



AI for Standards Library Search

Presented by Ayomide Fasemire, AstraZeneca





Meet the Speaker

Ayomide Fasemire

Title: Statistical Programmer Organization: AstraZeneca

Ayo is a statistical programmer in the oncology therapeutic area at AstraZeneca and a PhD researcher in Computational Structural Biology at the University of Warsaw, focusing on AI to accelerate drug discovery. He has experience on projects such as AI-based solutions for medical image analysis and currently developing technical solutions to improve standard adherence in clinical trial analysis and reporting.



Disclaimer and Disclosures

• The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.





Agenda

- 1. Intro: Current State of information retrieval
- 2. Data Preparation
- 3. Leveraging LLMs
- 4. Demo
- 5. Beyond Traditional
- 6. Takeaways and Limitations
- 7. The Future

Introduction

What is the state of traditional search tools? How efficient are they in retrieving info from standards Library?

Navigating a complex Standards library

- Clinical trials have grown in complexity, and so has the standards library expanded and evolved.
- Finding relevant information in such intricate landscape remains largely inaccessible for traditional search tools.







Limitations



cdise

They are inefficient for navigating complex standards library

#ClearDataClearImpact

Challenges in Mapping Non-Standard Datapoints

Manual Effort

Requires significant labor to review past studies for related mapping info

Inaccessible Meta-data

In many cases we are unable to retrieve relevant metadata

Missed Reuse Opportunities

Potential for standard reuse is not maximized

Inconsistencies

Discrepancies in how we present same information across studies





AI Unlocks Smarter Access to Clinical Standards

- LLMs are used to convert standards metadata into structured, queryable information.
- Map metadata to variables, enabling searchability.





AI Streamlines Search process



Search for information related to clinical trial data in natural language

冥

Returns a list of related variables with contents and exact location within standards specification.



Easy-access to standards enables adherence

• **Benefits**: ability to identify and infer dependencies between variables will enhance consistency in mapping new information across studies and Improve compliance.



Data Preparation

Preprocessing

Data standard metadata repository

 Rows with variable & label information (RAW, SDTM) was processed into Json format





Leveraging LLMs



#ClearDataClearImpact

Demo

An Interface for retrieving SDTM mapping details



cdise

........

....

......

........

....

. .

1

........

The second

.... 1/1 (Say

.....

1.41

. ..

.....

........

........

(deno

100

........

......... 12

........

1

.........

....

: .

.

:

.... ****

#ClearDataClearImpact

Beyond Traditional

Comparison with traditional search

LLMs outperform tradition search algorithm

Task: Retrieve RAW to SDTM mapping information for "Eye Prophylactics" when different combinations of synonym "Ocular care measures" was used for search

Traditional search output

No results found

LLM output

"content": {
"Library": "RAW",
"Dataset": "OCUTQS",
"Variable": "OCUTQS25",
"Label": "Eye Prophylactics Details",
"Map Definition": "Assign QSTESTCD with \"OCUTQS25\"",
"Map Mode": "APPEND",
"Map Order": 98.0,
"Dataset.1": "QS",
"Variable.1": "QSTESTCD",
"Label.1": "Question Short Name",
"Core": "Req"

},



Takeaways and Limitation



Data Quality

- The model output is largely dependent on the quality of internal metadata repository. Response are only as good as the source data.
- LLMs output are not 100% accurate, it is good practice to manually inspect identified studies for relevant metadata that might have been missed.
- Post processing steps might be required to present output in an expected way
- This approach could be expanded to include COLLECTION, ADaM, TLF standards metadata for fast and efficient information retrieval.



Future Directions



Conclusion

• As LLMs and specialized models improve, we can maximize them to navigate complex standards library more efficiently and conveniently.





Thank You!

