Generative AI

Critical issues for libraries

(What have we learned do far?)

SirsiDynix®
CONNECTIONS

Marshall Breeding
Independent Consultant, Author, and
Founder and Publisher, Library Technology Guides
https://librarytechnology.org/
https://twitter.com/mbreeding

Tuesday, October 22, 2024 || 11:50 am – 12:20 pm

Overview

Generative AI has quickly swept into the consumer, business, and educational spheres. It is essential for libraries to have a firm grasp on this new technology. As experts in information services, librarians must be aware of how AI-generated content has made its way into the scholarly literature and popular culture and must be prepared to advise their communities about its ethical, legal, and pragmatic use. To what extent should libraries integrate AI into their own systems and services?

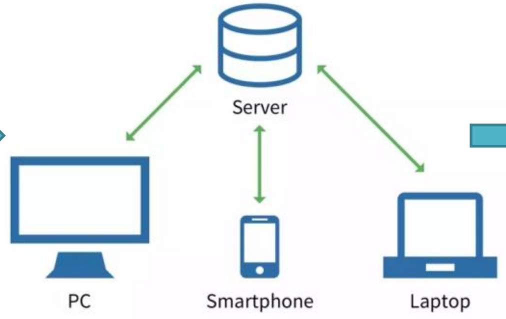
Marshall Breeding will explore these questions and reflect some recent trends and developments.

Paradigm shifts in Computing

Mainframe Computing

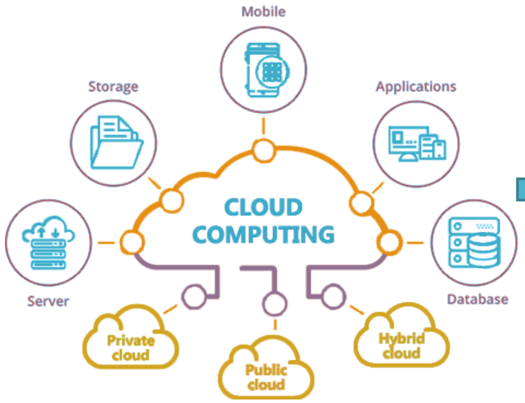
Highly efficient centralized computers, though with miniscule capacity by current standards

Client / Server

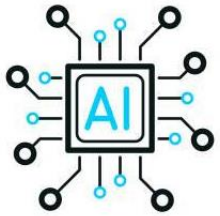


Distributed computing model to take advantage of desktop computers. Issues with scalability and support overhead.

Cloud Computing



No longer bound to the capabilities of single servers but can tap into massively distributed computational resources.



Powered by massive cloud computing capabilities

Paradigm Shifts in Library Automation

Business process automation: technology support for circulation, resource acquisition, cataloging, physical processing

Customer service delivery: resource discovery, online patron services, script-based chat based on knowledge bases

AI-enhanced services: AI-generated metadata at scale; automated image and video description; Answer patron questions (with citations) not just list citations; translations and multi-lingual services

Challenge:

How can libraries make appropriate use of AI and related technologies to improve services, optimize operations, and strengthen their position within their institutions and communities?

The mission of the library should drive technology services. Interest in AI should not derail existing successful services

AI use in libraries must be closely aligned with library missions, professional standards, legal requirements, and ethical concerns.

Out-of-the-box AI usually misses the mark

Generative AI Basics

Large Language Models

The most prominent framework driving Generative Artificial Intelligence

A type of Neural Network

Can perform natural language processing tasks

Incorporates massive amounts of data: Large number of parameters or (many billions) and pretraining tokens (many trillions)

Transformer model: encodes input into tokens, internal learning process using mathematical algorithms to discover relationships among tokens

Generative AI: AI models can generate new content: text, software code, images, video

Large Language Models a type of Generative AI trained on text and that can generate new textual content

Statistical approach: based on prompts, context, and the LLM: text generation based on predicting the next word, sentence, or passage needed to complete the response

Building an LLM

Based on a deep learning foundation model

Training: most LLMs are pre-trained on large-scale bodies of text. Convert the training material into tokens, feed the tokens into the model, which processes to tokens to understand and optimize relationships and meaning. Computationally intensive process usually handled by large-scale distributed GPU servers

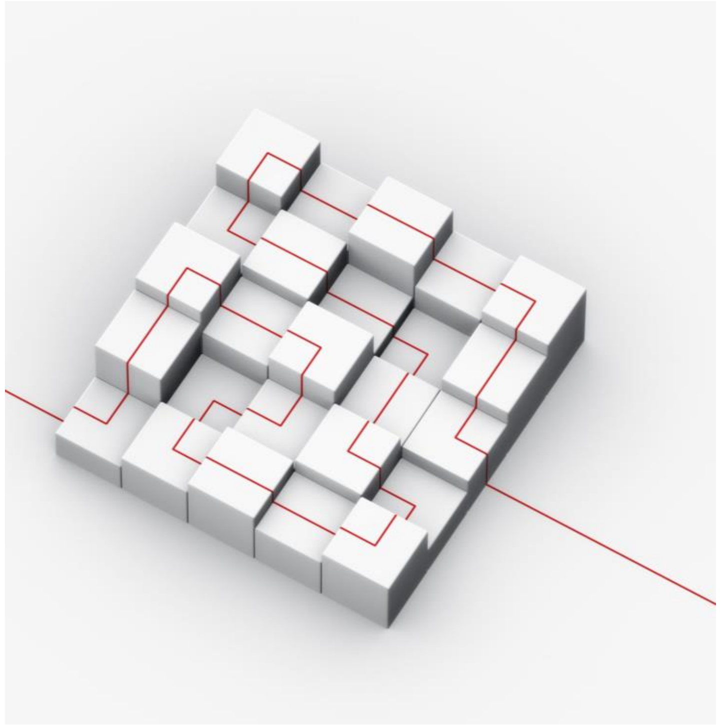
AI projects may or may not disclose the data used for training.

General content on the internet: Wikipedia, Redit, digitized books, open access articles, news. Similar scale of harvesting content on the Web as Google and other search engines.

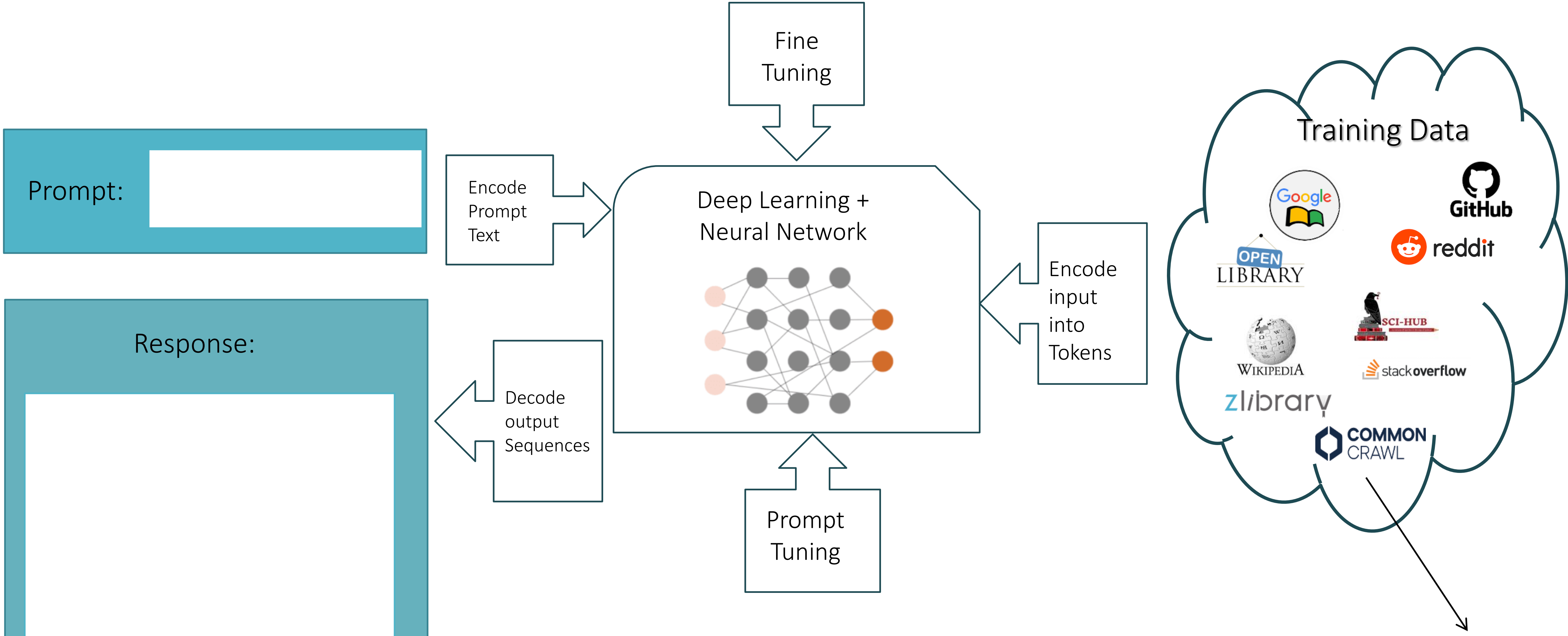
Legal and ethics concerns regarding permissions to incorporate copyrighted text in AI training data

Fine Tuning: process to optimize the LLM to perform specific types of activities

Prompt tuning: trains the model to perform tasks based on feeding it instructions or prompts



Generic Large Language Model



Over 250 billion pages spanning 16 years.
Free and open corpus since 2007.
Cited in over 10,000 research papers.
3–5 billion new pages added each month.
Primary training corpus in every LLM.
82% of raw tokens used to train GPT-3.
https://commoncrawl.org/

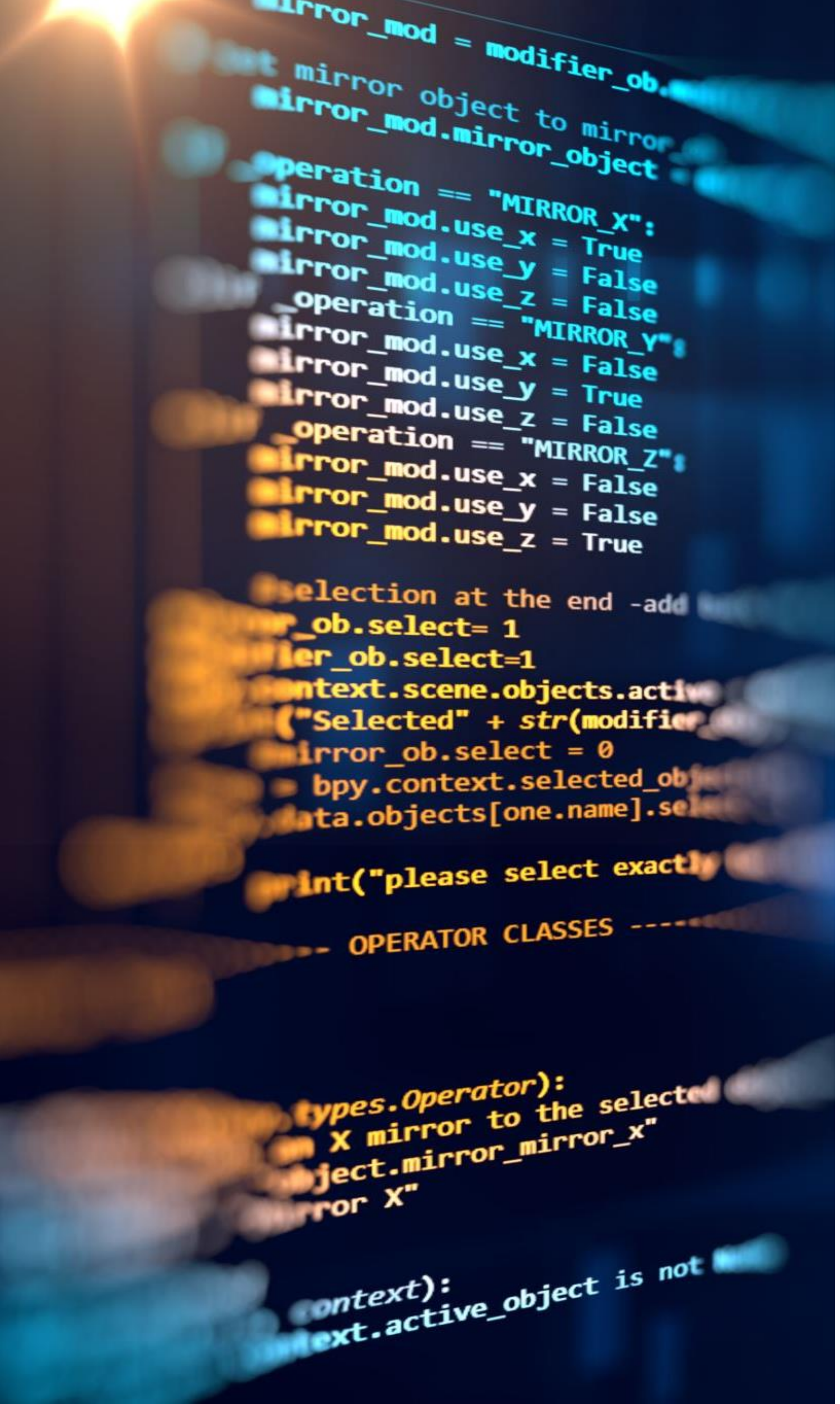
Examples of Large Language Models

GPT: developed by OpenAI (GPT-3, GPT-4) powering the ChatGPT service

PaLM2 (Pathways Language Model) developed by Google

LLaMA open source LLM developed by Meta (licensed for research and commercial use)

Jurassic-1 Developed by AI21 Labs.



LLM Capabilities

Generate responses in many languages or media, based on the content available in the training data

Problem solving: ability to emulate reasoning. Solve any variety of mathematical problems or inferences

Generate programming code (many languages)

Interact with structured data

Generate images, video, other multimedia content

Analyze text: summaries, concept extraction, analysis

What content can AI companies use?

Most of the AI companies do not publicly disclose what content is used for training their LLM

Some authors, artists, and publishers have sued OpenAI and others for copyright infringement

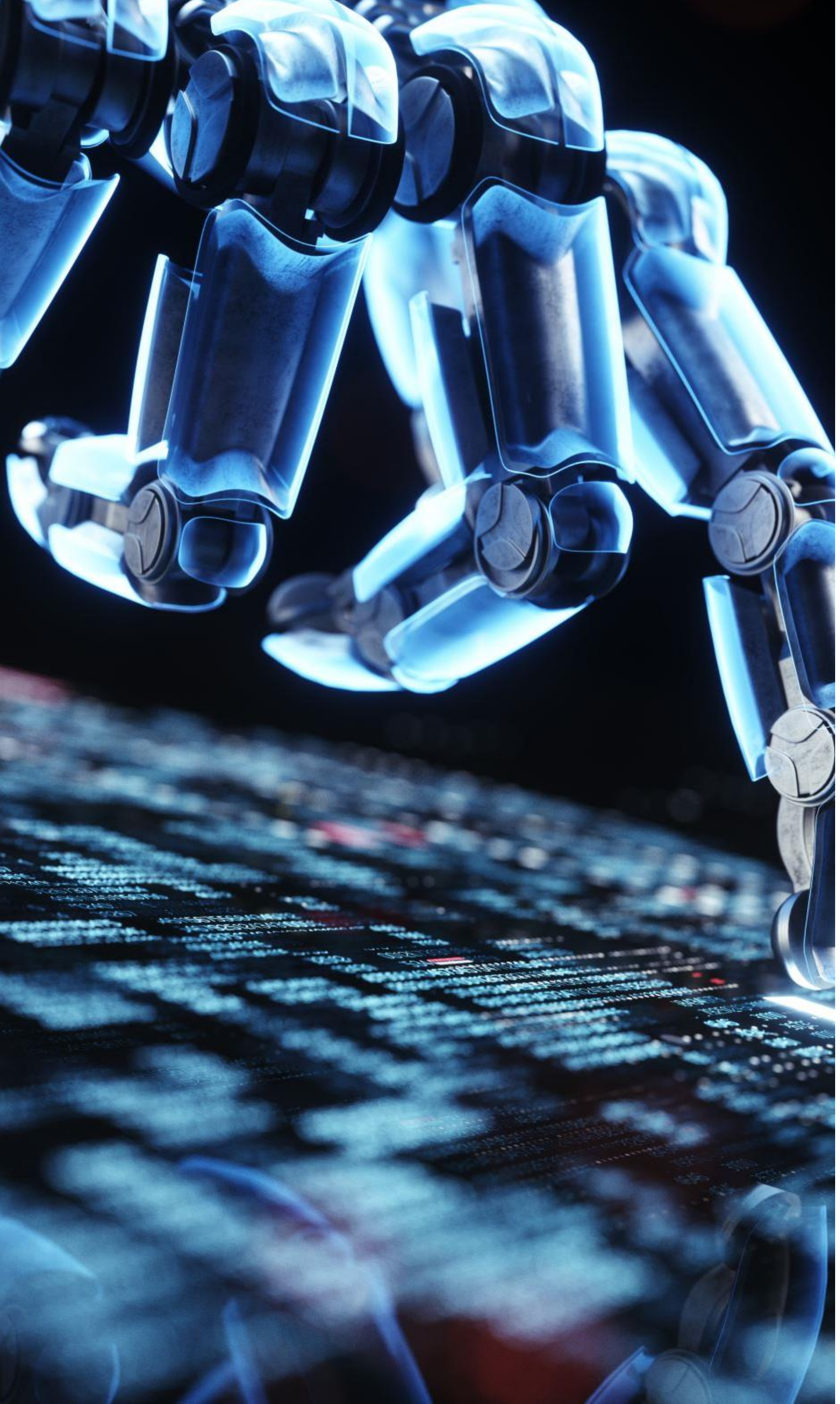
AI harvesting bots are aggressive

Private licenses between content owners/publishers and AI developers

Two Major Academic Publishers Signed Deals With AI Companies. Some Professors Are Outraged, Chronicle of Higher Education July 29, 2024
https://www.chronicle.com/article/two-major-academic-publishers-signed-deals-with-ai-companies-some-professors-are-outraged

New York Times sues OpenAI, Microsoft for using articles to train AI (December 27, 2023): “OpenAI and Microsoft used “millions” of Times articles to help build their tech, which is now extremely lucrative and directly competes with the Times’s own services”

https://www.washingtonpost.com/technology/2023/12/27/new-york-times-sues-openai-chatgpt



The (dis)information Ecosystem

An unfortunate reality:

- ❖ Reliable information is difficult to access due to paywalls or other obstacles
- ❖ Misinformation is abundant and easy to access
- ❖ Significant impact on what is presented through search engines and incorporated into AI training data
- ❖ Generative AI has the potential to create misinformation at scale
- ❖ It is often impossible to distinguish text, images, and video created via AI

AI enhanced Search

Delivers answers or summaries rather than links

Changes dynamics between search engines and websites

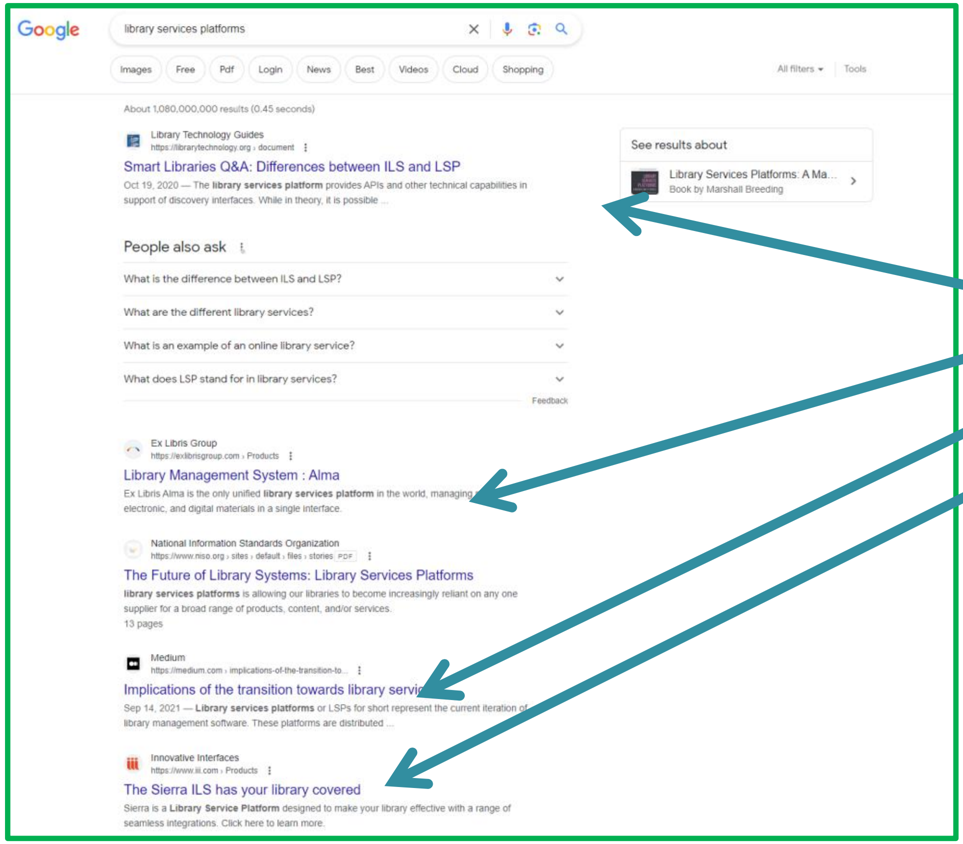
More than 50% of requests to web sites are from search engine bots

Website overhead: Tolerable if it results in increased traffic to website

What is the payoff to website operators if harvesting does not result in higher traffic?

Google Search presents delivers links to sponsored and organic results

library services platforms



Google

library services platforms

About 1,080,000,000 results (0.45 seconds)

Library Technology Guides
https://librarytechnology.org › document

Smart Libraries Q&A: Differences between ILS and LSP

Oct 19, 2020 — The **library services platform** provides APIs and other technical capabilities in support of discovery interfaces. While in theory, it is possible ...

See results about

Library Services Platforms: A Ma...
Book by Marshall Breeding

People also ask

What is the difference between ILS and LSP?

What are the different library services?

What is an example of an online library service?

What does LSP stand for in library services?

Ex Libris Group
https://exlibrisgroup.com › Products

Library Management System : Alma

Ex Libris Alma is the only unified **library services platform** in the world, managing electronic, and digital materials in a single interface.

National Information Standards Organization
https://www.niso.org › sites › default › files › stories PDF

The Future of Library Systems: Library Services Platforms

library services platforms is allowing our libraries to become increasingly reliant on any one supplier for a broad range of products, content, and/or services.
13 pages

Medium
https://medium.com › implications-of-the-transition-to...

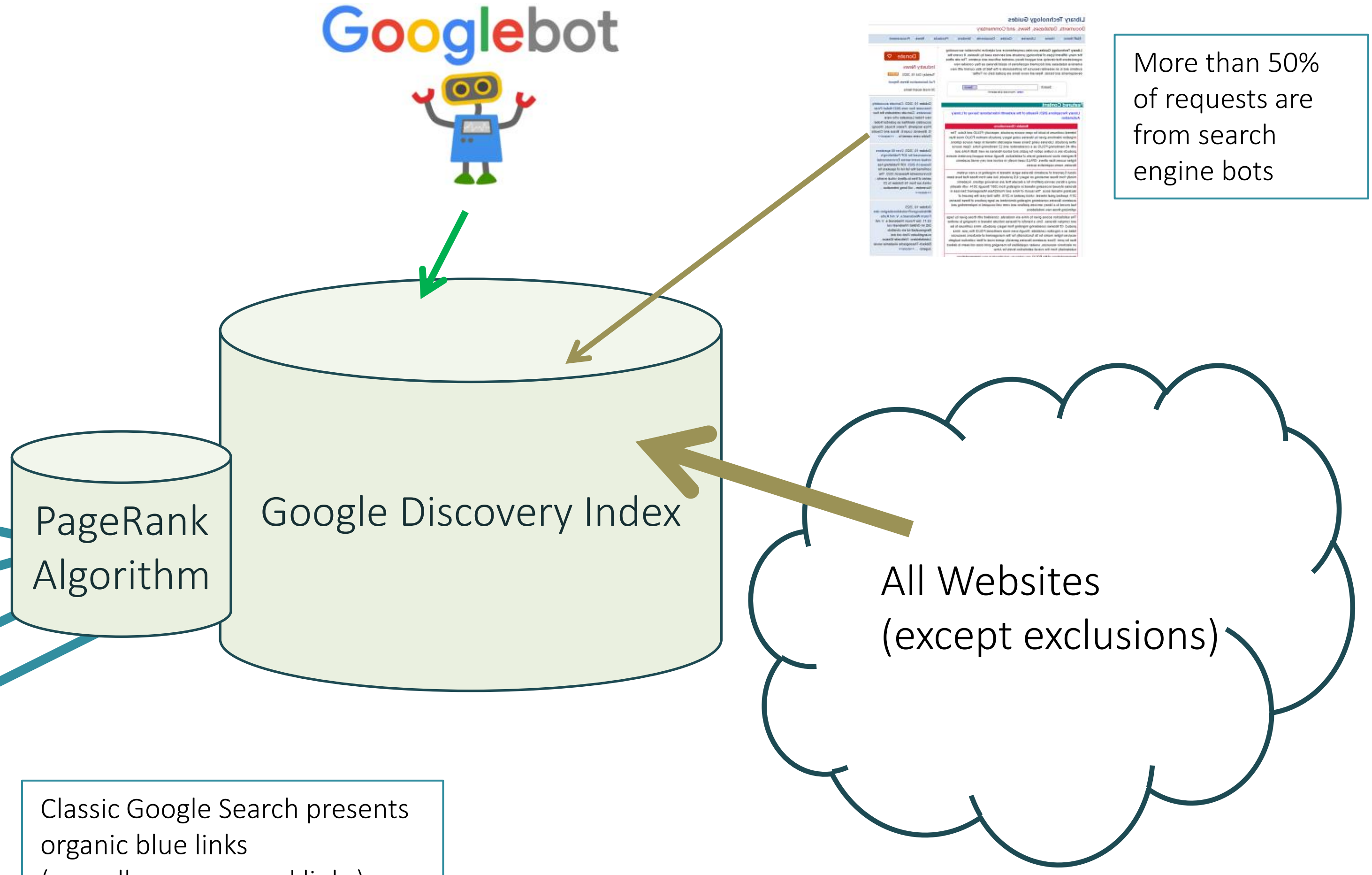
Implications of the transition towards library servi

Sep 14, 2021 — **Library services platforms** or LSPs for short represent the current iteration of library management software. These platforms are distributed ...

Innovative Interfaces
https://www.iii.com › Products

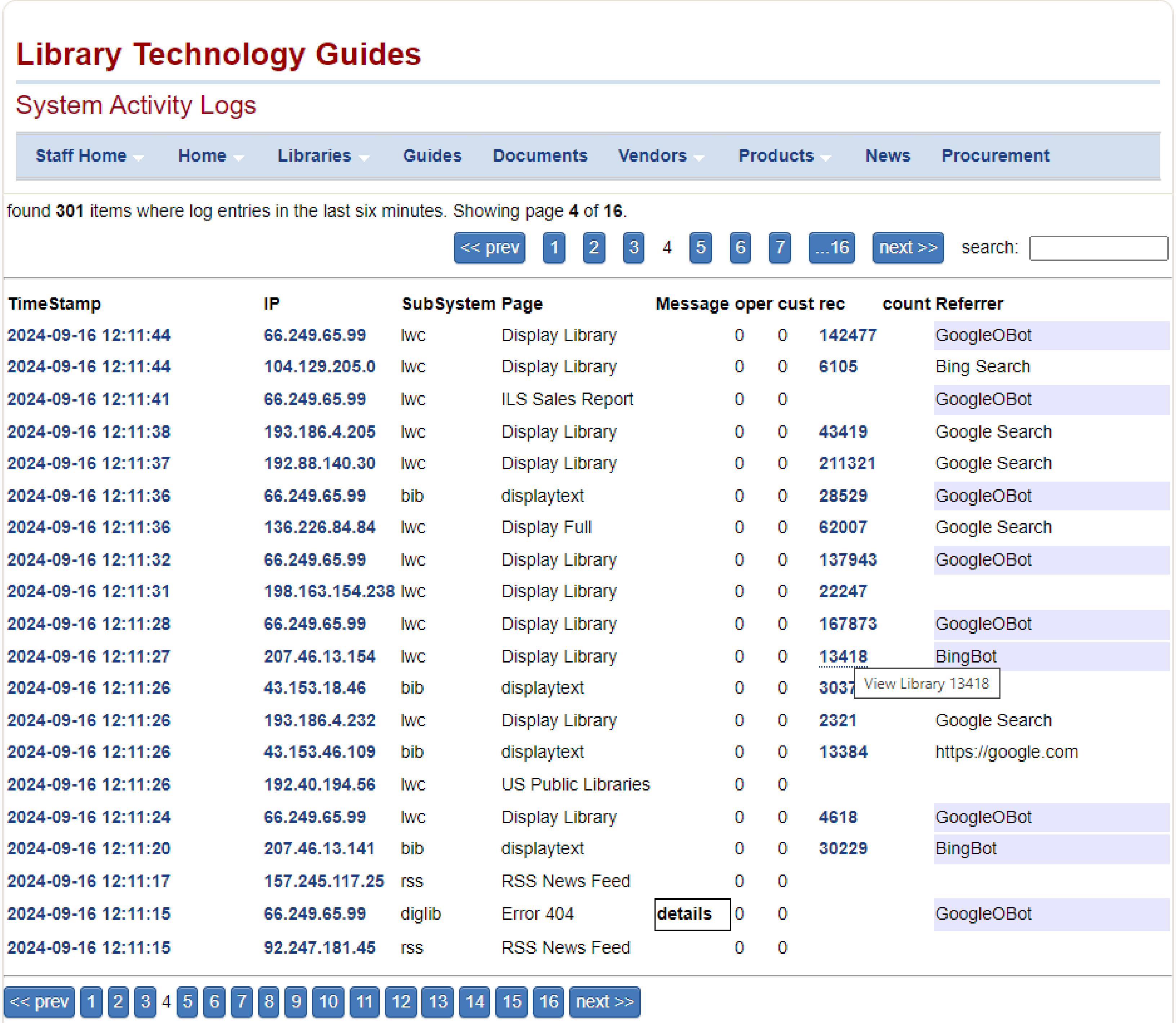
The Sierra ILS has your library covered

Sierra is a **Library Service Platform** designed to make your library effective with a range of seamless integrations. Click here to learn more.



More than 50% of requests are from search engine bots

Classic Google Search presents organic blue links
(as well as sponsored links)



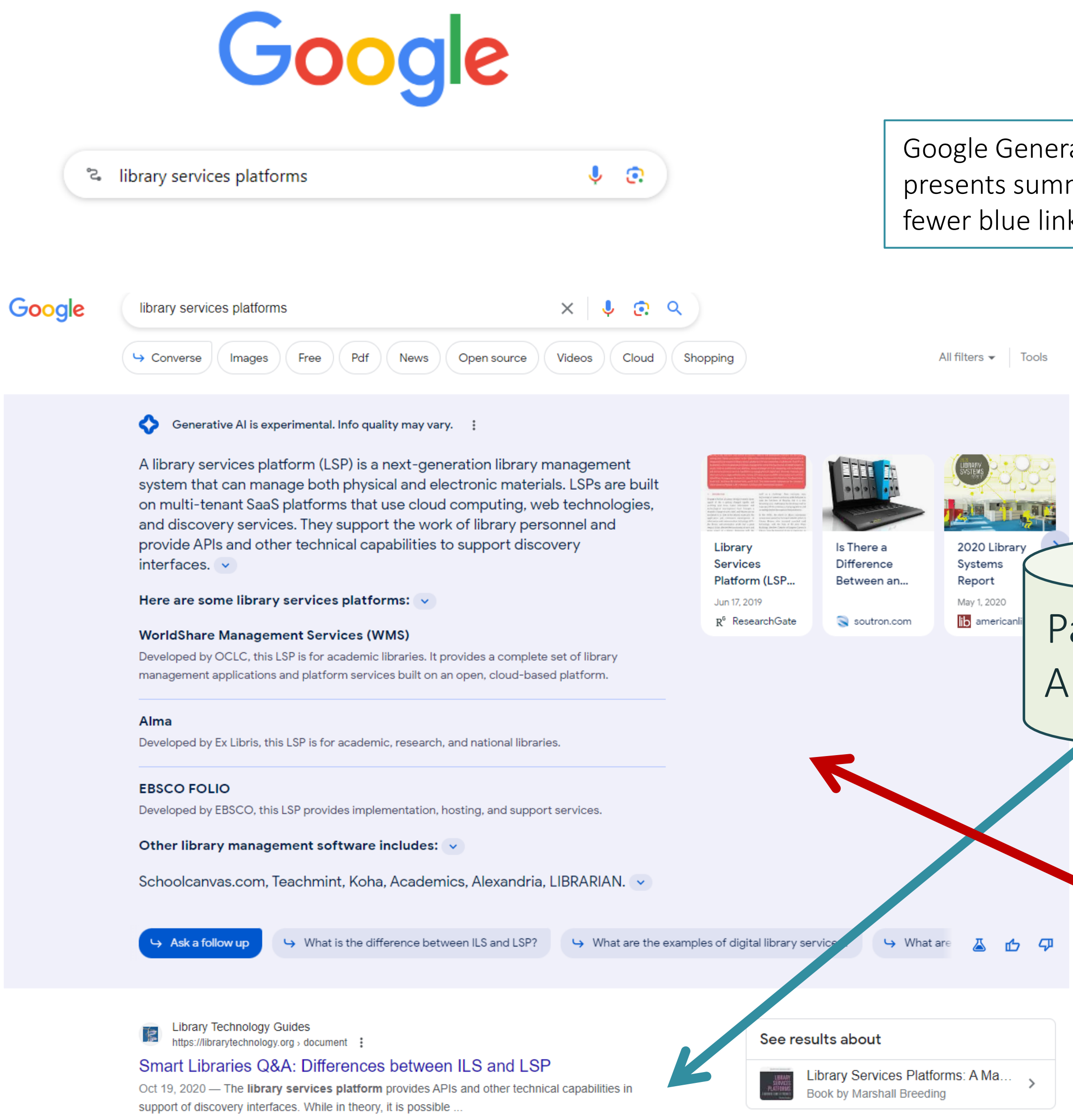
Library Technology Guides

System Activity Logs

found **301** items where log entries in the last six minutes. Showing page **4** of **16**.

TimeStamp	IP	SubSystem	Page	Message	oper	cust	rec	count	Referrer
2024-09-16 12:11:44	66.249.65.99	lwc	Display Library		0	0	142477		GoogleOBot
2024-09-16 12:11:44	104.129.205.0	lwc	Display Library		0	0	6105		Bing Search
2024-09-16 12:11:41	66.249.65.99	lwc	ILS Sales Report		0	0			GoogleOBot
2024-09-16 12:11:38	193.186.4.205	lwc	Display Library		0	0	43419		Google Search
2024-09-16 12:11:37	192.88.140.30	lwc	Display Library		0	0	211321		Google Search
2024-09-16 12:11:36	66.249.65.99	bib	displaytext		0	0	28529		GoogleOBot
2024-09-16 12:11:36	136.226.84.84	lwc	Display Full		0	0	62007		Google Search
2024-09-16 12:11:32	66.249.65.99	lwc	Display Library		0	0	137943		GoogleOBot
2024-09-16 12:11:31	198.163.154.238	lwc	Display Library		0	0	22247		
2024-09-16 12:11:28	66.249.65.99	lwc	Display Library		0	0	167873		GoogleOBot
2024-09-16 12:11:27	207.46.13.154	lwc	Display Library		0	0	13418		BingBot
2024-09-16 12:11:26	43.153.18.46	bib	displaytext		0	0	3037		
2024-09-16 12:11:26	193.186.4.232	lwc	Display Library		0	0	2321		Google Search
2024-09-16 12:11:26	43.153.46.109	bib	displaytext		0	0	13384		https://google.com
2024-09-16 12:11:26	192.40.194.56	lwc	US Public Libraries		0	0			
2024-09-16 12:11:24	66.249.65.99	lwc	Display Library		0	0	4618		GoogleOBot
2024-09-16 12:11:20	207.46.13.141	bib	displaytext		0	0	30229		BingBot
2024-09-16 12:11:17	157.245.117.25	rss	RSS News Feed		0	0			
2024-09-16 12:11:15	66.249.65.99	diglib	Error 404	details	0	0			GoogleOBot
2024-09-16 12:11:15	92.247.181.45	rss	RSS News Feed		0	0			

Library Technology Guides system log. Shaded entries are search or AI bots



library services platforms

library services platforms

Converse | Images | Free | Pdf | News | Open source | Videos | Cloud | Shopping | All filters | Tools

Generative AI is experimental. Info quality may vary.

A library services platform (LSP) is a next-generation library management system that can manage both physical and electronic materials. LSPs are built on multi-tenant SaaS platforms that use cloud computing, web technologies, and discovery services. They support the work of library personnel and provide APIs and other technical capabilities to support discovery interfaces.

Here are some library services platforms:

WorldShare Management Services (WMS)

Developed by OCLC, this LSP is for academic libraries. It provides a complete set of library management applications and platform services built on an open, cloud-based platform.

Alma

Developed by Ex Libris, this LSP is for academic, research, and national libraries.

EBSCO FOLIO

Developed by EBSCO, this LSP provides implementation, hosting, and support services.

Other library management software includes:

Schoolcanvas.com, Teachmint, Koha, Academics, Alexandria, LIBRARIAN.

Library Services Platform (LSP... Jun 17, 2019 ResearchGate

Is There a Difference Between an... soutron.com

2020 Library Systems Report May 1, 2020 americanli

Ask a follow up | What is the difference between ILS and LSP? | What are the examples of digital library servic | What are

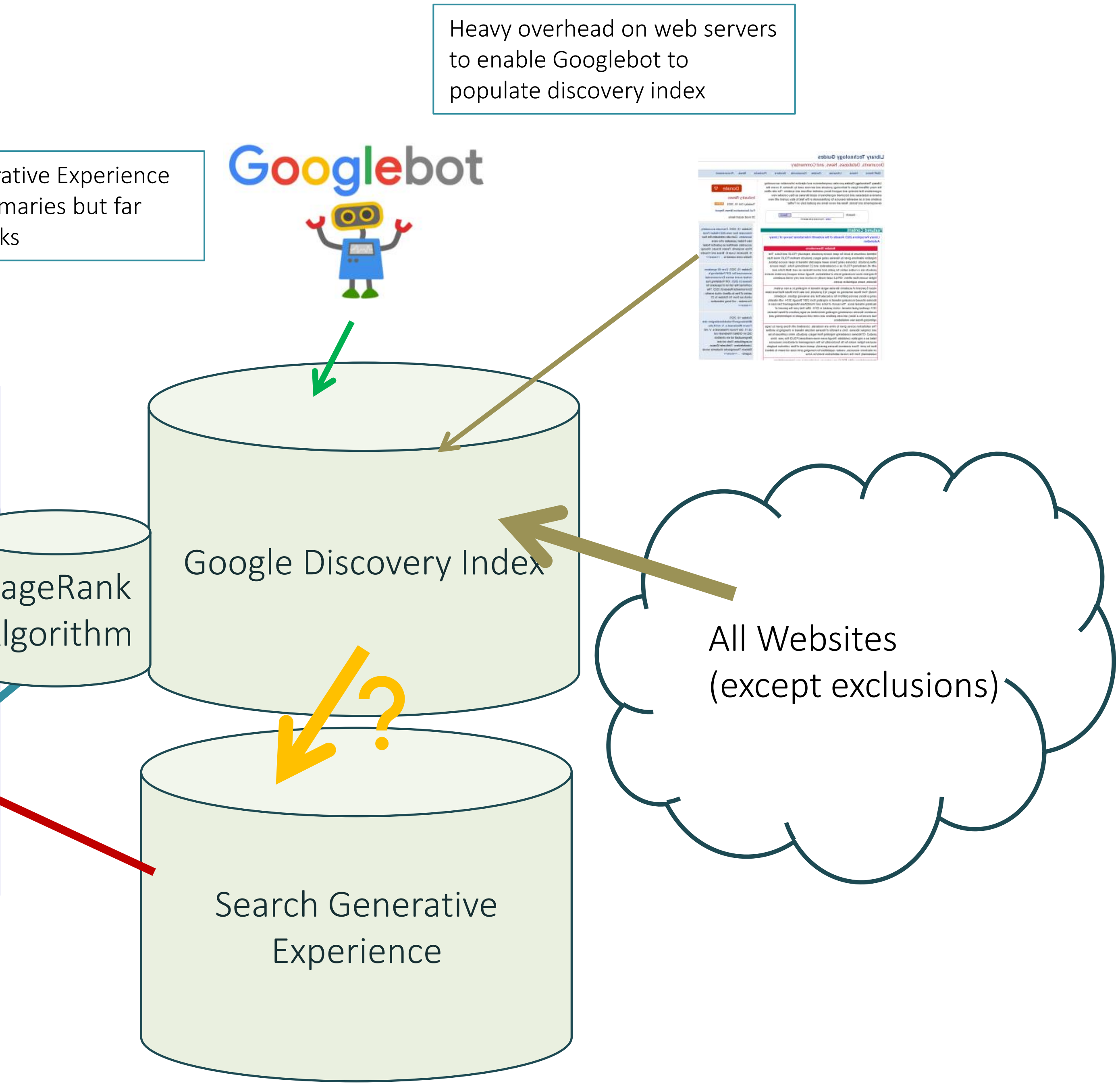
Library Technology Guides
https://librarytechnology.org › document

Smart Libraries Q&A: Differences between ILS and LSP

Oct 19, 2020 — The **library services platform** provides APIs and other technical capabilities in support of discovery interfaces. While in theory, it is possible ...

See results about

Library Services Platforms: A Ma...
Book by Marshall Breeding


Googlebot
PageRank Algorithm
Google Discovery Index
?
All Websites (except exclusions)
Search Generative Experience

Google Generative Experience presents summaries but far fewer blue links

Heavy overhead on web servers to enable Googlebot to populate discovery index

Environmental Impact for AI and Cloud Computing

From ChatGPT-4 "Large-Scale Models: Models with tens or hundreds of billions of parameters, such as GPT-3, require more substantial resources. For instance, GPT-3, with its 175 billion parameters, was trained on a supercomputer-like setup with thousands of NVIDIA V100 GPUs over weeks.

As of my last update in April 2023, training state-of-the-art LLMs could require thousands of GPU or TPU hours, translating into substantial electricity costs and CO2 emissions. The exact figures vary widely based on the specific configurations and efficiency optimizations applied. Training a model like GPT-3 from scratch could cost millions of dollars in cloud computing resources alone."

The Staggering Ecological Impacts of Computation and the Cloud: the Cloud now has a greater carbon footprint than the airline industry. A single data center can consume the equivalent electricity of 50,000 homes. At 200 terawatt hours (TWh) annually, data centers collectively devour more energy than some nation-states. (https://thereader.mitpress.mit.edu/the-staggering-ecological-impacts-of-computation-and-the-cloud/)

Note: Deploying and training LLMs consumes more computational resources than ongoing use. Incremental differences between general search engines, AI enhances search, and conversational AI services.

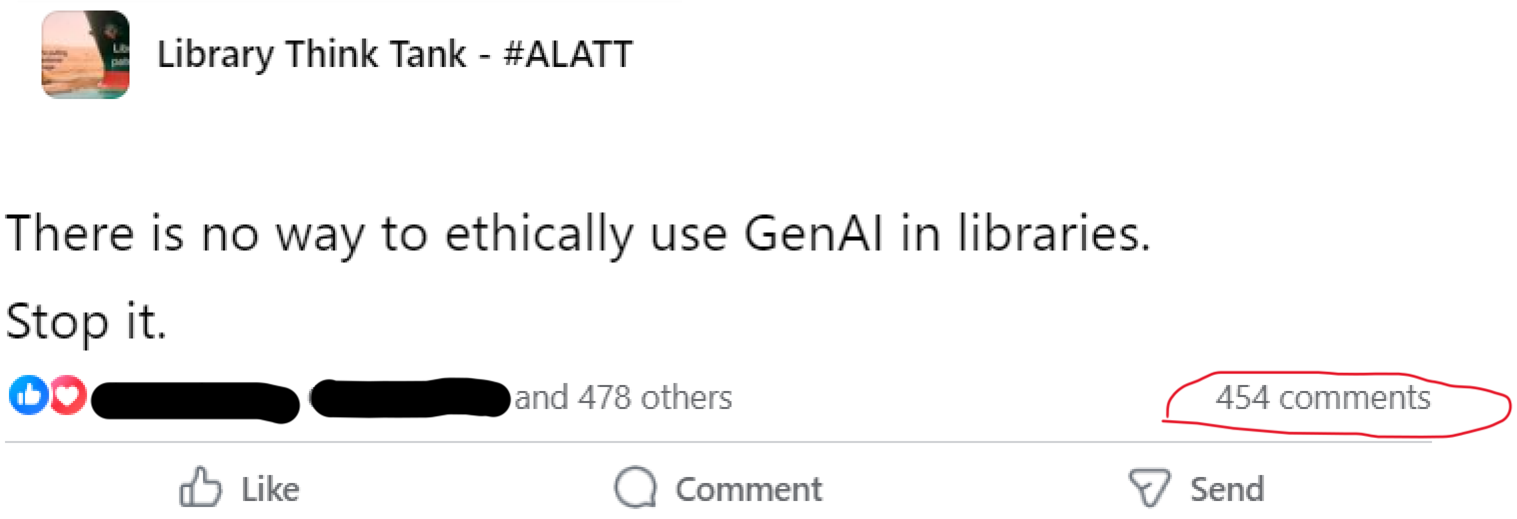
Potential Problems

Bais: can reinforce stereotypes

Hallucinations: factual errors

Attribution or Consent: does the LLM have permission to use all resources in its training data

How do users use generated content ethically or legally? (especially in scholarly and educational settings)



Library Think Tank - #ALATT

There is no way to ethically use GenAI in libraries.

Stop it.

[redacted] [redacted] and 478 others

454 comments

AI and Libraries

Many of the flaws of LLMs are inconsistent with library use, which requires objective, factual, and reliable outcomes

Library-specific use of Generative AI requires construction of additional mechanisms to ensure accuracy and mitigate bias.

Scholarly and other reliable sources to feed appropriate results to LLM which then generates responses

Prompt frameworks provide guardrails for factual outcomes

Ability to cite specific sources within training data as part of responses

Designed to eliminate or minimize hallucinations and bias

Generative AI in Academica

Generative AI is already being used to (co)-write scholarly papers, perform supporting tasks, and in the review and publication process

AI tools can streamline the work of researchers, administrators, and students, but must be used within ethical and bounds

Use of ChatGPT and other Generative AI services is widespread

- Student papers > peer-reviewed scholarly papers

Difficult or impossible to detect: Many false positives

Institutions increasingly develop policies on appropriate use

Most Scholarly publishers have policies on how generative text can be incorporated and cited.

Automated AI detection has not been proven to be reliable: false positives can be catastrophic for a student’s academic career.

Libraries respond to AI

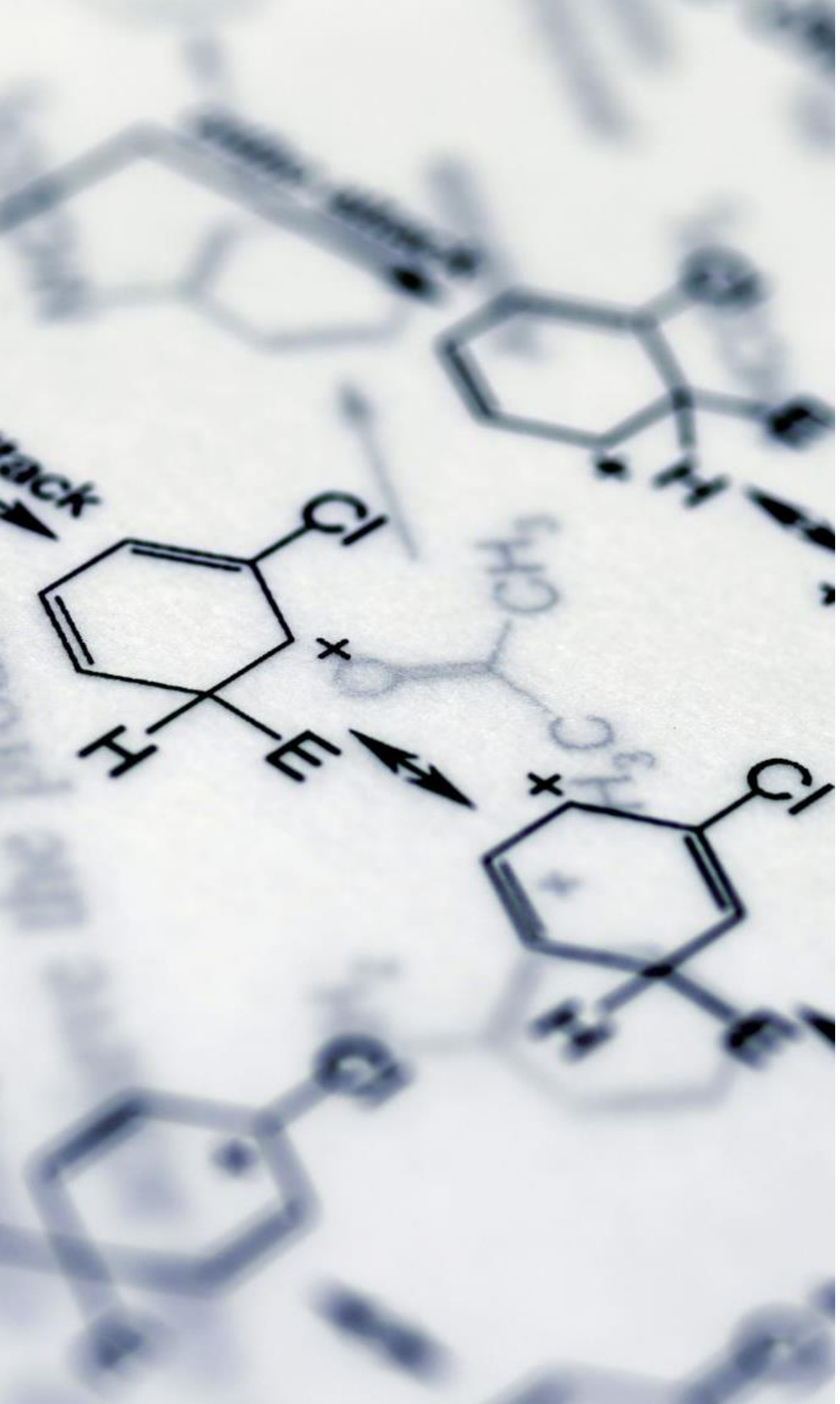
Extend scope of bibliographic instructional to include AI

Provide perspective on information literacy to include new challenges brought by generative AI

Guide users to AI tools that can improve processes for research and writing

Incorporate AI to increase capacity for collection management

Just like any new automation tool: helps the library to direct their efforts to more impactful activities as it streamlines routine work.



Scholarly AI Tools

ScholarGPT: developed by XPixel

Dimensions Research GPT (ChatGPT overlay)

AskYourPDF: plugin for ChatGPT

ScholarAI: powered by GPT-4 Turbo

Hundreds of others – new tools continually announced

Retrieval augmented Generation

A framework model to create generative AI responses based on information retrieved from reliable resources

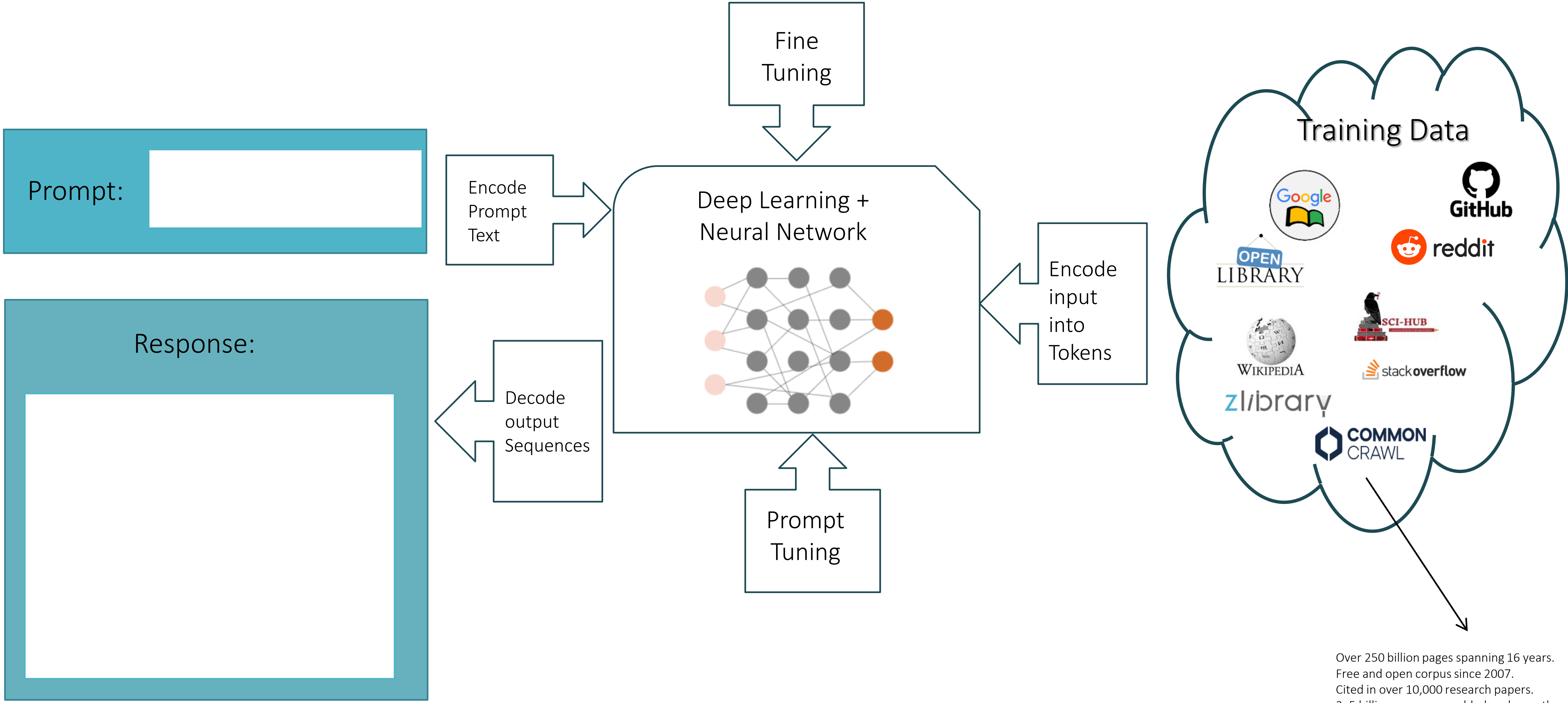
Selected documents or statements from knowledge base are presented to the context window of the LLM

Avoids relying on the LLM for factual statements

The training data underlying the LLM incudes unreliable information that should not be re-presented through the RAG framework

Leverages the LLM for generation of text

Generic Large Language Model



Prompt:

Response:

Encode Prompt Text

Deep Learning + Neural Network

Fine Tuning

Prompt Tuning

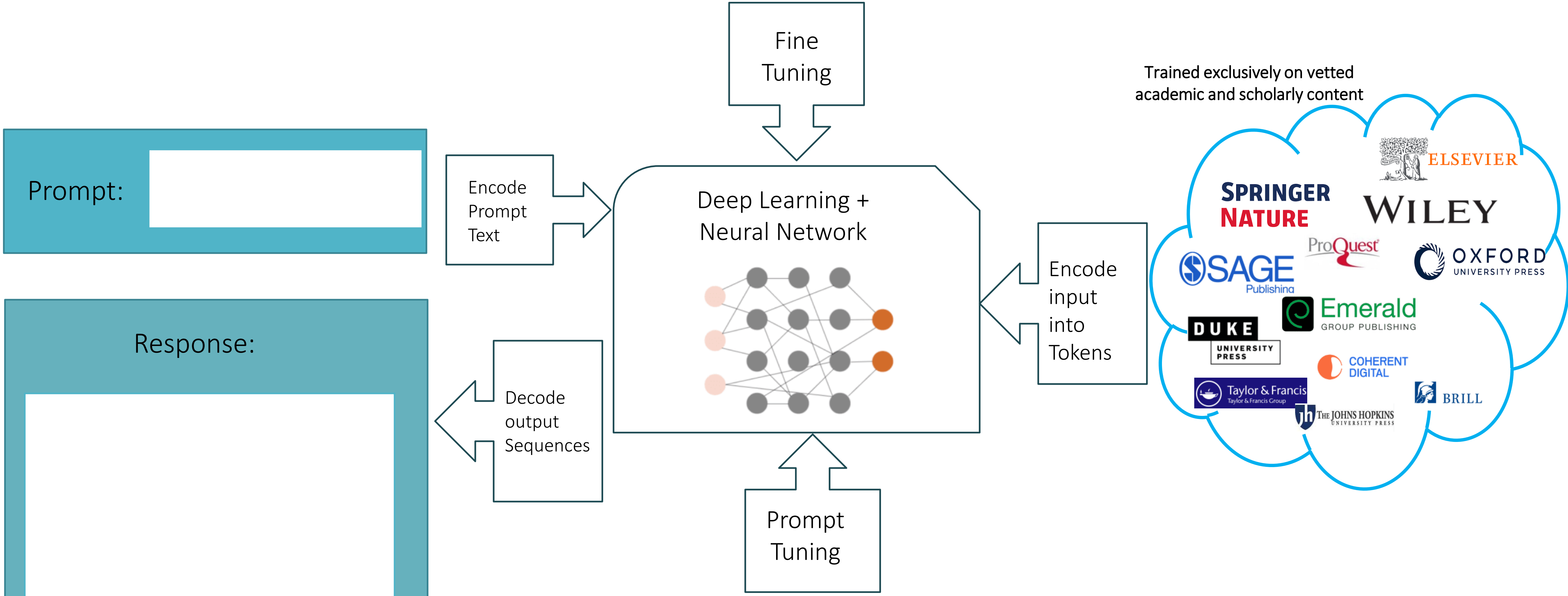
Encode input into Tokens

Decode output Sequences

Training Data

Over 250 billion pages spanning 16 years.
Free and open corpus since 2007.
Cited in over 10,000 research papers.
3–5 billion new pages added each month.
Primary training corpus in every LLM.
82% of raw tokens used to train GPT-3.
https://commoncrawl.org/

Library-specific Large Language Model?



Prompt:

Response:

Encode Prompt Text

Fine Tuning

Deep Learning + Neural Network

Encode input into Tokens

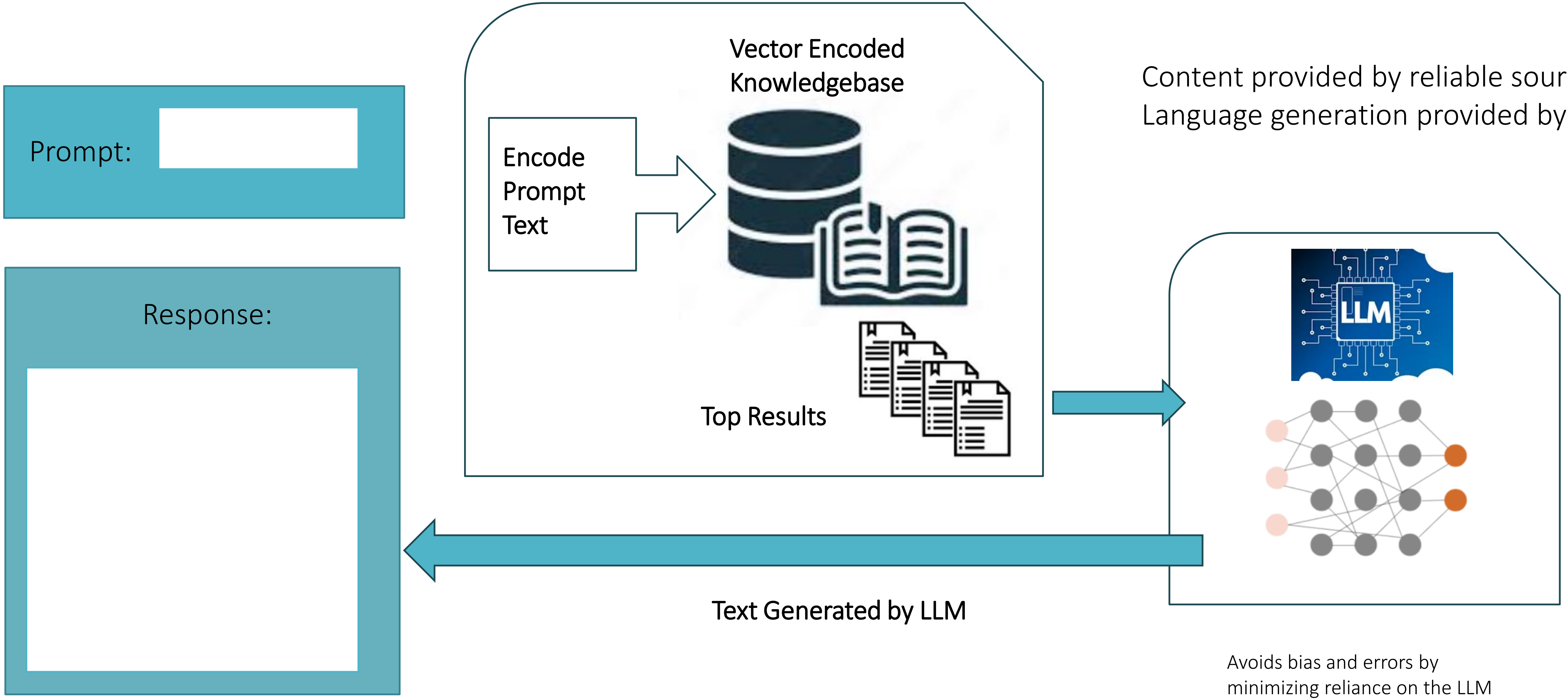
Decode output Sequences

Prompt Tuning

Trained exclusively on vetted academic and scholarly content

Not necessarily a practical alternative since it would be extremely costly to build and maintain

Retrieval Augmented Generation



Prompt:

Response:

Vector Encoded Knowledgebase

Encode Prompt Text

Top Results

Text Generated by LLM

Content provided by reliable sources
Language generation provided by LLM

Avoids bias and errors by minimizing reliance on the LLM for facts and content

AI Powered Library Discovery

Beyond current index-based discovery services

Ongoing access to full text of scholarly articles based on keyword retrieval and relevancy ranking

New capabilities to return results based on concepts even when not found in the query text.

Cross-language searching

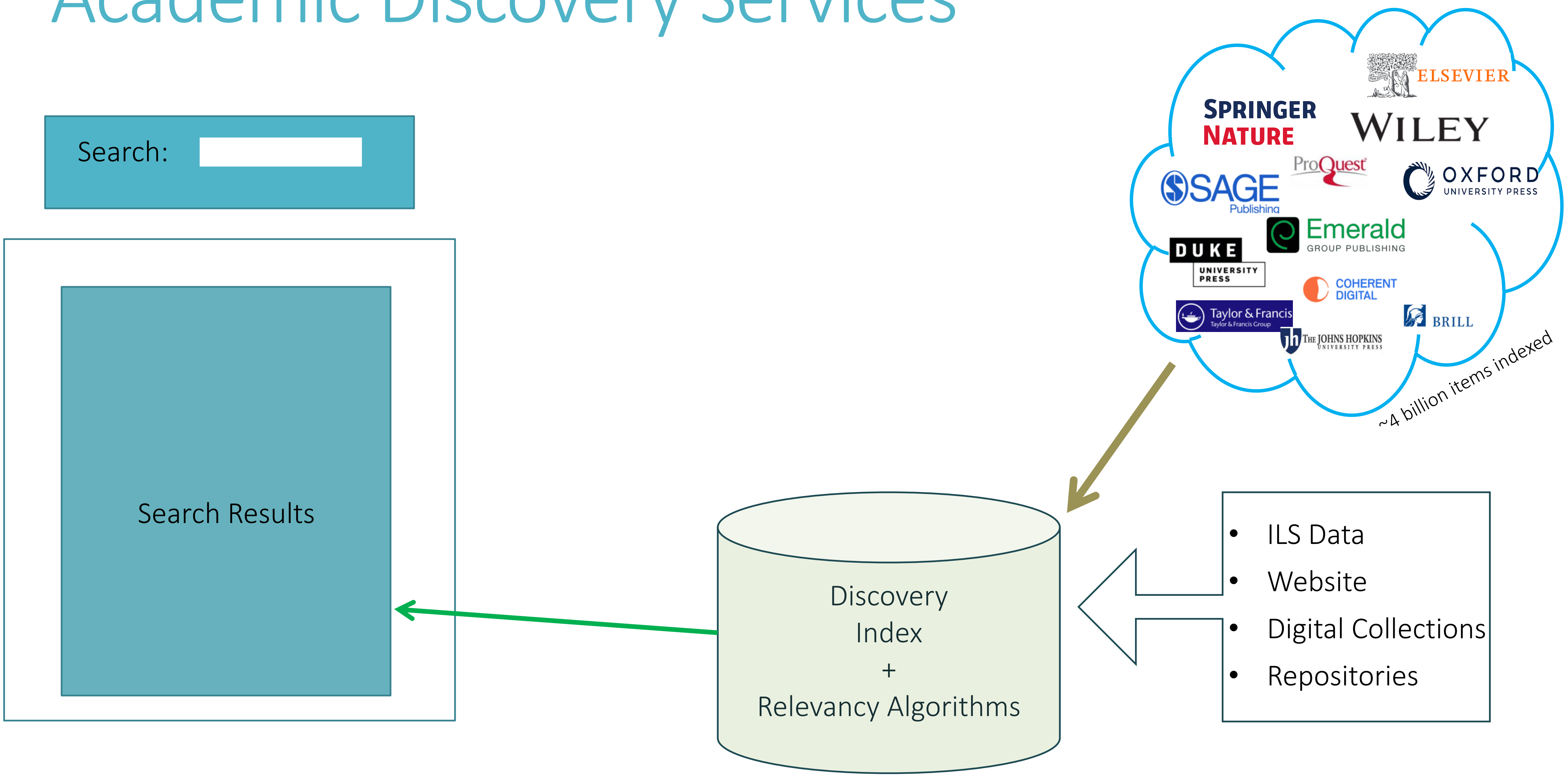
Summary and interpretation of results

Ability to explore topics with new tools and interfaces

Extract and cite relevant portions of text for incorporation in research papers

Many other capabilities that streamline user research and writing

Academic Discovery Services



~4 billion items indexed

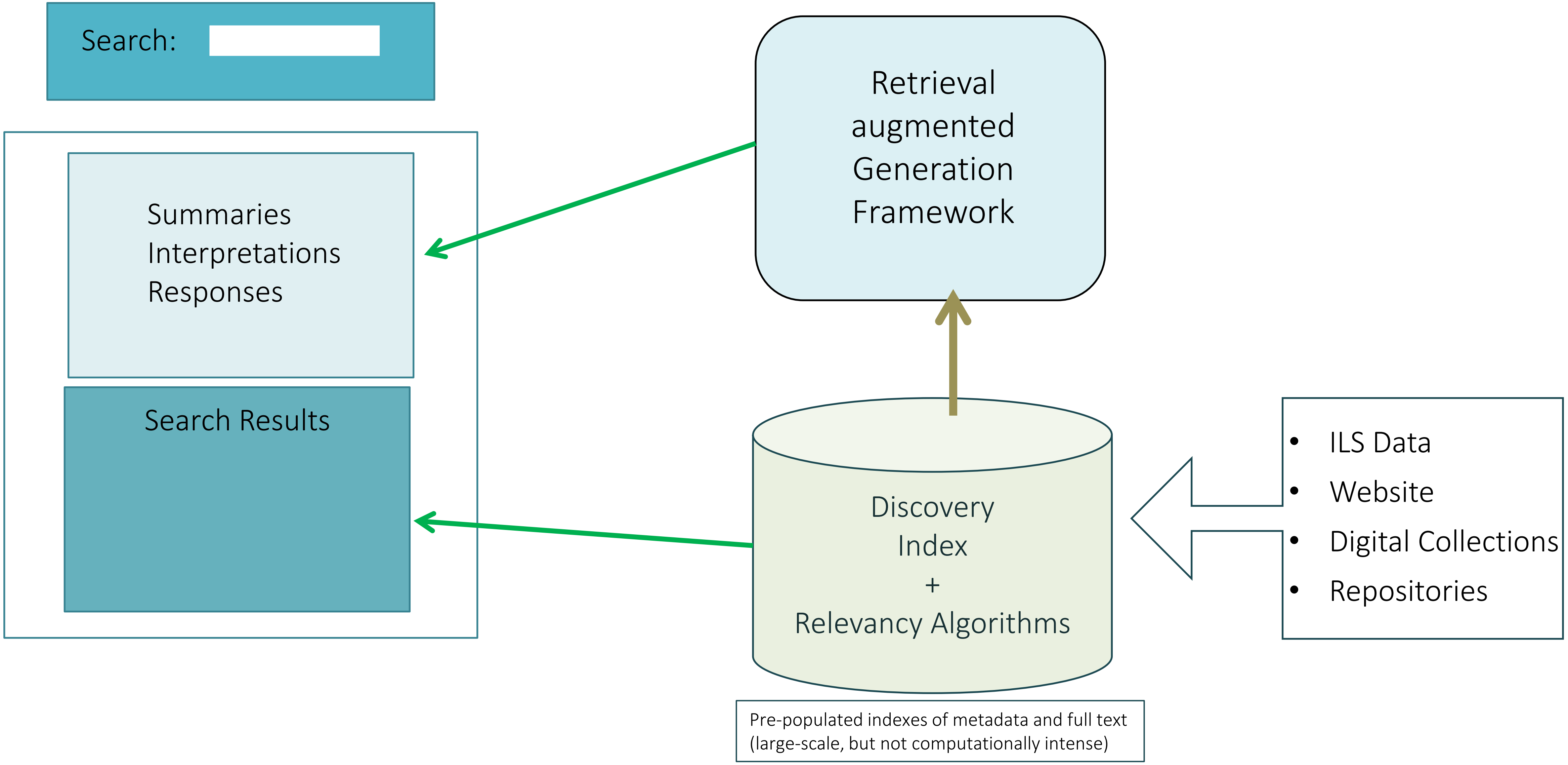
Search:

Search Results

Discovery
Index
+
Relevancy Algorithms

- ILS Data
- Website
- Digital Collections
- Repositories

Academic Discovery Services



Search:

Summaries
Interpretations
Responses

Search Results

Retrieval augmented Generation Framework

Discovery Index
+
Relevancy Algorithms

- ILS Data
- Website
- Digital Collections
- Repositories

Pre-populated indexes of metadata and full text (large-scale, but not computationally intense)

AI-powered Library Chat

Feasible to create conversational chat services based on native generative AI services geared to operational questions: hours, locations, policies, personnel.

Less feasible for chat services to respond to research questions that require live access to the web and online resources

RAG architecture can be used with Library Chat services for research questions if the framework includes access to adequate reference and research resources

Must be thoroughly tested to ensure high quality results

Use by library workers

Library patrons will increasingly be using AI. Librarians will need to understand these technologies to provide professional services

Use AI-based tools to improve productivity in selected areas of work

Incorporate in professional work and administrative tasks

- Specific tasks that can be reviewed and revised before incorporating into official work products

Using AI in Library Technical Services

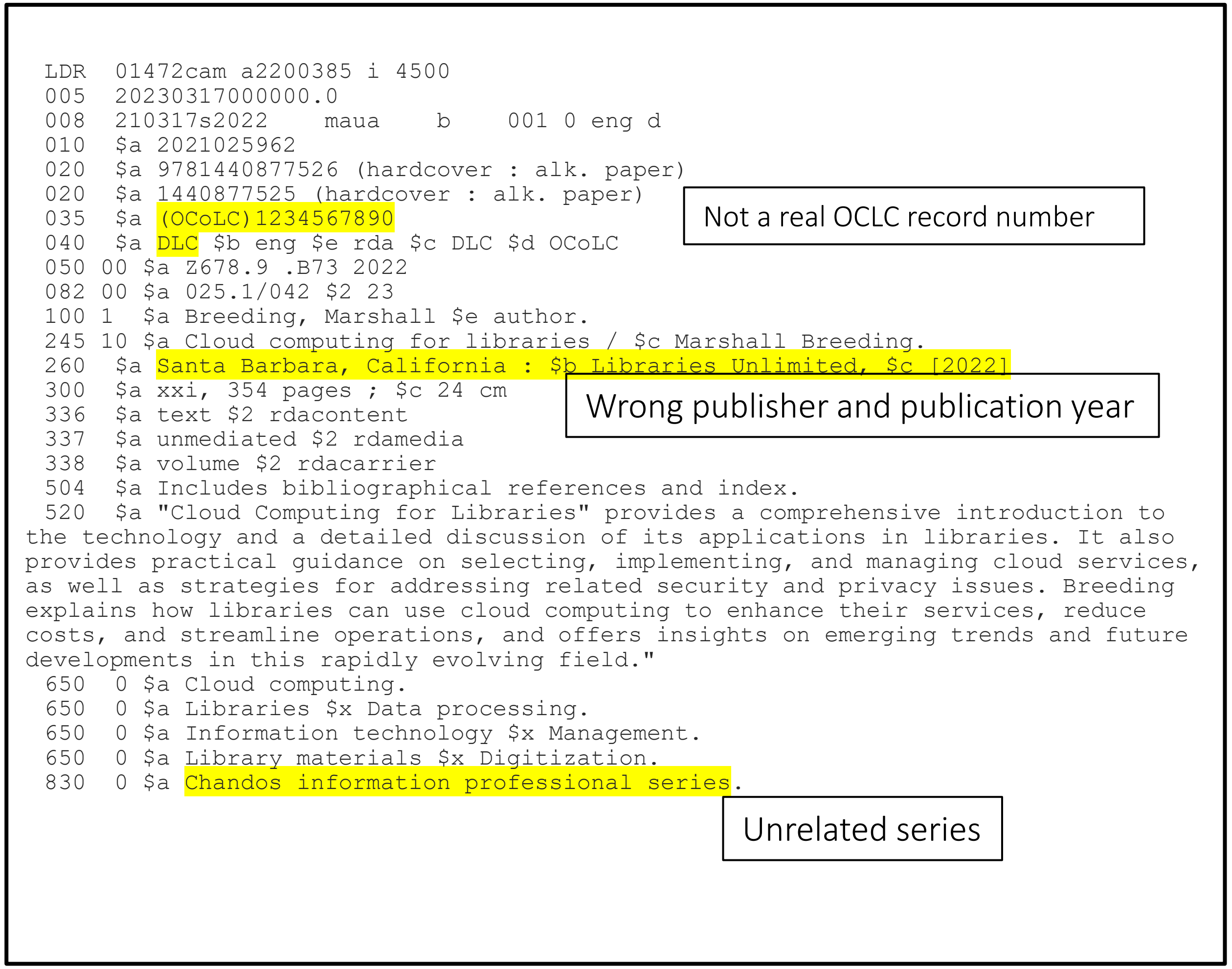
Can Generative AI be used to support or replace aspects of cataloging and metadata services?

Current versions of Generative AI are not capable of professional-level creation of MARC records but can make impressive looking records.

More targeted AI technologies may be able to perform more complete and accurate cataloging based on official documentation and a body of high-quality bibliographic records for reference.

MARC Record produced by ChatGPT

GPT-3



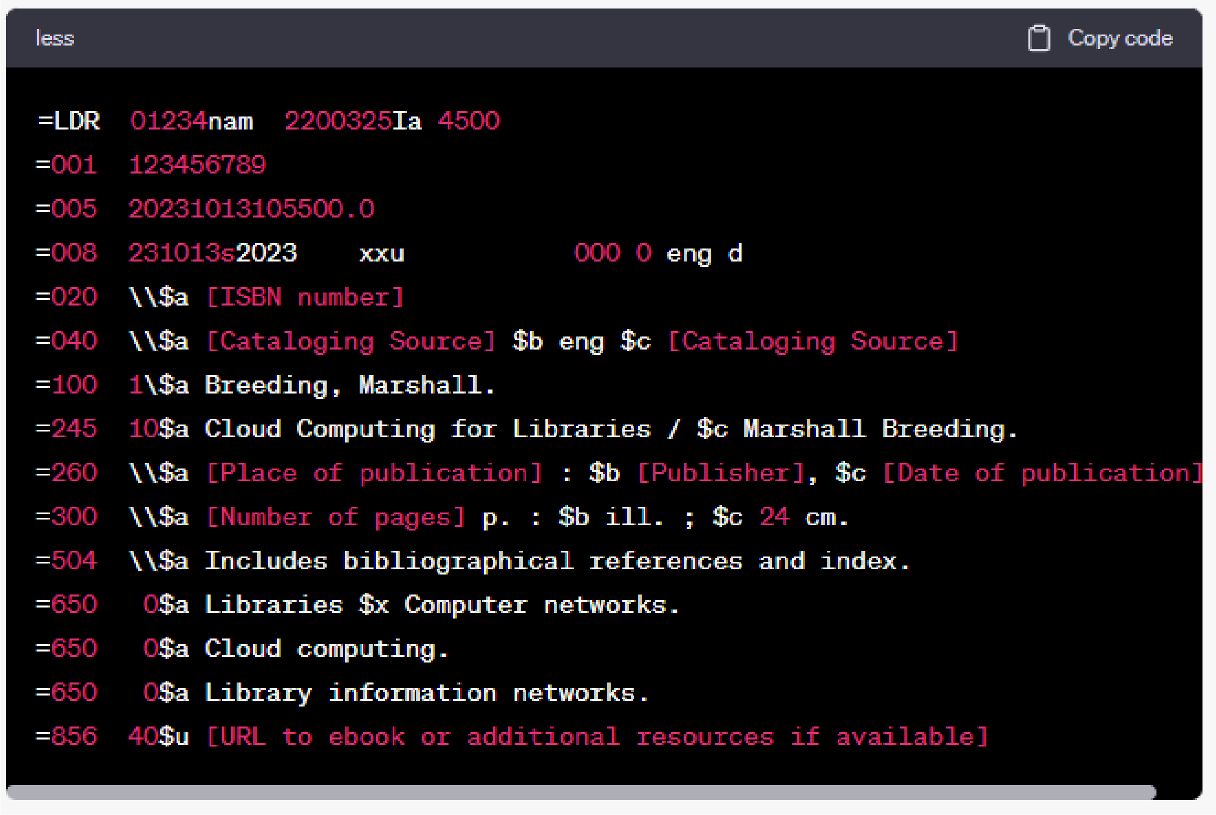
```
 LDR  01472cam a2200385 i 4500
 005  20230317000000.0
 008  210317s2022    maua    b    001 0 eng d
 010  $a 2021025962
 020  $a 9781440877526 (hardcover : alk. paper)
 020  $a 1440877525 (hardcover : alk. paper)
 035  $a (OCoLC)1234567890
 040  $a DLC $b eng $e rda $c DLC $d OCoLC
 050 00 $a Z678.9 .B73 2022
 082 00 $a 025.1/042 $2 23
 100 1  $a Breeding, Marshall $e author.
 245 10 $a Cloud computing for libraries / $c Marshall Breeding.
 260  $a Santa Barbara, California : $b Libraries Unlimited, $c [2022]
 300  $a xxi, 354 pages ; $c 24 cm
 336  $a text $2 rdacontent
 337  $a unmediated $2 rdamedia
 338  $a volume $2 rdacarrier
 504  $a Includes bibliographical references and index.
 520  $a "Cloud Computing for Libraries" provides a comprehensive introduction to
the technology and a detailed discussion of its applications in libraries. It also
provides practical guidance on selecting, implementing, and managing cloud services,
as well as strategies for addressing related security and privacy issues. Breeding
explains how libraries can use cloud computing to enhance their services, reduce
costs, and streamline operations, and offers insights on emerging trends and future
developments in this rapidly evolving field."
 650  0 $a Cloud computing.
 650  0 $a Libraries $x Data processing.
 650  0 $a Information technology $x Management.
 650  0 $a Library materials $x Digitization.
 830  0 $a Chandos information professional series.
```

Not a real OCLC record number

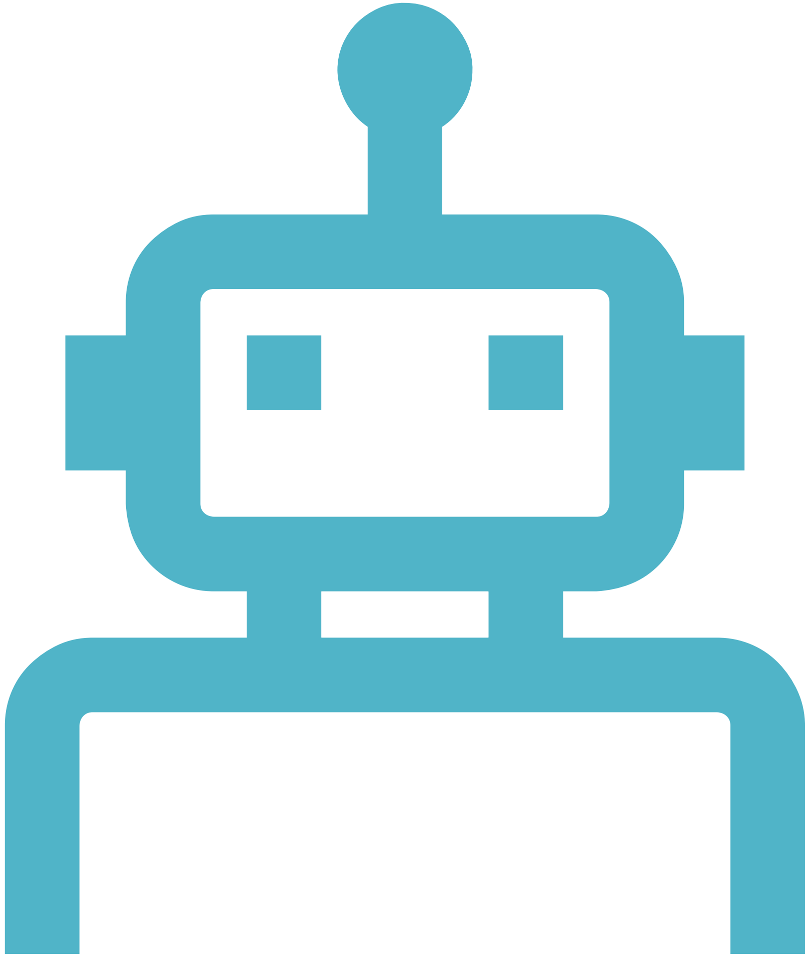
Wrong publisher and publication year

Unrelated series

GPT-4



```
=LDR  01234nam  2200325Ia 4500
=001  123456789
=005  20231013105500.0
=008  231013s2023    xxu            000 0 eng d
=020  \\$a [ISBN number]
=040  \\$a [Cataloging Source] $b eng $c [Cataloging Source]
=100  1\$a Breeding, Marshall.
=245  10$a Cloud Computing for Libraries / $c Marshall Breeding.
=260  \\$a [Place of publication] : $b [Publisher], $c [Date of publication]
=300  \\$a [Number of pages] p. : $b ill. ; $c 24 cm.
=504  \\$a Includes bibliographical references and index.
=650   0$a Libraries $x Computer networks.
=650   0$a Cloud computing.
=650   0$a Library information networks.
=856  40$u [URL to ebook or additional resources if available]
```



Cataloging Copilot?

Cataloging, like computer programming, involves detailed coding, following complex rules

Requires extensive expertise and training

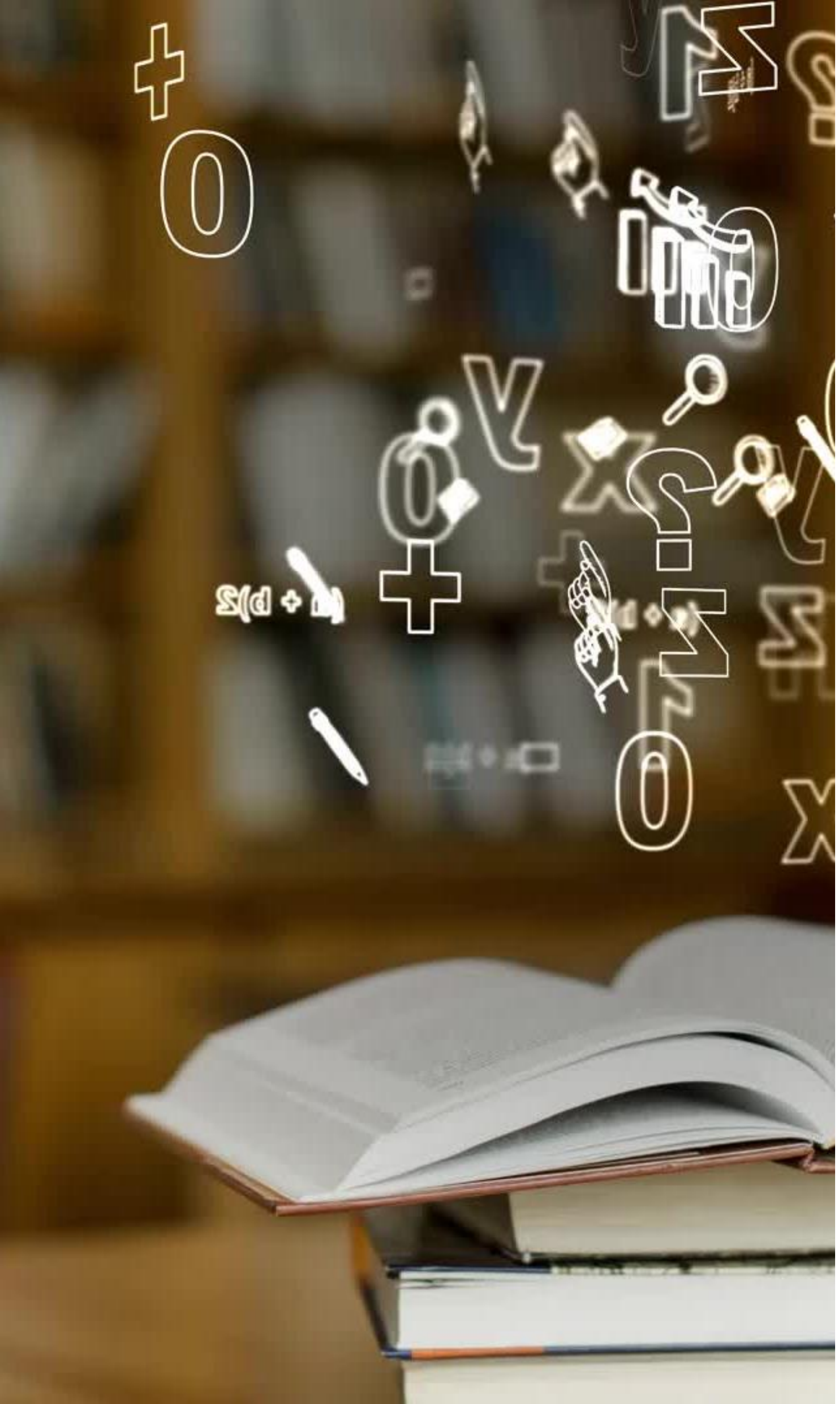
Current AI services cannot produce adequate bibliographic records

Can AI technologies be used to create a Copilot environment to accelerate productivity and to help less experienced catalogers produce high-quality records?

Possible training inputs would include official documentation: Cataloger’s Desktop, RDA, AACR2, BIBFRAME, subject headings (LC, Dewey, Sears, etc.), LC Name Authorities, etc.

Generate bibliographic records in desired formats (MARC21, BIBFRAME, etc.) that could be reviewed by cataloging personnel.

Also expect capabilities to be designed that can automatically generate metadata for digital objects based on full text using summarization, concept extraction, and image recognition.



Looking ahead

AI has already disrupted many existing trends, in both positive and negative ways

Libraries are increasingly engaged with AI technologies and projects. They are experimenting with locally developed tools and implementing vendor-provided services

Libraries are leading discussions surrounding ethics and AI and dealing with the fallout of misinformation.

The use of AI-powered services is increasingly widespread. AI literacy is an important component of digital literacy and bibliographic instruction.

Libraries are engaging with vendors to imagine AI-based services consistent with academic values and public interest.

AI will increasingly power the next level of library automation. Essential to use any available tools to work efficiently and deliver expected services.

Libraries can use AI to make a positive difference to counter misinformation and bias that increasingly pervades society.