

Reimagining Discovery

Transforming Access to Collections with AI-Driven Exploration

CAROLYN CAIZZI, AMY DESCHENES, & STU SNYDMAN

CNI, DECEMBER 2024



SLIDES

Our Challenge

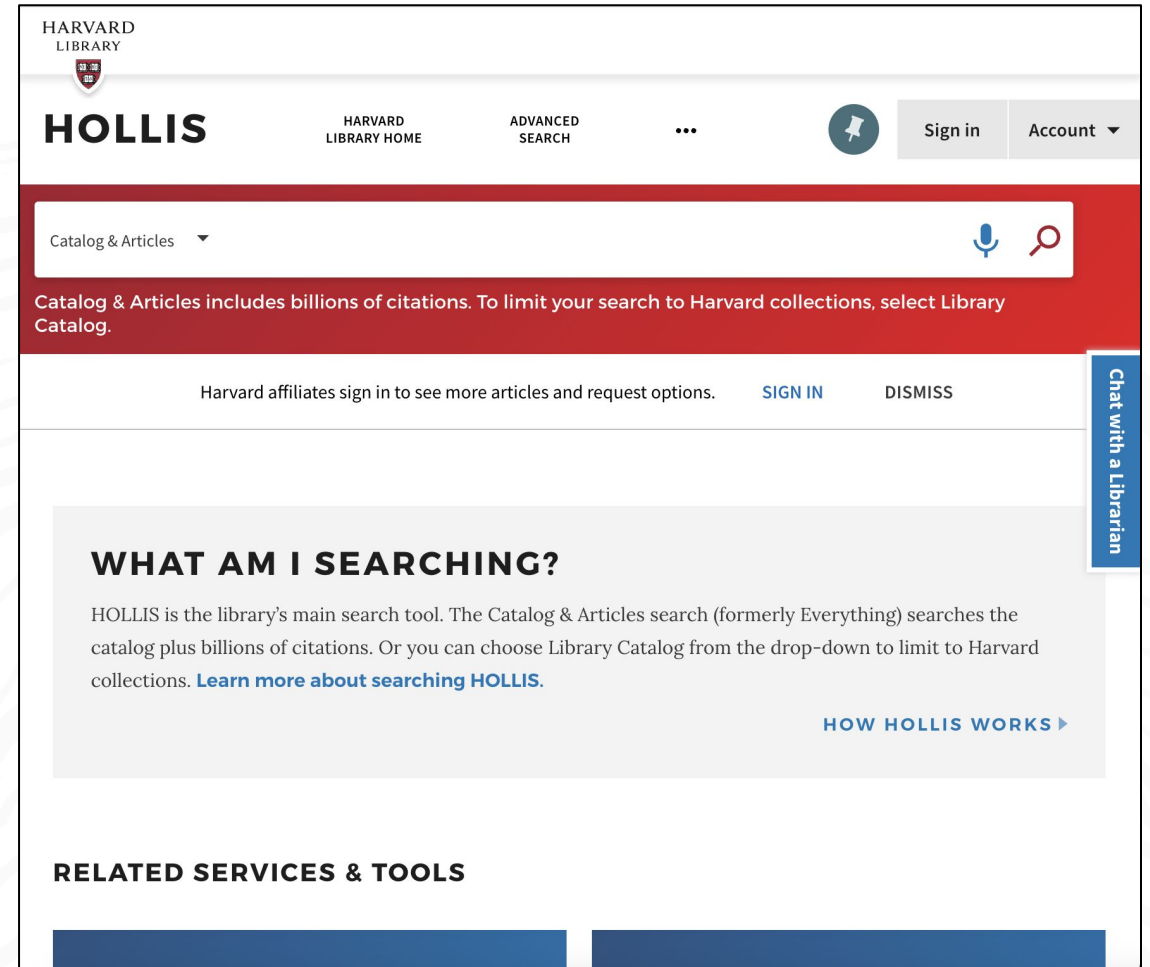
The grand challenge of the Harvard Library is to enable researchers and students to find the right content to support world-class research, teaching and learning, amidst a massive and ever-growing sea of information.

And to ensure users have that the Library is connecting them to *trustworthy* resources.



Scope of Library Discovery at Harvard (in part)

- 1 billion journal articles
- 14M print titles, representing 19M physical volumes (also includes scores, maps, audio/video, etc.)
- 9M ReCAP partner titles – print
- 4M ebooks & journals
- 2M images
- 15K archival finding aids
- 10K geospatial layers



The screenshot shows the Harvard HOLLIS search interface. At the top, the Harvard Library logo is on the left, and navigation links for 'HOLLIS', 'HARVARD LIBRARY HOME', and 'ADVANCED SEARCH' are in the center. On the right, there are 'Sign in' and 'Account' buttons. Below the navigation is a search bar with a dropdown menu set to 'Catalog & Articles'. A red banner below the search bar states: 'Catalog & Articles includes billions of citations. To limit your search to Harvard collections, select Library Catalog.' Below this is a message for Harvard affiliates to sign in. The main content area features a section titled 'WHAT AM I SEARCHING?' which explains that HOLLIS is the library's main search tool and provides instructions on how to use the search filters. A 'Chat with a Librarian' button is visible on the right side of the interface.

Feedback on Current Discovery Environment

"It's intimidating for new people who have never browsed a special collection to even know where to begin, so a way to access a search tool easily from the library website might help a lot."

- Faculty member, Biology

"Streamlining HOLLIS could go a long way. The main reason why I don't consult many materials as often as I know I could is because I get overwhelmed with some of the inconsistencies in how information appears."

- PhD Student, Music

"Finding Harvard materials is very onerous. Images in particular are a nightmare. I much prefer to get images directly from Google."

- Faculty member, History of Art & Architecture

Applying AI to Library Discovery



Searching by keywords and returning a list of links is antiquated and unsatisfying



Users will increasingly expect to interact with search via natural language rather than by keywords



Results will be context-aware answers



Talk with HOLLIS (October 2023-January 2024)

An experimental generative AI chat interface to search HOLLIS that:



Uses natural language conversation instead of keywords and interprets semantic search input



Returns a concise, context-aware answer



Locates books on the shelf and provides information about their availability

Our project goals included:



Build skills on our team in implementing generative AI tools for library applications



Build a prototype to help learn from users



Develop a pathway to the next generation of library search at Harvard

Reimagining Discovery (3 year project)

INNOVATE & BUILD

- Discovery system for distinctive collections (*Collections Explorer*)
- Integrate semantic search, natural language processing, & generative AI

REPLACE

- *Harvard Digital Collections & HOLLIS Images*

UPGRADE & ENHANCE

- Update *HOLLIS for Archival Discovery* to Arclight
- Update LibraryCloud to enhance machine access to collections
- Enhance & integrate full-text searching everywhere
- Enhance metadata creation with AI (handwriting recognition, speech to text, & image tagging)

IMPACTED SYSTEMS

All library discovery systems will be involved in this project.

CONTENT TYPES

Text
Single Images
Page-turned Objects
Audio (w/transcripts)
Video (w/transcripts)
Spatial Data
Born-Digital
Research Data
Publications
3D/XR

Adjacent Work

DRS	Virtual Reading Rooms
Viewers & Players	Handwriting Recognition R&D
Streaming Delivery Service	Standardized Rights
IIIF/MPS	Statements

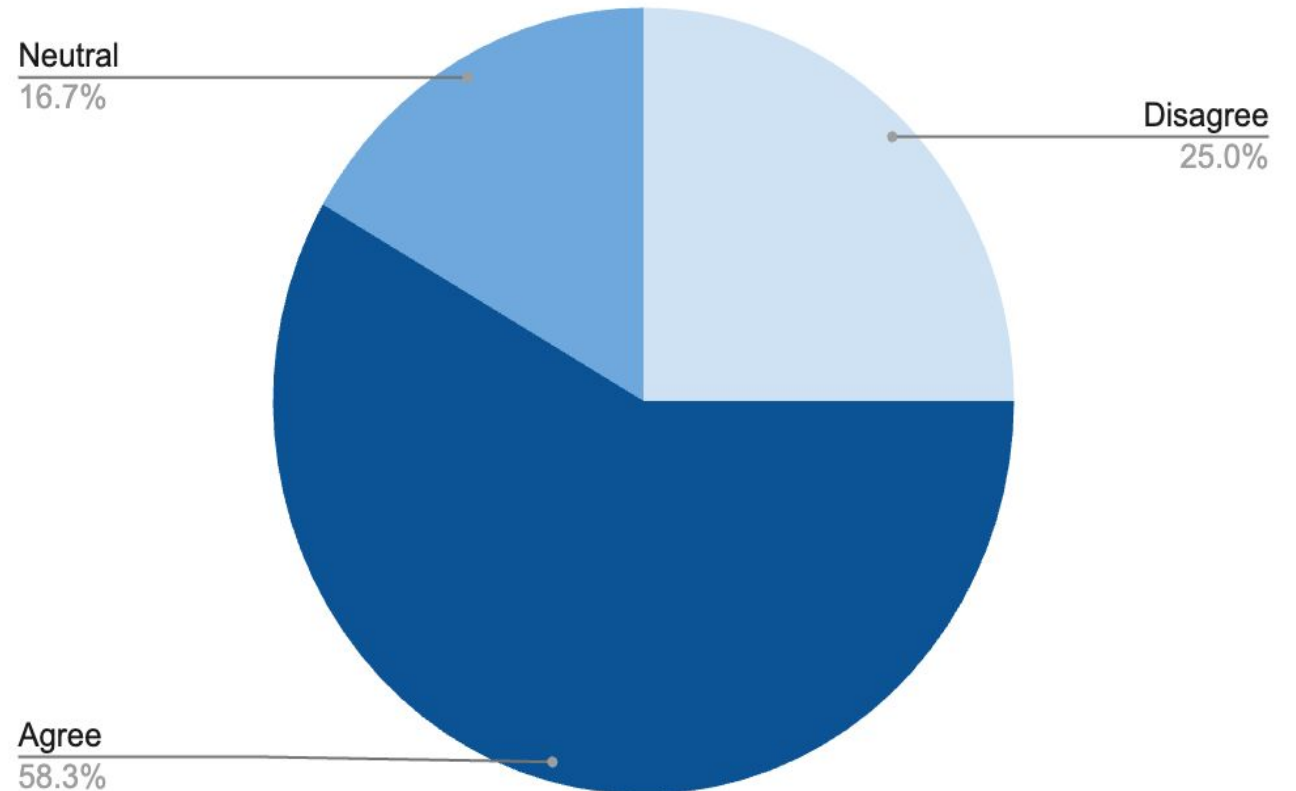
Research & Strategy

Influence of ChatGPT on search preferences

Students & researchers are **using traditional search tools less often**, in favor of ChatGPT and other search tools that support natural language queries.

Generative AI chatbots reduce my reliance on traditional research tools
(like Google or HOLLIS)

From URC Fall 2024 Survey on AI
Use with 200+ students



Exploratory Research & Planning

Research conducted in 2023 with special collections users and on Talk with HOLLIS.

3 Key Findings from the User Research

1. Users want more ***explanation*** of why these results are in the output.
2. Users want ***recommendations*** for related materials or search terms.
3. New special collections researchers are ***unsure how to start***.

Mozilla.AI
Partnership at the
beginning of the project

Product Strategy for *Collections Explorer*

1) **Audience**

A person who wants to take their first step to easily explore special collections.

2) **Goals**

Conduct natural language searching of special collections; offer AI-generated related prompts and item-level summaries; access special collections materials.

3) **Value proposition**

This tool helps you take the first step in finding Harvard Library's unique materials.

4) **Validation**

Prompt examples; guiding text on homepage; AI features.

5) **Measurement**

Analytics on UI elements to measure use; successful usability testing.

User Centered Work

Keeping the work user-centered will ensure that we build something usable and useful.

- Embedded **feedback survey**
- **User testing** for usability and accessibility with Harvard ID holders and members of the public
- **Analytics & metrics** on key features
- **Integrated** student representatives for insights and outreach

Product Insights Interns

Embedded Students

- Meet every other week with project team members and also complete asynchronous tasks.
- Mix of special collections and casual browsers.
- Compensated as student workers.
- Give feedback on preliminary designs.



UX Design

HARVARD LIBRARY

Collections Explorer

What does Harvard have on the history of germ theory development? →

This system is experimental. We recommend you double-check the responses.

Include individual items in my results

DIGITAL COLLECTION

Contagion

May include Digitized Materials

Why Did I Get This Result? ▼

DIGITAL COLLECTION ITEM
Hancock family papers

Contagion

May include Digitized Materials

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc porta ante id diam vestibulum vehicula eu quis nisl. Donec imperdiet ut nulla a malesuada. Nam ac lacus vestibulum, ultricies erat eu, dictum ipsum. In et dui id magna ornare laoreet. Vivamus efficitur feugiat ipsum, nec tincidunt quam gravida nec. Curabitur condimentum velit et imperdiet

Why Did I Get This Result? ▼

FINDING AID

COVID-19 Response ephemera

May include Digitized Materials

Why Did I Get This Result? ▼

FINDING AID ITEM
Hancock family papers

Letter by Secretary of International Medical Congress, drawing Attention to Address by Professor Pasteur, relative to Germ Theory in connection with Inoculation of Animals as Means of protecting them against Virulent Diseases

May include Digitized Materials

Why Did I Get This Result? ▼

FINDING AID ITEM
Hancock family papers

The inhibitory action of Listerine. St. Louis: Lambert Pharmacal Co., 1908.

May include Digitized Materials

Why Did I Get This Result? ▼

COLLECTION WEBSITE

Why Did I Get This Result? ▼

YOU MIGHT ALSO TRY

How did public education campaigns of the 19th century intersect with germ theory?

What are techniques for antiseptic surgery?

How does animalcular theory relate to germ theory?

How did germ theory change healthcare?

What is the role of mass media in promoting germ theory about tuberculosis?

NEED ARTICLES OR BOOKS

View in HOLLIS

M A A

Comments 66%

Search

available within the full dataset returned

M #28 · Current Prototype
Meg McMahon 5 days ago
The title goes to the item page associated with the item in the CollEx Search Results

M #22 · Current Prototype
Meg McMahon 6 days ago
This is 50px from the filter bar
2 replies

M #20 · Current Prototype
Meg McMahon 7 days ago
This Link would go right to the Digital Materials Tab in Finding Aids.

L #14 · Promo ad
libraryux 2 months ago
What if this was above the filters, but the fonts and box height were the same as the three gray ones.

L #13 · Promo ad
libraryux 2 months ago
This icon is a bit difficult to see. What if it was just a magic wand or a simple sparkle instead?

L #12 · Promo ad
libraryux 2 months ago
What if this said "Try out Collections Explorer" instead.

L #11 · Promo ad
libraryux 2 months ago
What if this said "We're Building Something New for Special Collections" instead

Give us your feedback.

Demo of Collections Explorer

HARVARD LIBRARY



Collections Explorer helps you discover Harvard Library's unique materials.

Ask about any topic and find Harvard Library's historical materials, images, maps, datasets, and more



This system is experimental. Give us your [feedback](#).

Are there any materials that relate to the history of advertising?

What musical recordings are available to listen to publicly?

What exists related to Black theater in Harvard's collections?

HOW THIS SYSTEM WORKS ▶

HARVARD LIBRARY



LOCATIONS & HOURS

DATABASES

ASK A LIBRARIAN

REPORT A PROBLEM

HARMFUL LANGUAGE STATEMENT

HOLLIS

HOLLIS IMAGES

HOLLIS FOR ARCHIVAL DISCOVERY

CURIOSITY

[Accessibility](#) [Privacy](#)

For questions and comments about this site, contact the Harvard Library Office of the Librarian at office_of_the_librarian@harvard.edu.

Artificial Intelligence

- **Embedding model** for searching and getting results, aka *semantic retrieval*
- **Large language models** for the *Why Did I get this Result, You Might Also Try, & Need Books or Articles* features

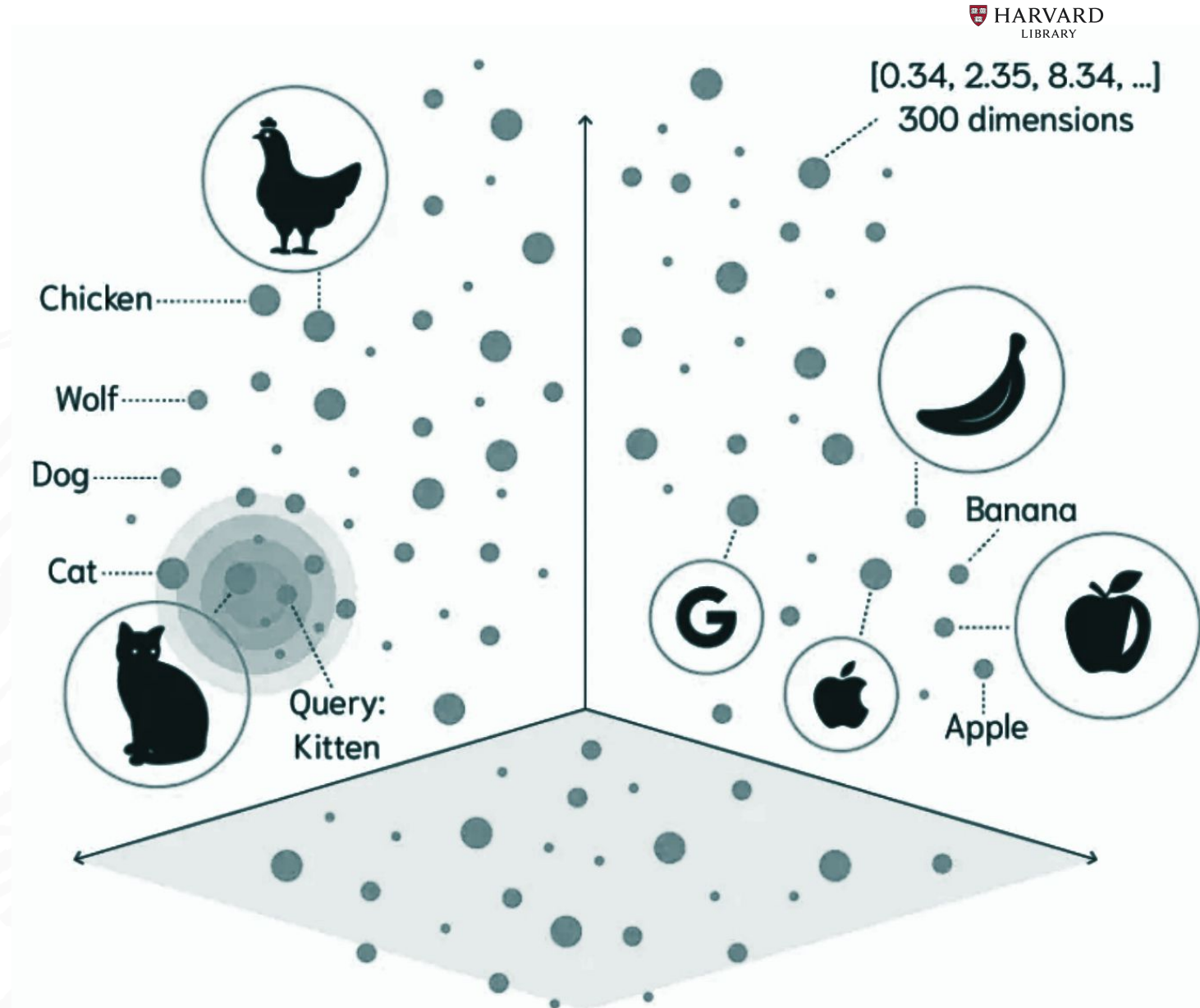
Semantic Retrieval

Semantic Retrieval

Helps you find what you are looking for even if you don't use the exact words.



Embedding Model
BAAI/bge-base-en-v1.5



Large Language Models ✨

COLLECTIONS EXPLORER

collections related to documented types of flowers from different ecosystems throughout history →

This system is experimental. We recommend you double-check the responses.

Include individual items in my results

 FINDING AID

The Archives of Rudolf and Leopold Blaschka and the Ware Collection of Blaschka Glass Models of Plants

May Include Digitized Materials

The finding aid "The Archives of Rudolf and Leopold Blaschka and the Ware Collection of Blaschka Glass Models of Plants" is related to your query as it documents the creation and growth of the Ware Collection of Blaschka Glass Models of Plants, which includes over 4,300 models representing more than 780 plant species. This collection, known as the "Glass Flowers," provides a documented type of flowers from different ecosystems throughout history.

[Collapse ▲](#)

 DIGITAL COLLECTION
[Botanical Illustrations](#)

Collection Overview

Includes Digitized Materials


[Why did I get this result? ▼](#)

 FINDING AID ITEM

[The Archives of Rudolf and Leopold Blaschka and the Ware Collection of Blaschka Glass Models of Plants](#)

Manuscript draft for the Ralph H. Cheney publication, Descriptive Handbook and Guide to the Non-Flowering Plants and the Plants

YOU MIGHT ALSO TRY

Historical collections of flower specimens from diverse ecosystems 


Documented types of flowers found in various habitats over time

Archives of floral biodiversity across different geographical regions

Herbarium collections showcasing the evolution of flower species

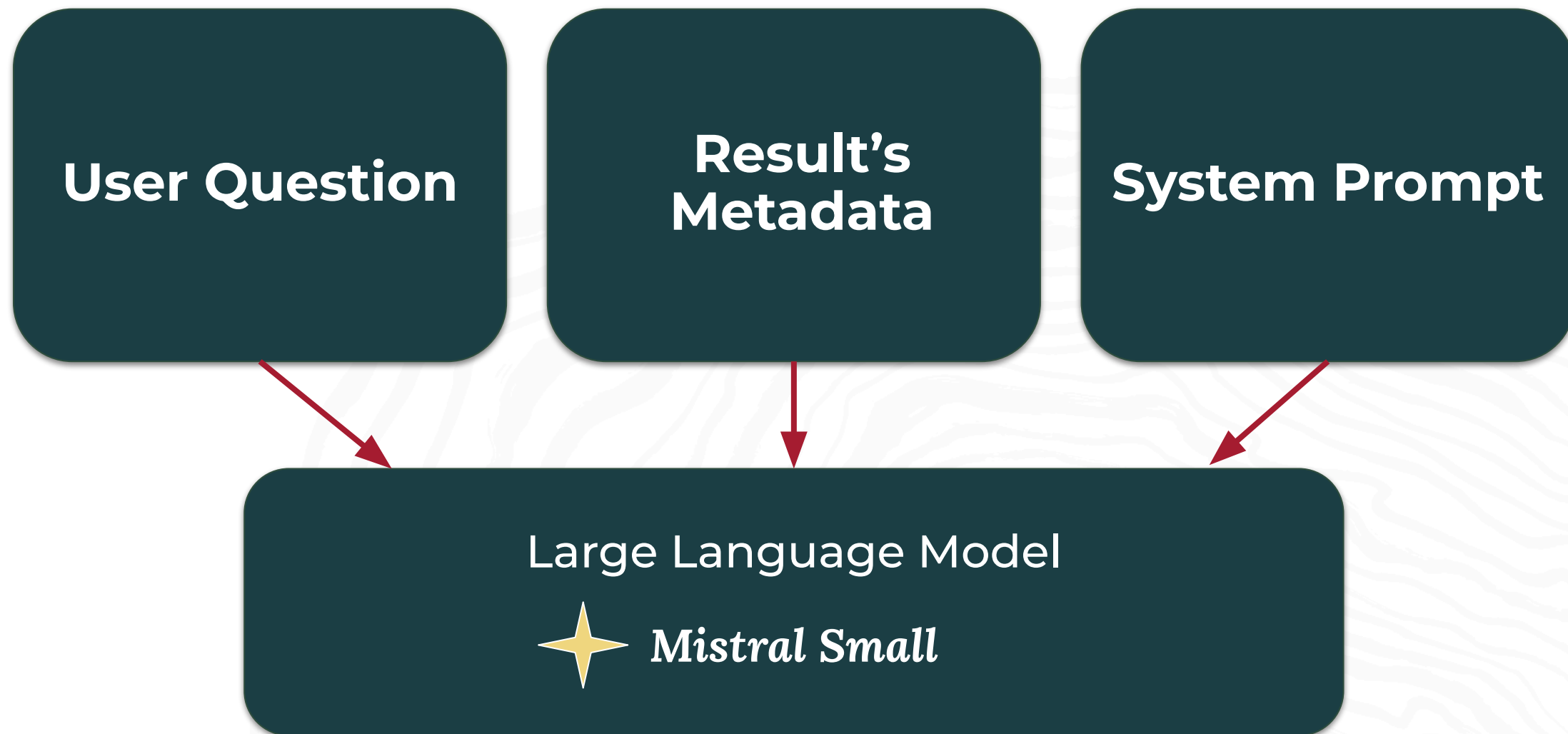
Chronicled varieties of flowers from distinct ecological zones throughout history

NEED ARTICLES OR BOOKS

[View Results in HOLLIS](#) 

Give us your feedback.

Why Did I Get This Result



You Might Also Try & View Results in HOLLIS

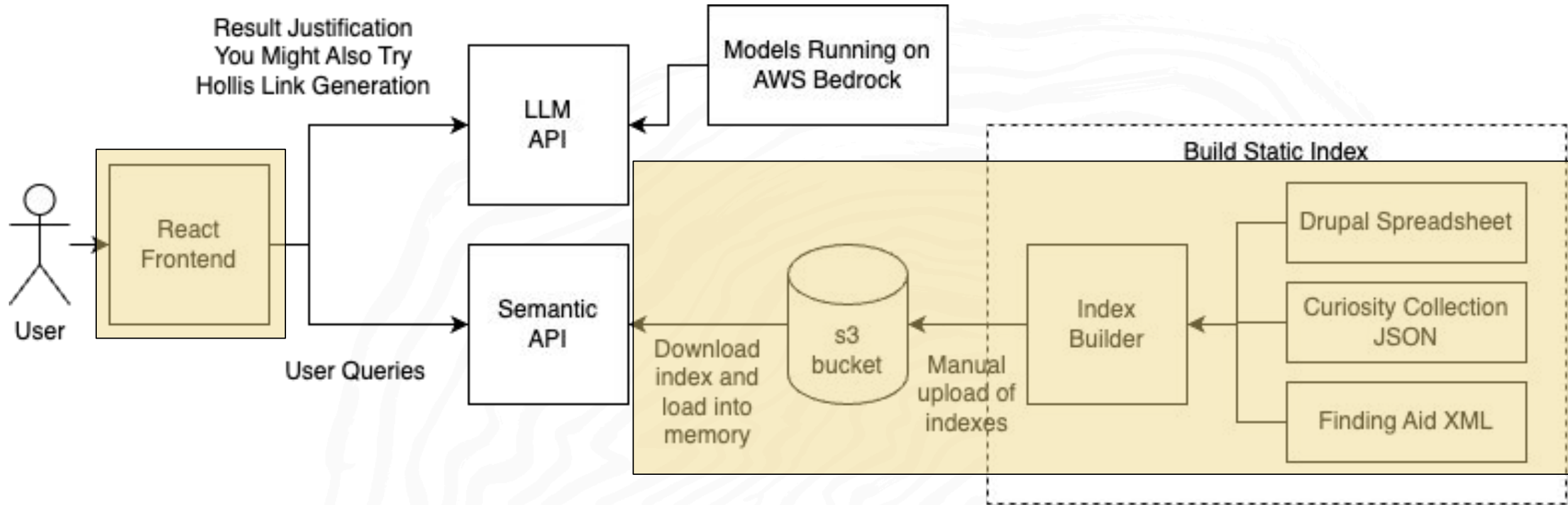
User Question

System Prompt

Large Language Model

★ *Claude 3 Haiku*

Current Components



Insights from User Testing

Insights from User Testing: AI

AI Features

- Users expect "Why Did I Get This Result?" to be visible without clicking.
- Advanced researchers prefer explanations that focus on item description, avoiding relevance judgments.
- Users appreciate "You Might Also Try" but want prompts tailored more to Harvard Library collections.

AI Transparency

- Clear identification of AI-powered features.

Semantic Search Adjustments

- Non-relevant results often stem from additional terms like "in Widener Library"; use "stop-word" to ignore these and get better results.

Insights from User Testing: Search Results

Preference for Individual Items

- Participants prefer individual items in results over collection-level descriptions due to time efficiency.
- Collection-level records are avoided because reviewing them is perceived as time-consuming.

Digitization Status

- Participants want immediate clarity on an item's digitization status without navigating multiple pages.
- Ability to only see digitized items.

Minimal Filters

- Helps primary audience understand scope of collections
- Builds trust, especially with our secondary audience.
- Digitized/Non-Digitized, Year, Repository, Language

Wrapping Up

Key Considerations



- User-centered decisions
- Openness, trust, & transparency
- Budgeting for AI
- Environmental impact
- Evolving technology
- Future of discovery

Reimagining Discovery

Transforming Access to Collections with AI-Driven Exploration

CAROLYN CAIZZI, AMY DESCHENES, & STU SNYDMAN

CNI, DECEMBER 2024



SLIDES