenti umani in linguaggio naturale, sono stati sviluppati sin dagli anni Sessanta (Eliza)

Le loro funzionalità hanno avuto una svolta nel decennio scorso con le tecniche di elaborazione del linguaggio naturale (NLP) e gli strumenti di machine learning (Alphy)

Potenziati vantaggi per i discovery tool attualmente caratterizzati da falsa facilità di uso, opacità nel ranking dei risultati e scarso supporto alla funzione esplorativa



Requisiti per l'uso di sistemi di IA nelle biblioteche (e dovunque)

L'analisi e la messa a punto di sistemi di IA richiedono:

- ☐ investimenti economici;
- ☐ infrastruttura IT adeguata;
- ☐ dati disponibili e ben organizzati;
- ☐ personale formato;
- ☐ chiarezza sugli obiettivi;
- ☐ accettazione culturale del cambiamento



processo costoso

Data, the new oil (of libraries)...



Se, da una parte, prima di intraprendere qualsiasi progetto, occorre chiedersi se la biblioteca abbia le risorse, le competenze e i dati per usare gli strumenti di IA con profitto, dall'altra, alcuni fatti ci fanno capire il “potere” in mano alle biblioteche stesse, ossia:

- dati **strutturati**
- collezioni digitalizzate **accessibili** (talvolta in pubblico dominio)
- dati di genere diverso («le biblioteche interessano perché aiutano a migliorare la **diversità** dei dati. Le collezioni bibliotecarie coprono lingue, epoche, territori e generi che nei dataset commerciali sono spesso sottorappresentati», risposta di ChatGPT 5.4 Pro, 11 marzo 2026)



HARVARD
LAW SCHOOL

Institutional Data Initiative

contact@institutionaldatainitiative.org

Library Staff A-Z

On this page

[Our Team](#)

[News About Our Work](#)

[Library Staff](#) → Institutional Data Initiative

The Institutional Data Initiative is a research initiative at Harvard Law School Library. We work with knowledge institutions—from libraries and museums to cultural groups and government agencies—to refine and publish their collections as data to facilitate responsible AI training.

Visit the [IDI website](#) to learn more.

Our Team

Le biblioteche come hub e la funzione dei cataloghi

Le biblioteche rappresentano non soltanto i luoghi fisici nei quali si conserva e rende disponibile il materiale bibliografico, e non di rado documentale, in forma analogica e digitale, ma anche i luoghi più idonei in cui accertare l'autorevolezza delle fonti per mezzo di strumenti di ricerca; nonostante le modifiche di questi ultimi - da cataloghi cartacei a OPAC, da OPAC a next generation catalogue e a discovery tool - le loro funzioni sono rimaste immutate e rispondono al desiderio umano, prima che bibliotecario, di attuare l'Universal Bibliographic Control (UBC). La sfida attuale è quella di reinterpretare l'utopia del controllo bibliografico universale all'interno di un vasto contesto informativo che interagisce con sistemi "intelligenti"