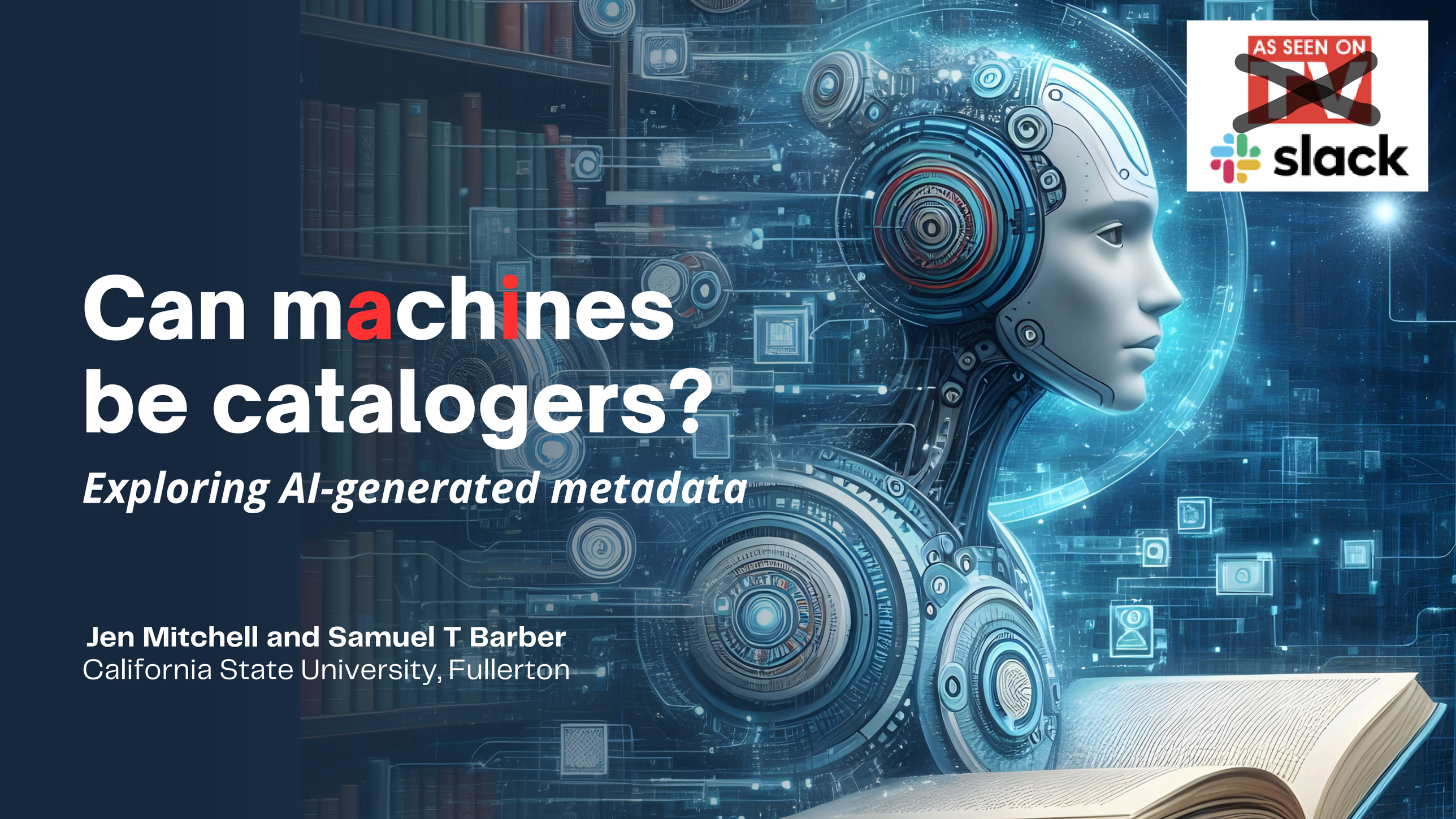


Can machines be catalogers?

Exploring AI-generated metadata

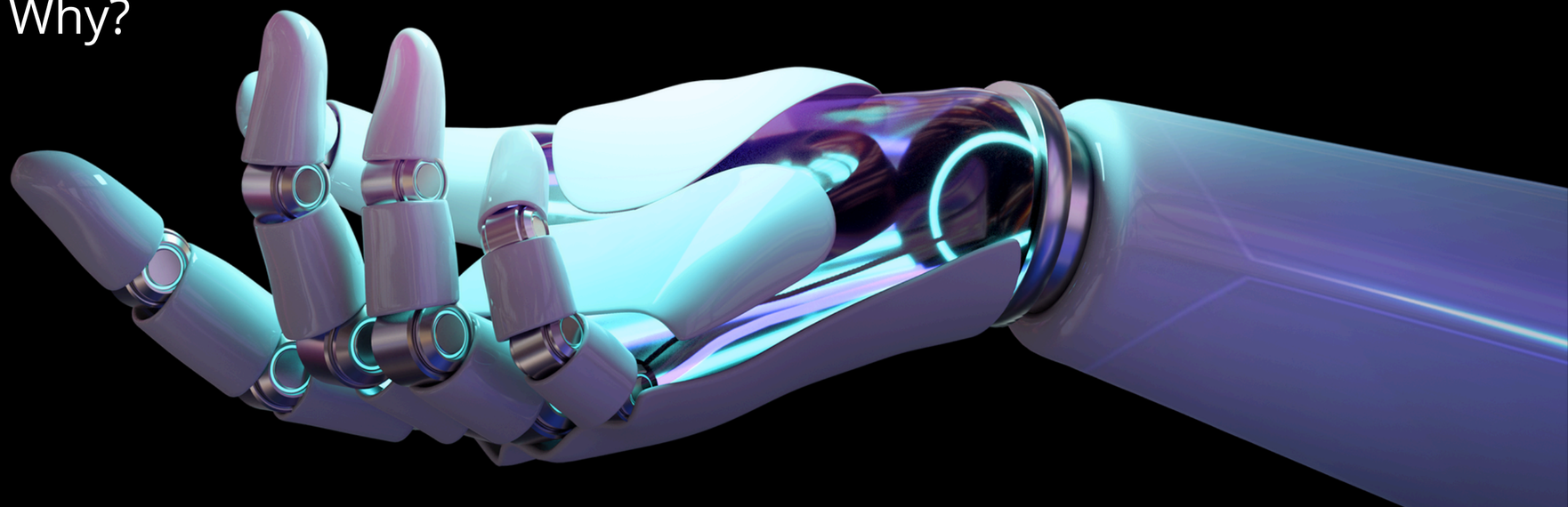
Jen Mitchell and Samuel T Barber
California State University, Fullerton

Introduction

February 27, 2024. A Polish librarian posts at 2am (PST) on Alma-L to advise they have found 28 AI-enhanced CZ records in Alma. (Currently, there are now 420).

The CSUF Cat Team immediately conduct an investigative training party

We did NOT consult first with either Ex Libris documentation, nor the announcement post on the Ex Libris blog. Why?



Methodology

We retrieved the AI-enhanced CZ records via an advanced search (scoped to CZ) for Keyword “Exl-AI”.

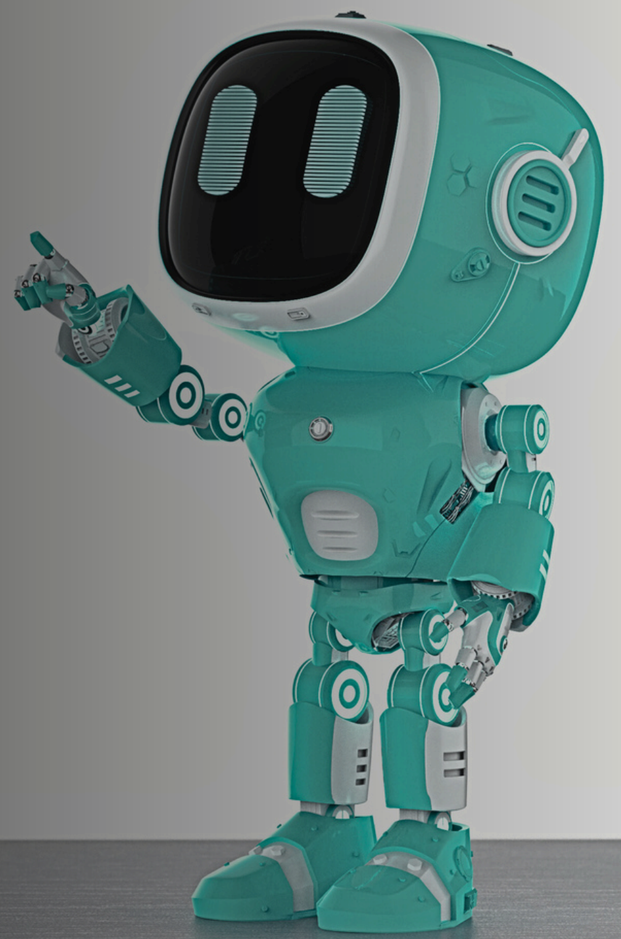
The 035 prefix for AI-enhanced records is $a (Exl-AI)

Initial Findings

We correctly identified 2 out the 3 elements that are AI-enhanced:

- 520 Summary note ✓
- 650 LCSH ✓

The third element is Language, with the intention to add missing codes to the 008 and, where necessary, a 041 field.



Initial Findings

Generally, though reasonably impressed, we noted:

- The LCSH were quite poorly handled. Geographical subdivisions, for e.g., were utilized inconsistently or incorrectly.
- For works on Ecuadorian agriculture, for example, the AI-enhancement generated useful 650's for "Agriculture", "Banana trade", and "Aerial photography in agriculture".
- However, an opportunity for increased granularity was missed by omitting a $z Ecuador.

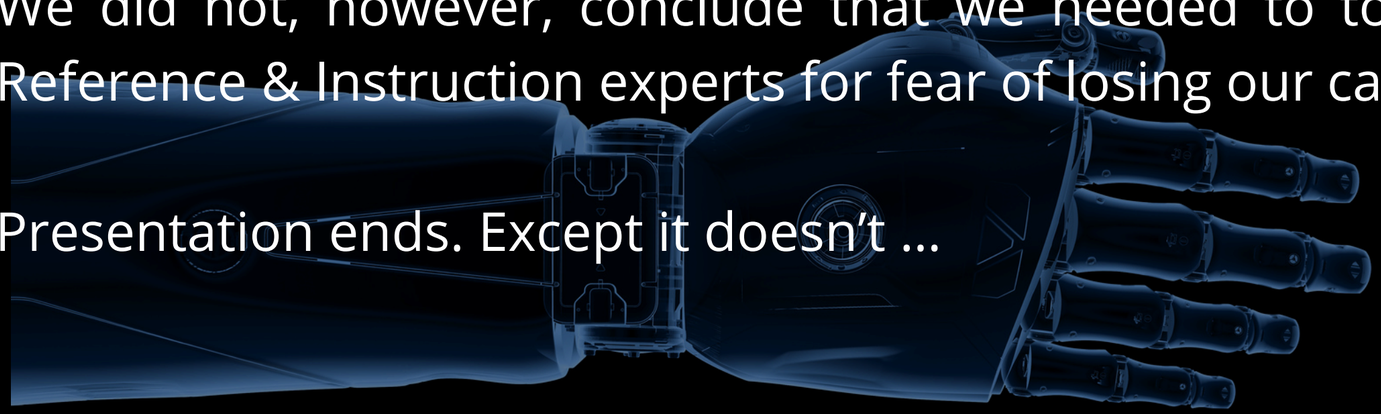
Initial Conclusion

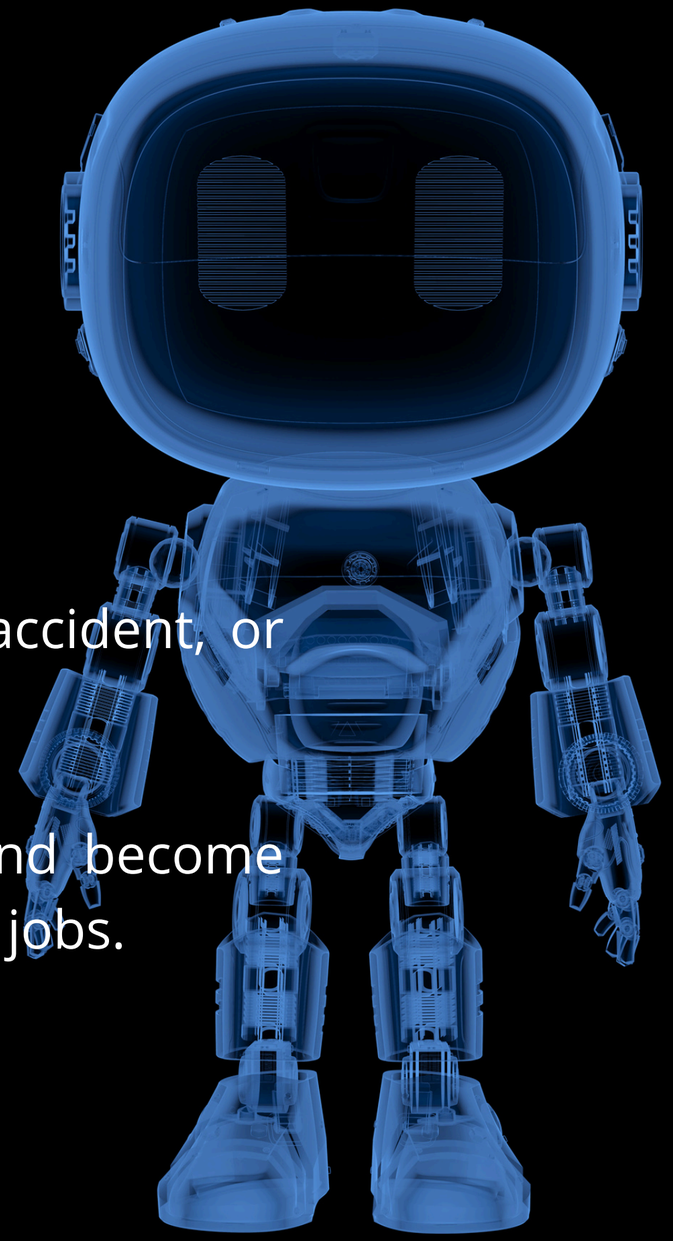
Generally, we thought the results were OK.

ANY enhancement of CZ records, whether created by AI, by accident, or even by instances of divine intervention is most welcome.

We did not, however, conclude that we needed to tool up and become Reference & Instruction experts for fear of losing our cataloging jobs.

Presentation ends. Except it doesn't ...



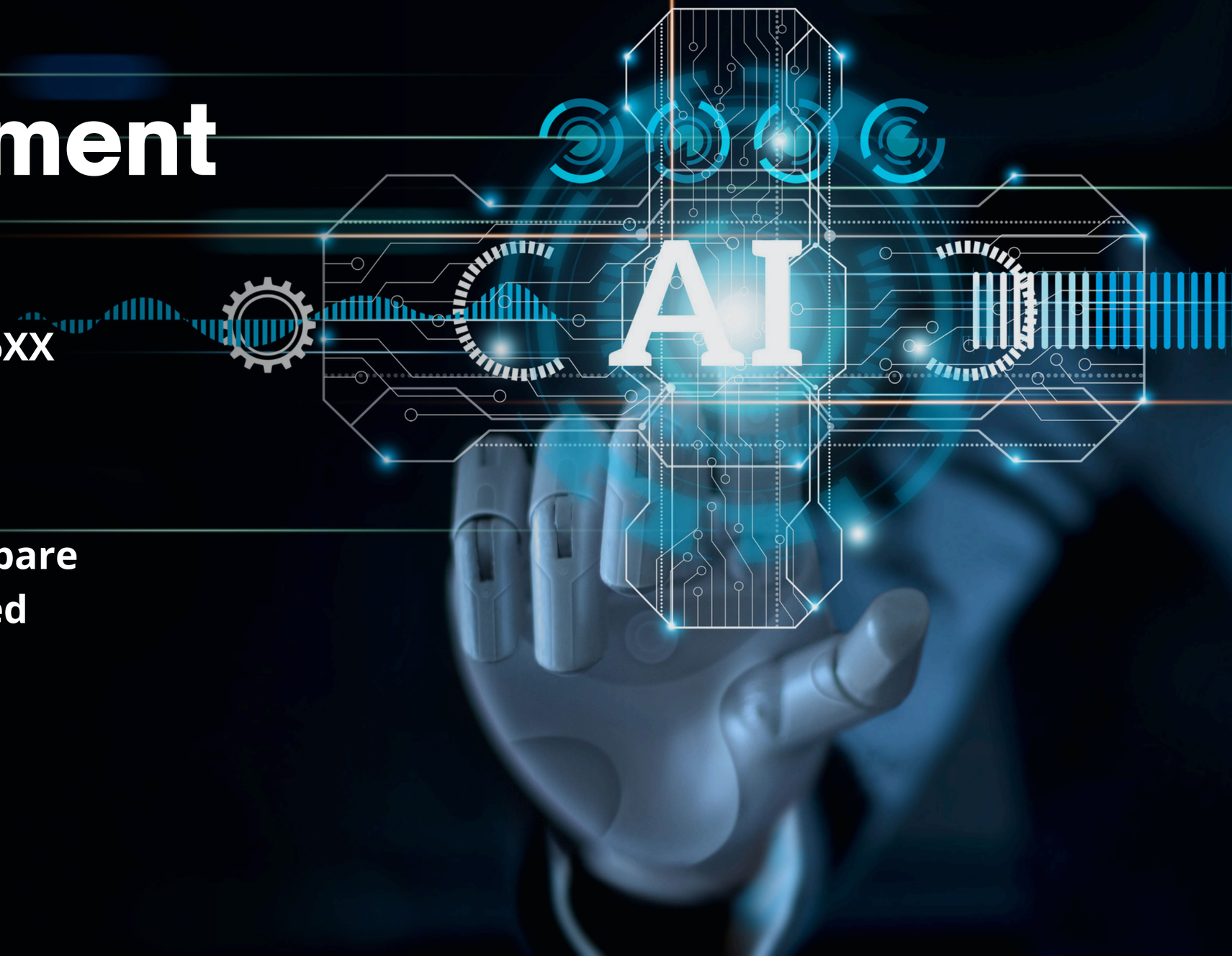


Exploring AI-generated Metadata

The Experiment

Can AI accurately generate metadata for the 520 and 6XX MARC fields?

How do these AI tools compare with Ex Libris's AI-generated metadata?



What’s the Difference?

LLMs

Narrower subset of generative AI

Focused on understanding and generating “human-like” text

Generative

Broader category of AI with more diverse output

Generates text, images, videos, code

Both generate content but differ in scope and application

What’s the Difference?

LLMs

BLAH
BLAH!!
BLAH

Generative

Both generate content but differ in scope and application

The Tools

perplexity

Gemini

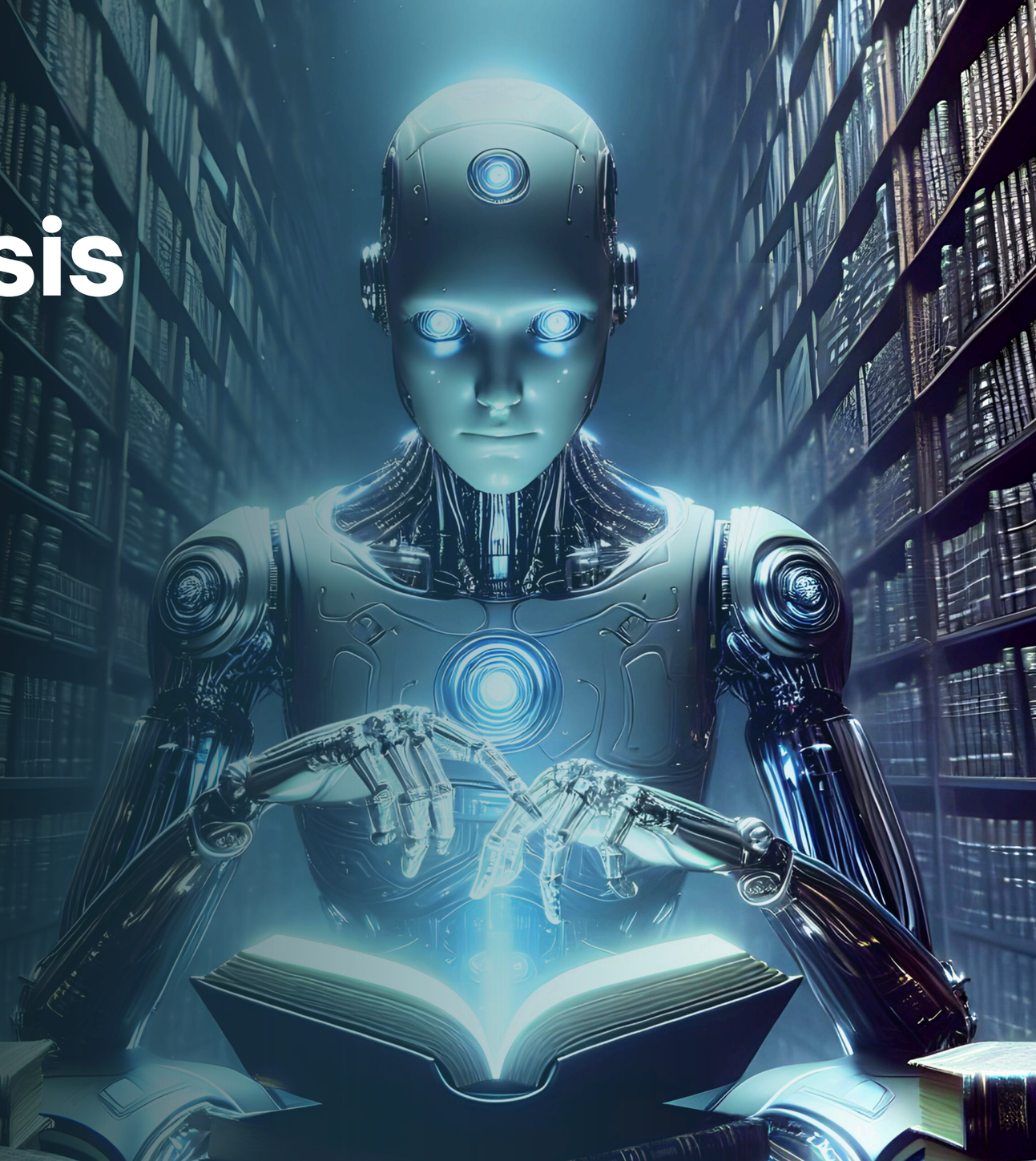
BLACKBOX AI

ChatGPT

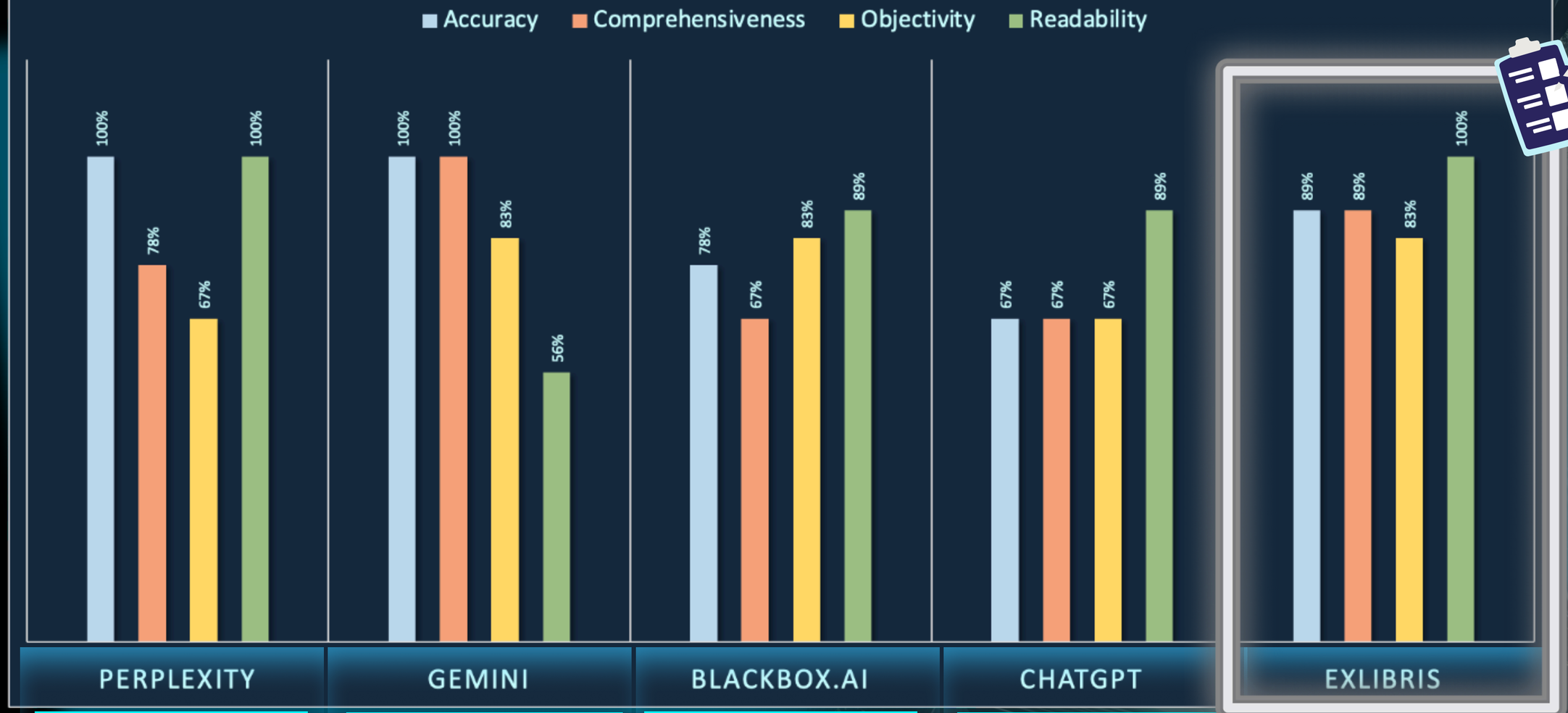
Criteria for Analysis

Book Summaries (520)

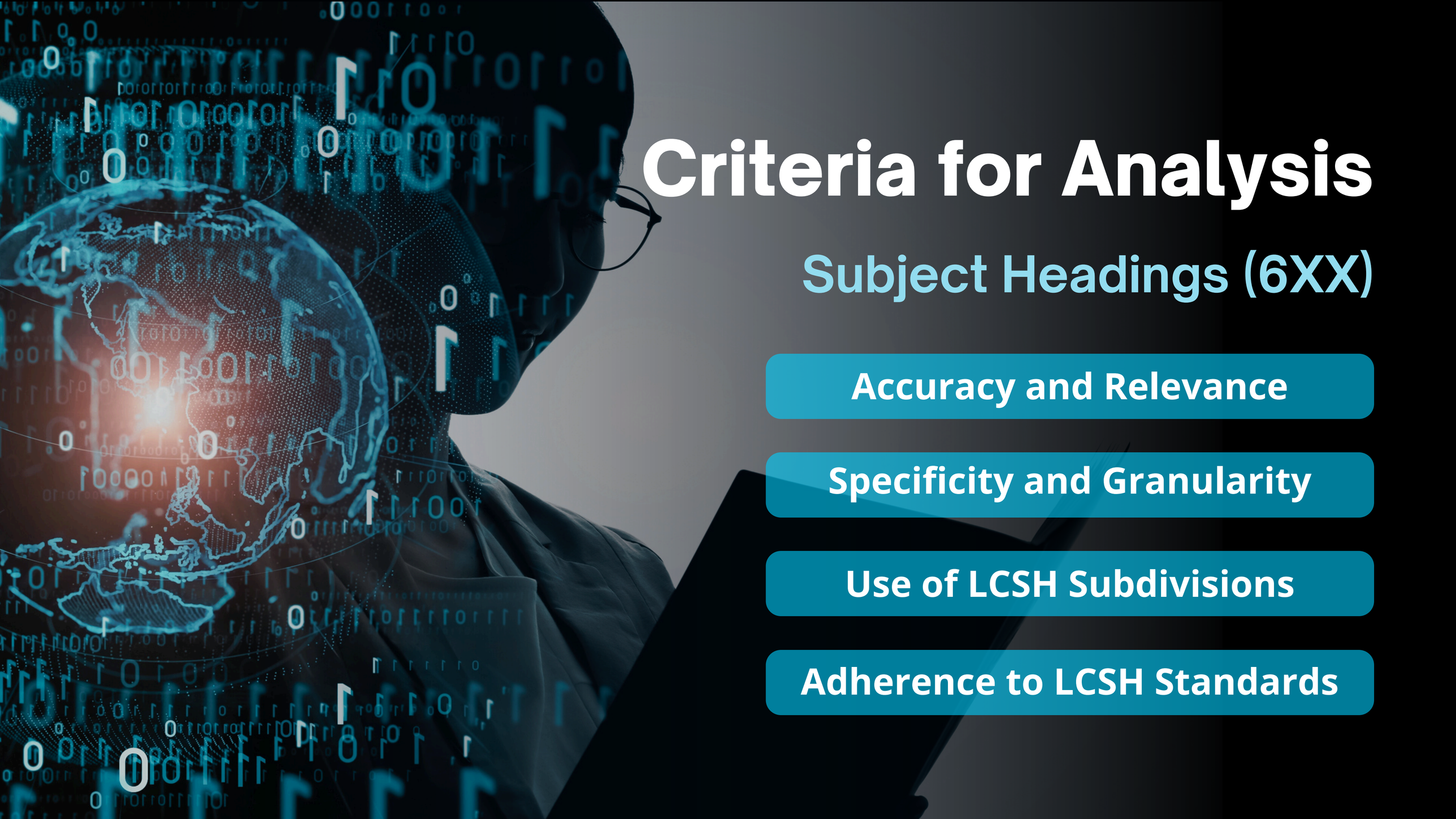
- **Accuracy**
- **Comprehensiveness**
- **Objectivity**
- **Readability**



BOOK SUMMARIES (520)



PERPLEXITY	GEMINI	BLACKBOX.AI	CHATGPT	EXLIBRIS
Concise, readable, accurately captures themes. Slightly lacking in comprehensivness & objectivity.	Excels in comprehensiveness & objectivity. Excessivness length reduces readability & usefulness.	Good objectivity & readability. Slightly lacking in comprehensiveness & accuracy.	Scored moderately across criteria, hallucinating some info, but had good readability.	Concise, accurate, balanced & readable. High performance most likely stems from having partial or full access to book texts.



Criteria for Analysis

Subject Headings (6XX)

- Accuracy and Relevance
- Specificity and Granularity
- Use of LCSH Subdivisions
- Adherence to LCSH Standards

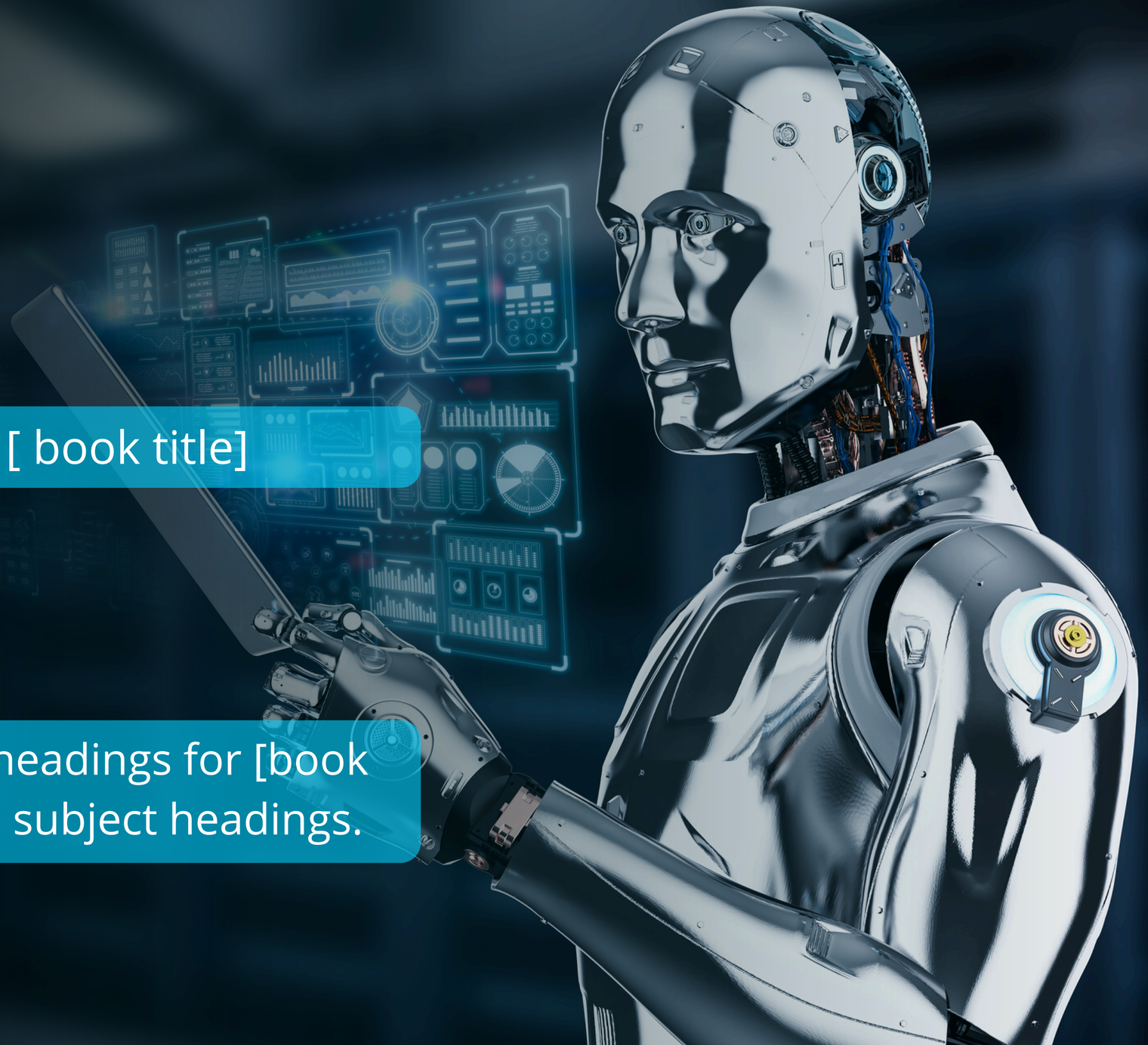
Prompts

Book Summaries

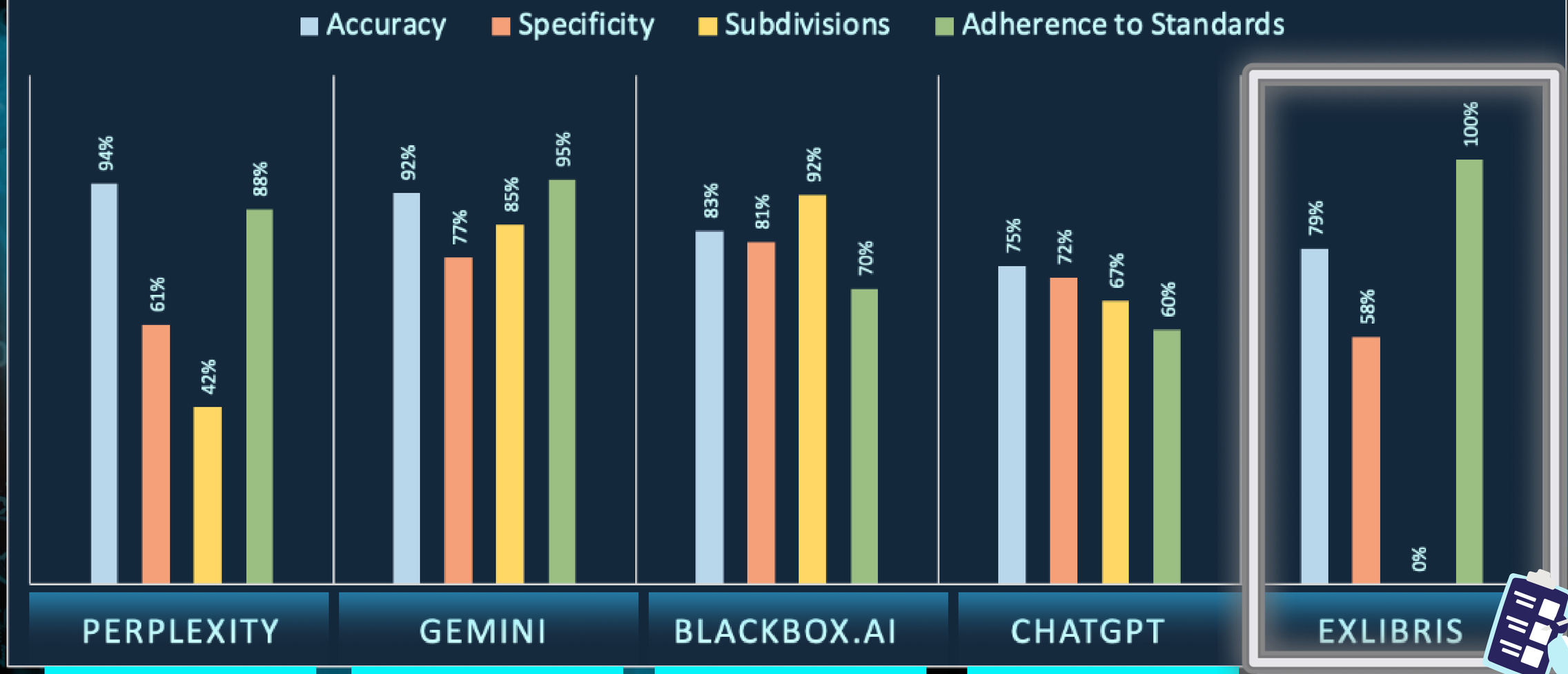
Generate a book summary of [book title]

Subject Headings

Generate appropriate subject headings for [book title] using Library of Congress subject headings.



SUBJECT HEADINGS



PERPLEXITY
High accuracy & adherence to standards. Struggled with specificity & use of subdivisions.

GEMINI
Performed well across all categories. Most balanced & capable tool for the task.

BLACKBOX.AI
Excelled at specificity & subdivisions but had lower accuracy & standards adherence.

CHATGPT
Modest accuracy & specificity, less reliable in adhering to standards & applying subdivisions correctly.

EXLIBRIS
Follows standards broadly but headings lack detail & no use of subdivisions.

Performance Ranking Overall

Book Summaries & Subject Headings

TOP PERFORMERS

GEMINI

PERPLEXITY

MIDDLE PERFORMERS

BLACKBOX.AI

EXLIBRIS

BOTTOM PERFORMER

CHATGPT

Best Performer Overall

Gemini

HIGH ACCURACY

Consistently scores high in accuracy for both book summaries and subject headings.

COMPREHENSIVE

Provides most detailed and comprehensive book summaries, though length could be reduced.

FOLLOWS STANDARDS

Shows excellent adherence to Library of Congress standards in subject headings.

SPECIFICITY & SUBDIVISIONS

Effectively uses detailed and precise categories for subject headings.

AI Self-Evaluation

perplexity

As an AI tool, my suitability for AI cataloging and generating content for MARC fields 520 and 650 lies in my ability to provide concise and accurate summaries for MARC field 520, which involves brief descriptions of materials, and to assign subject headings based on established thesauri like Library of Congress Subject Headings for MARC field 650. The effectiveness in these roles hinges on summarizing information effectively and assigning appropriate subject headings to aid users in locating materials based on specific subjects or topics.

AI Self-Evaluation

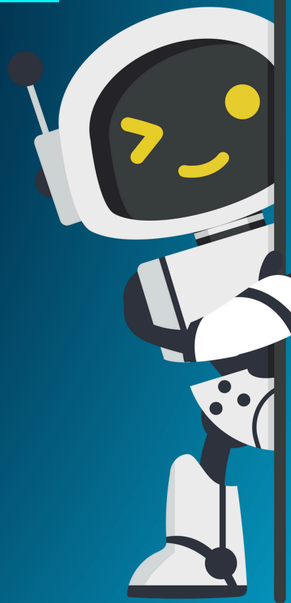
Gemini

While I cannot replace human expertise, I can be a valuable tool for AI cataloging due to my ability to access real-time information, summarize text, understand language, and adapt writing style. However, I cannot perform critical evaluation or understand specific cataloging rules, requiring human oversight to ensure accuracy and adherence to MARC standards.

AI Self-Evaluation

BLACKBOX AI

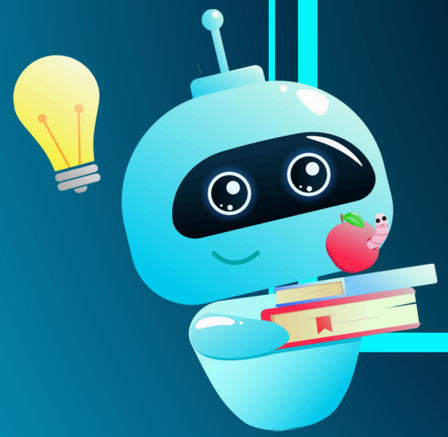
Based on my capabilities and design, I am well-suited to serve as an AI cataloger ... due to my focus on interpretability and explainability ... My ability to detect specific features and generate visualizations can help ensure accurate and relevant content for MARC fields, particularly for MARC field 650 which requires specific subject headings and geographic information. My high accuracy in predictive tasks and interpretable models can provide reliable and accurate content ... while also avoiding the potential harm caused by overly complicated black box models...



AI Self-Evaluation

ChatGPT

ChatGPT can be utilized for generating content for MARC fields 520 and 650 as a natural language generation tool. It offers assistance in producing descriptive information, but users should ensure accuracy and adherence to cataloging standards by reviewing and verifying the output. As MARC fields require structured data, integration with other tools may be necessary to convert the generated content into the appropriate format. Fine-tuning or customization might be needed to align with specific cataloging requirements... Ultimately, the responsibility for accuracy... rests with the user...



Conclusions

Promise Amid Challenges

AI-generated bibliographic metadata shows potential, but evaluating the tools across multiple areas is essential.

Different tools may excel in different areas. Understanding their strengths and weaknesses is crucial.

Further development, particularly in areas of specificity and subdivisions is needed.