Taylor & Francis

AI and Libraries: A New Era in Scholarly Resource Management

CONCERT 2024

Dr. Paolo Lombardi - Director of Artificial Intelligence Innovation @ Taylor & Francis

paolo.lombardi@tandf.co.uk

Outline

1. The evolving role of librarians in the age of AI
2. Impact of AI on the research process
3. Modern semantic search and its implications
4. Retrieval Augmented Generation and its implications
5. Research automation and the future of research consumption

Taylor & Francis

Impact of AI

How research and libraries are changing

Librarians: A Keystone of Scholarly Production

Source: Image AI generated by the speaker using GPT-4o

Librarians are at the forefront of managing scholarly communication.

They bridge the gap between researchers and information, ensuring knowledge is:

- Accessible,
- Organized,
- Preserved.

Their pivotal role advances scholarship and fosters innovation in the academic community.

Today's Librarians: Digital Navigators

Guide users through the overwhelming sea of information:

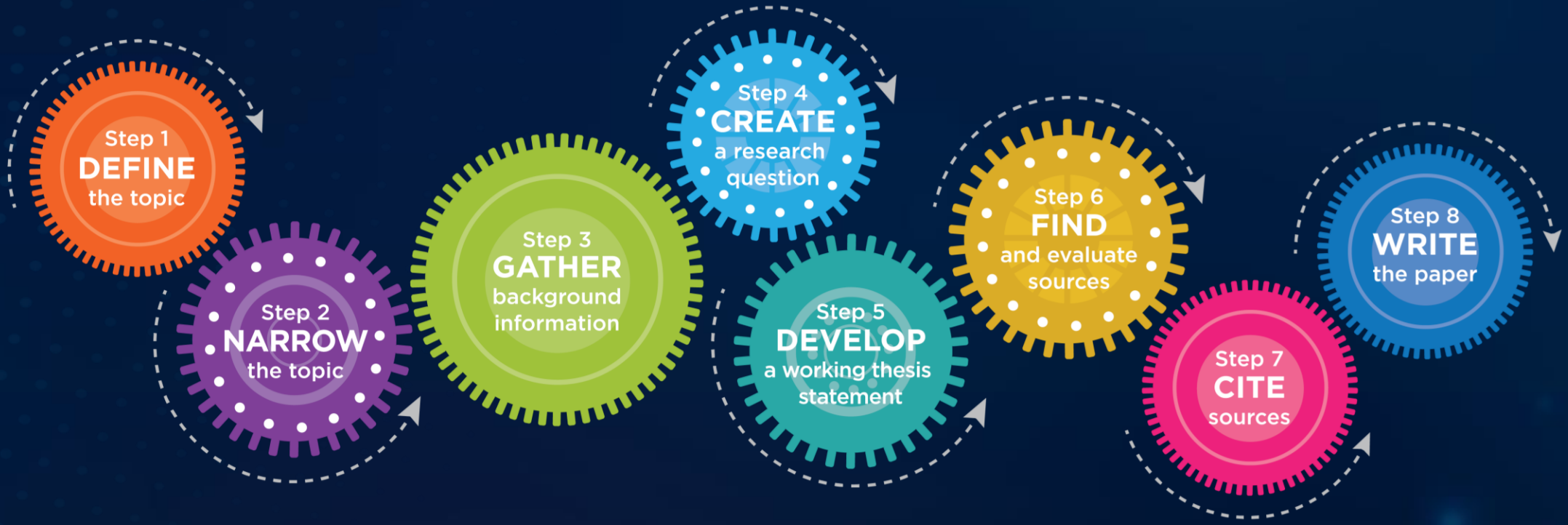
- Curate digital collections
- Navigate electronic databases
- Understand complex digital platforms
- Understand and manage metadata
- Help users develop digital skills

Facilitate the entire lifecycle of information—from creation and dissemination to preservation and access.

Embrace AI tools to enhance discovery and search, making finding scholarly content more intuitive and efficient.

Source: IkonStudio / Alamy Stock Vector

How Researchers Have Accessed & Used Content So Far



SEARCH
Books and Articles

1. By keyword
2. By author
3. By journal

SCREEN
Read abstract or summary → Decide if it is worth reading the rest.

UNDERSTAND
Read full body, connect with other knowledge or sources, test, reflect, learn.

SUMMARIZE
Cluster sources by topics, methods, and contributions; create Literature Reviews.

Taylor & Francis

Impact of AI

SEARCH

- Concepts instead of just keywords
- ‘Chat with content’ and iterative exploration
- Brainstorming new ideas

Advantages
- More relevance
- Gain time

SCREEN

- Classify papers by type or topic
- Extract important points
- Check the references more efficiently

Advantages
- More focus
- Gain time

UNDERSTAND

- Explain graphs, tables, formulas
- Connect to related articles and books
- Accessibility: e.g., translate, alt-text, or text-to-voice

Advantages
- More ideas
- Gain time

SUMMARIZE & PUBLISH

- Writing assistant
- Create clusters of related sources
- Plain Language Summaries

Advantages
- Communicate more effectively
- Gain time

Taylor & Francis

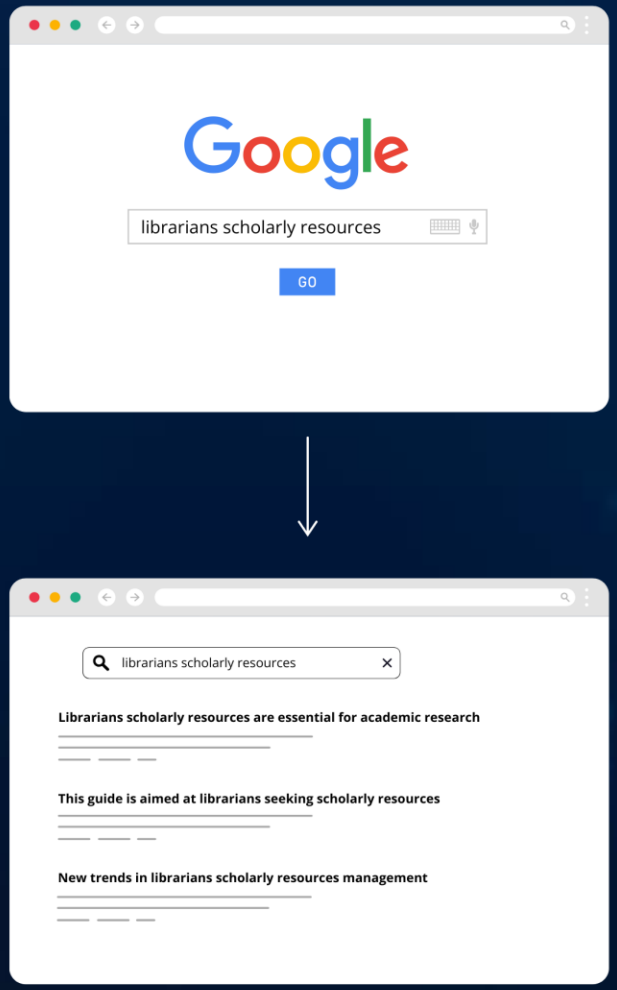
Semantic Search

Its evolution in the last 25 years

2000s: Keyword-Based Search

Up to the early 2000s, search engines like Google primarily relied on keyword matching. These systems worked by identifying the words entered by a user and matching them with documents that contained those exact terms.

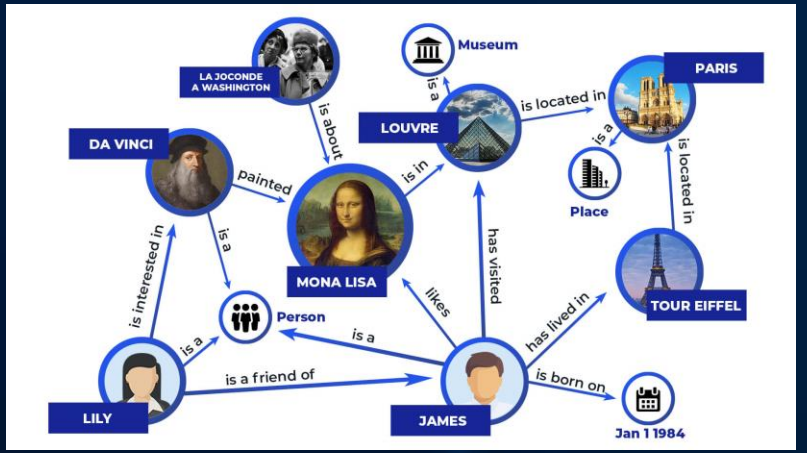
- **Strengths:** Fast and efficient for simple searches where exact word matches were sufficient.
- **Limitations:** Limited understanding of context or intent. It couldn't differentiate between words with multiple meanings (e.g., "apple" as a fruit vs. "Apple" as a company) and didn't capture relationships between terms.
- **Focus**: Matching exact terms, or weighted frequency (LSI).
- **Example**: Searching "librarians scholarly resources" returns documents with those words, regardless of context.



2010s: Metadata and Knowledge Graphs

In the 2010s, semantic search took a major leap with the introduction of **knowledge graphs** (KGs) and advancements in **NLP (Natural Language Processing).** Document elements were recognised and tagged as **entities** (people, places, concepts) so that search algorithms could process their **relationships**. NLP was used to recognise the intent of the user and process natural language queries.

- **Strengths:** KGs and NLP allowed to process queries according to their meaning and intent, allowing for more natural interactions.
- **Limitations:** Though more powerful, these systems were still largely rule-based and required manual updates to knowledge graphs.
- **Focus**: Understanding entities and relationships.
- **Example**: Search engines recognize "librarians" as entities and relate them to scholarly resources, potentially providing answers, not just links.

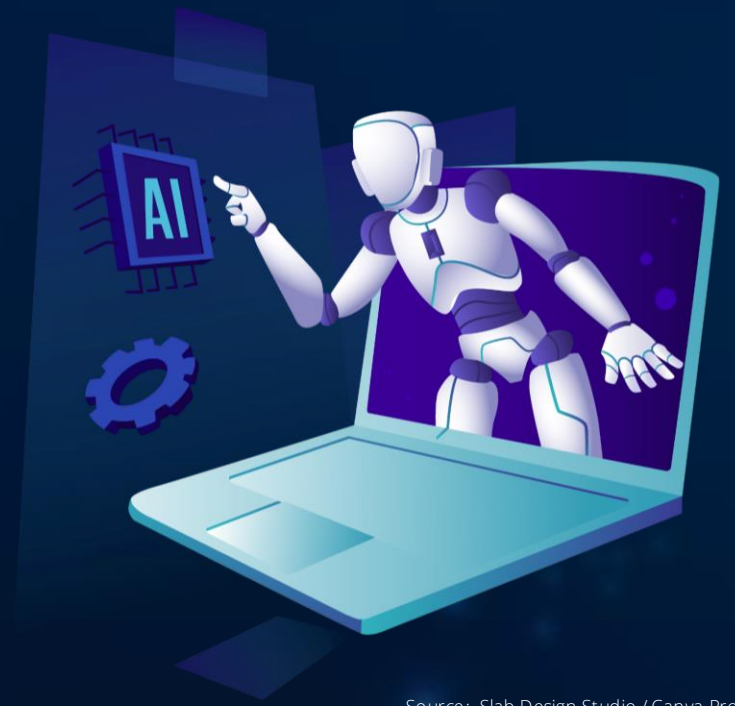


Source: https://www.atulhost.com/what-is-knowledge-graph

2020s: AI & Deep Learning (Transformers, BERT, GPT)

In the 2020s, **deep learning** and **transformer models** like **BERT** (Bidirectional Encoder Representations from Transformers) and **GPT** (Generative Pretrained Transformer) revolutionized semantic search. These models enabled search engines to deeply understand context, intent, and even infer meaning from user queries without relying on predefined relationships.

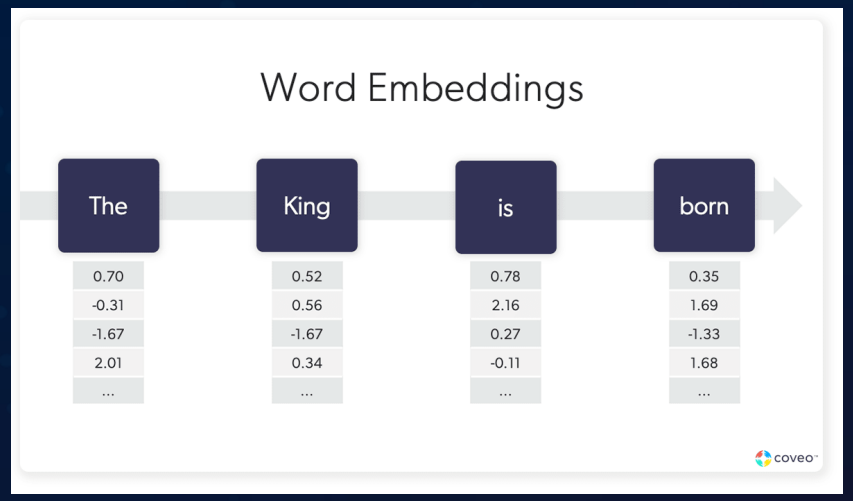
- **Strengths:** These AI models can handle complex queries, understand synonyms, and even anticipate user needs based on **context**.
- **Limitations:** Though advanced, these models are resource-intensive and require massive datasets and computational power.
- **Focus**: Contextual understanding, deeper intent analysis.
- **Example**: A query like "How do librarians manage AI in scholarly resources?" can return comprehensive, context-aware answers, possibly with AI-generated summaries and citations.

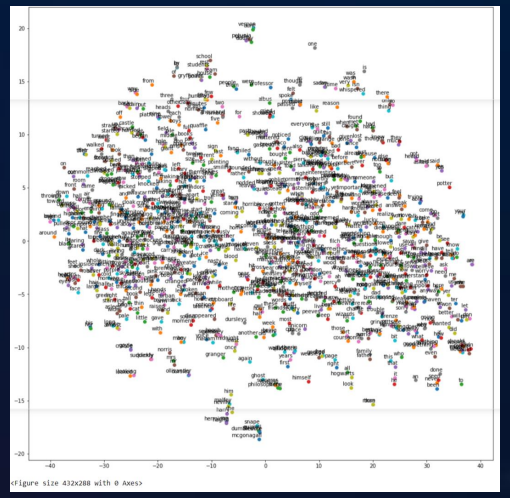


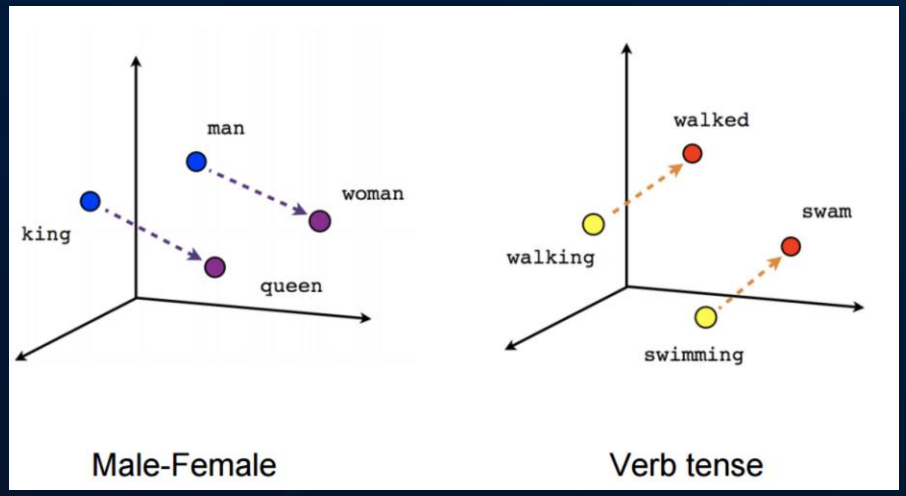
Source: Slab Design Studio / Canva Pro

Embeddings: Concepts in the Space of Meaning

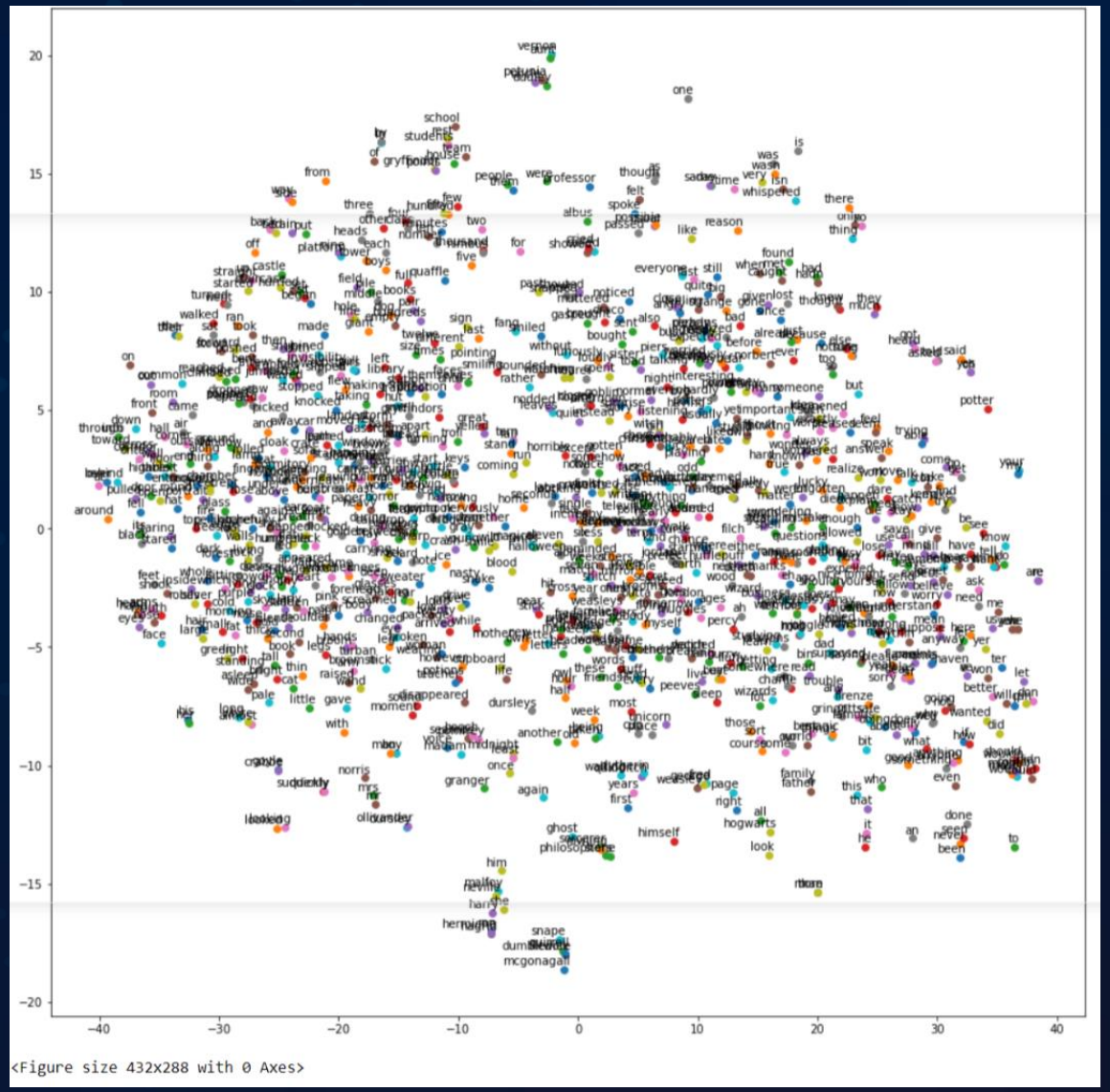
The	King	is	born	Queen
1	0	0	0	0
0	0.25	0	0	0.75
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1







Embeddings: Concepts in the Space of Meaning



A representation of concepts with 2-dimensional embeddings.

State of the art embeddings today use 1024 to 4096 numbers (dimensionality of the vector space) for each concept.

Similar concepts are close together, different concepts are far from each other.

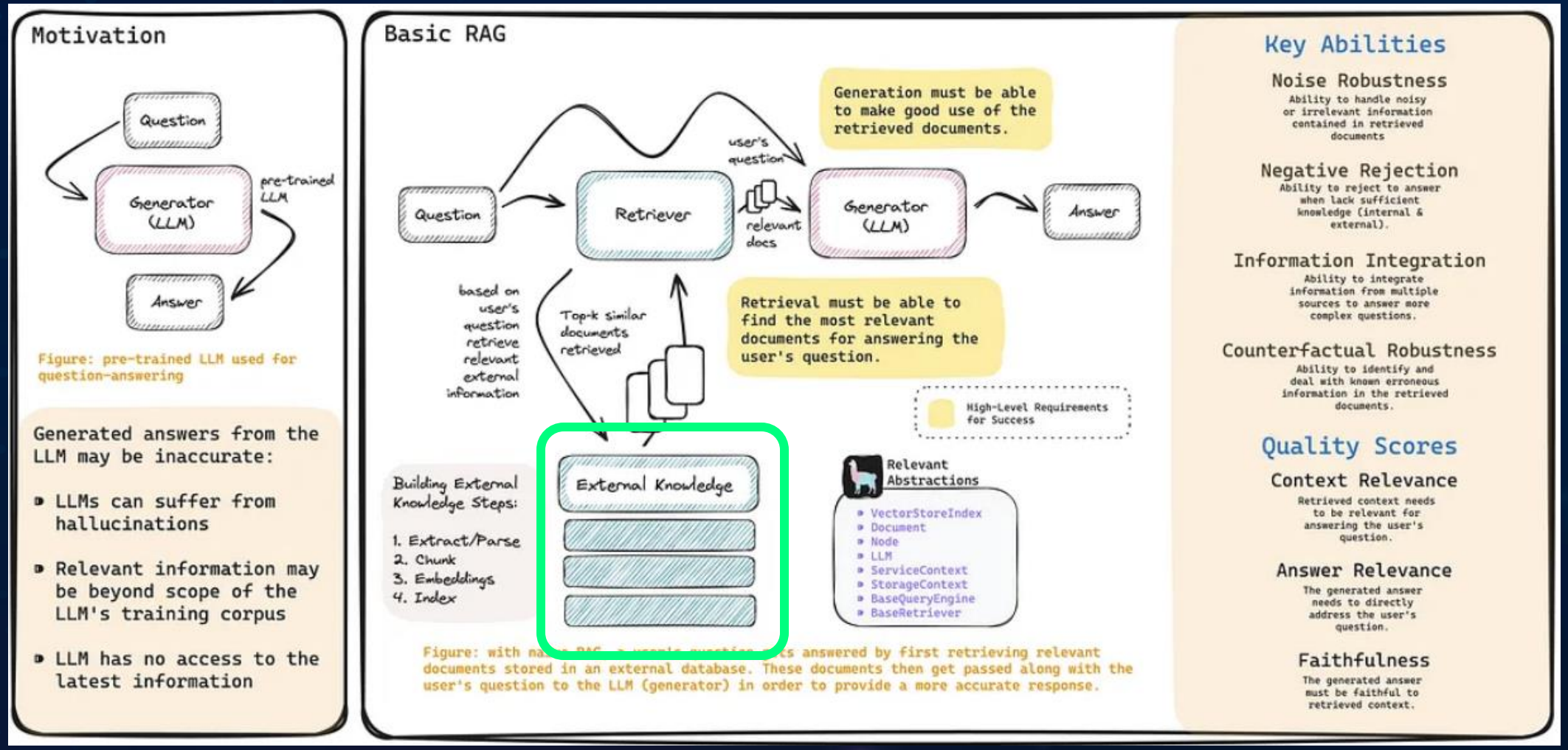
Main advantage: search by proximity in the space of meaning instead of having to process the words and spellings.

Taylor & Francis

RAG AI Systems

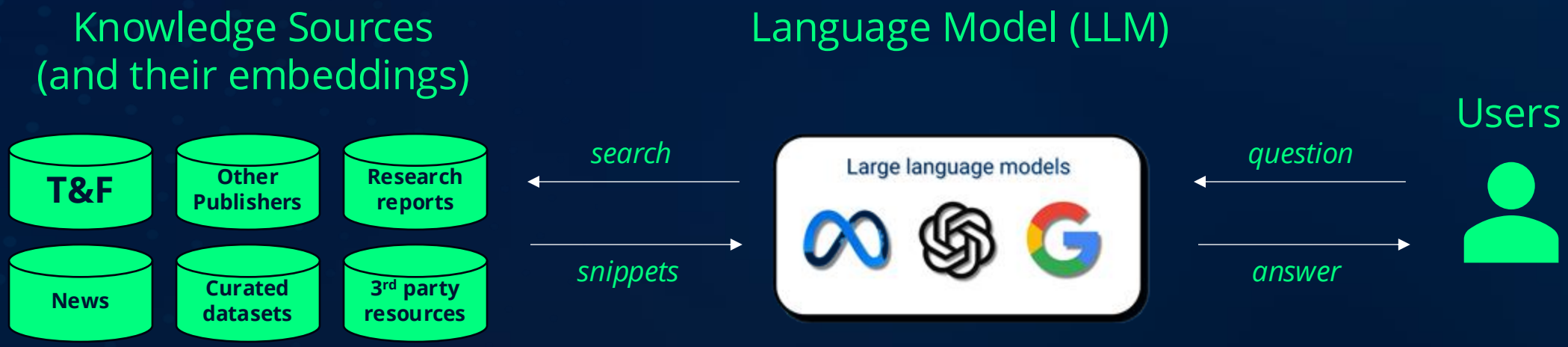
Retrieval Augmented Generation

Retrieval Augmented Generation



Source: Andrei (LlamaIndex Blog) / Medium.com

RAG: Collaborative Interaction of Sources and LLM



Provide:

- Validated knowledge (peer review)
- Updates (journals)
- Verifiable source (author information)
- Source attribution
- Negative rejection (exclude hallucinations)

Provide:

- Human-like interaction
- Context awareness
- Information integration and summarisation
- Ability to discern complex questions
- Ability to activate other tools (calculator, web…)

Advantages of RAG

Does NOT require expensive AI training

Preserves attribution – compatible with academic practices

Easy and relatively cheap to implement

Allows easy addition or removal of sources (retractions, expired subscriptions)

Can be targeted to the audience (department, educational, etc)

Managing the Knowledge Sources

It will become increasingly important to include the right knowledge sources in interactive AI systems:

1. Identify information sources needed by researchers
2. Assess or guarantee their quality → Precision, Accuracy
3. Ensure sufficient coverage, inclusion, diversity → Recall
4. Connect them via frontend agents / applications

Who will curate and provide the frontend agent / application (chatbot)?

Possible providers:

- Chatbot Producers?
- Bibliometrics Providers?
- Publishers?
- Institutions?
- Distribution Platforms?
- Startups?
- … Libraries?

Google

THOMSON REUTERS

MSCI

Announced Mar-Oct 2024

OpenAI

FT FINANCIAL TIMES

News Corporation

Announced Apr-Oct 2024

science OS

Founded 2023

Clarivate Web of Science™ AI Assistant

Launched Sept 2024

Scopus® AI

Change the way you view knowledge

Launched Aug 2023

Taylor & Francis

Taylor & Francis

The Future

Research automation

AI at Taylor & Francis

We are conducting and partnering for AI R&D along the entire research process

Cross-functional activity involving Product & Technology, Editors, Legal, Policy and Ethics

Objectives:

1. Unlock all our content, including pre-digital-era foundational books and articles
2. Enable researchers to access the content that is relevant to them more easily
3. Enable librarians to manage knowledge sources more efficiently and selectively
4. Enable the use of Open Access content and other selected content for AI applications, in a way that is respectful of authors' rights

The Automated Researcher

2024-9-4

The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

Chris Lu[1,2,*], **Cong Lu**[3,4,*], **Robert Tjarko Lange**[1,*], **Jakob Foerster**[2,†], **Jeff Clune**[3,4,5,†] **and David Ha**[1,†]

[*]Equal Contribution, [1]Sakana AI, [2]FLAIR, University of Oxford, [3]University of British Columbia, [4]Vector Institute, [5]Canada CIFAR AI Chair, [†]Equal Advising

- Generates novel research ideas
- Writes code
- Executes experiments
- Visualizes results
- Describes its findings
- Writes a full scientific paper
- Runs a peer review of its own paper

Source: Tatiana Egorova / Alamy Stock Vector

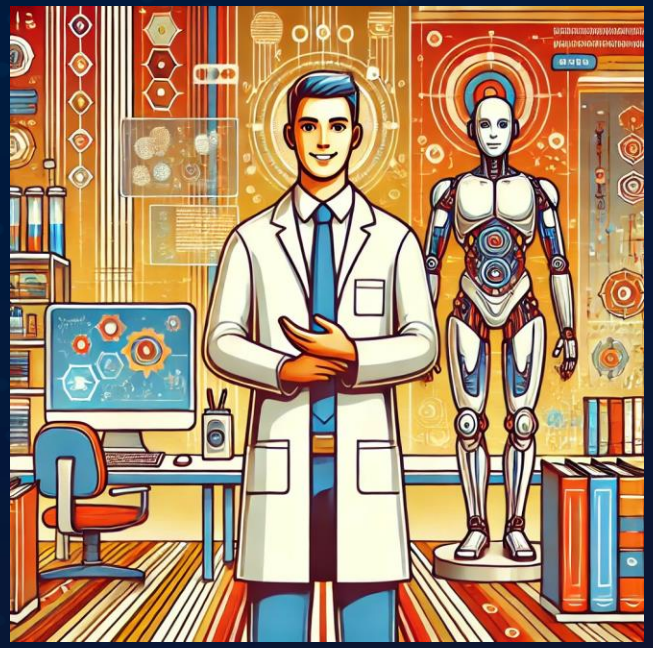
Taylor & Francis

Importance of Source Selection and Curation

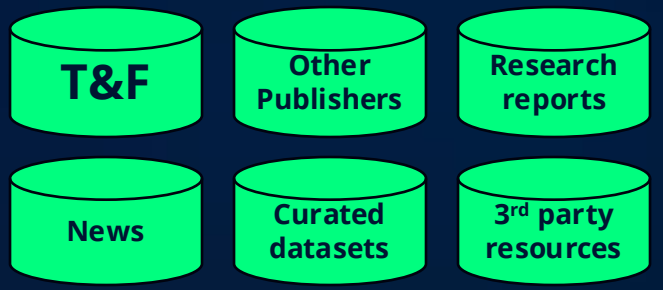
Knowledge Sources
(and their embeddings)

Human Researchers

UI



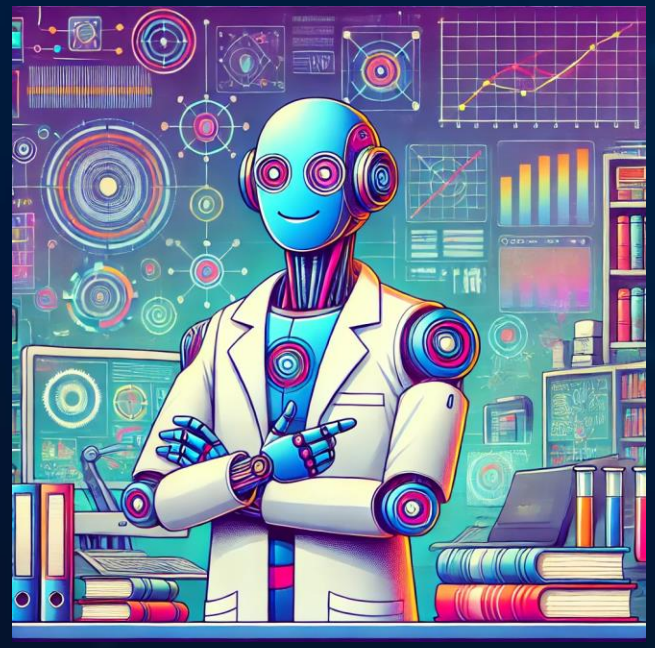
Source: Image AI generated by the speaker using GPT-4o



- Guarantee integrity
- Factuality and verifiability
- Diversity and inclusion
- Shared understanding
- Collaboration

API

AI Researchers



Source: Image AI generated by the speaker using GPT-4o

Taylor & Francis

Thank you

Dr. Paolo Lombardi - Director of Artificial Intelligence Innovation @ Taylor & Francis

paolo.lombardi@tandf.co.uk