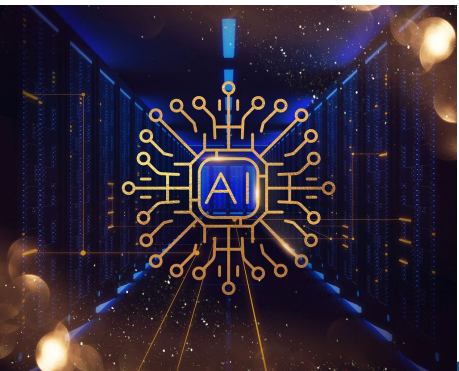
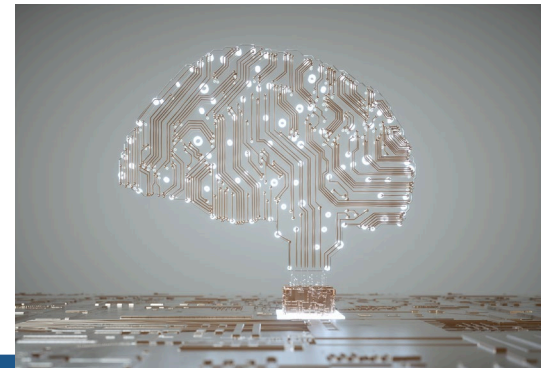


Using AI to improve technical services workflows

Presented by: Alvin Stockdale
Senior Serials Specialist
National Library of Medicine



[This Photo](#) by Unknown
Author is licensed under
[CC BY](#)



Today's agenda

- Creating chat prompts to format the 505 field for electronic Cataloging in Publication (ECIP) records.
- Generating python script to programmatically select the best OCLC record to process large amounts of books acquired through approval plans.



[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)

Formatting the 505 field for ECIPs

- What are ECIPs?
- Table of contents (TOC) can vary in style and format
- Paste your TOC in ChatGPT and it will recognize it as a TOC
- Chat prompt(s) can reformat the data to adhere to MARC standards
- Improve results by supplying a small example of a well-formed 505 field.

Steps in ChatGPT

- Copy your ECIP's TOC and paste in ChatGPT. It should recognize they are book chapters
- Then use a chat prompt like this:
These are chapters from a book. I need the chapter list to be reformatted. I want chapter names and author affiliations removed. I only want the chapter titles followed by the associated authors, separated by space / space. I do not want the chapters in bullets or numbered. I want all the data joined together. Chapters should be separated by space--space. Please add a leading space in front of colons. Here are two chapters that show the format I want you to use for the entire list:
AI Driven System for Early Detection and Diagnosis of Cataracts by Image Recognition and Machine Learning Algorithms / Pramod Kumar, Rashda Rahman, Mohit Sharma --
Federated Learning for Breast Cancer Classification : Study of Decentralized Mammography Classification / Ngai Yiu Enoch Mok, Yasmeeen George

Input

- Chapter : AI Driven System for Early Detection and Diagnosis of Cataracts by Image Recognition and Machine Learning Algorithms

Pramod Kumar, Department of Computer Science and Engineering, Glocal University, Saharanpur, India., India

Rashda Rahman, Glocal University, India

Mohit Sharma, Glocal University, India

- Chapter : Federated Learning for Breast Cancer Classification: Study of Decentralized Mammography Classification

Ngai Yiu Enoch Mok, Nanyang Technological University, Australia

Yasmeen George, Monash University, Australia

ChatGPT response

AI Driven System for Early Detection and Diagnosis of Cataracts by Image Recognition and Machine Learning Algorithms / Pramod Kumar, Rashda Rahman, Mohit Sharma -- Federated Learning for Breast Cancer Classification : Study of Decentralized Mammography Classification / Ngai Yiu Enoch Mok, Yasmeen George

Will it hallucinate?

- Hallucinations occur when ChatGPT generates fake data
- Asking to reformat data seems very reliable
- Asking ChatGPT to generate data is where you must be careful

Live Demo

Next Steps

- Explore the best chat prompt to properly format table of contents data no matter what your source format is
- Create a tool external to ChatGPT colleagues can use
- Share tool with other libraries
- Could be used on e-books as well

Automating the selection of OCLC records to process large approval plans

- Periodically process large approval plans
- A small number of staff process the books as they come in
- Staff search OCLC for each individual record and then bring them into our catalog individually
- Wanted to see if this work could be automated
- Vendor provides a spreadsheet with ISBN, Title, and price

Steps for automation

- Save ISBNs from vendor spreadsheet to .txt file
- Use that file for batch search in OCLC Connexion. This webpage from SUNY is helpful: [How do you batch export bibliographic records from OCLC to Alma? - FAQs \(libanswers.com\)](#)
- Sort the batch results in your local save file by cataloging language
- Only export records with a cataloging language of “eng”

What will your export file contain?

- All OCLC records that have your supplied ISBN in the 020 field (could be \$a or \$z)
- Print format and electronic format records
- Many duplicate records in both formats

What do I want from the file (in order)

1. Print based records only
2. Full encoding level records
3. Non-full level records but from National libraries
4. Non-full level records created by the vendor

Generating python script with ChatGPT

- Don't need much python experience
- Just need python installed and know how to run a script
- Think of how you would tell if a record is electronic based, or full cataloging level, and explain that criteria in your chat prompt.

Getting rid of the e-books from your OCLC file

- Chat prompt #1: I have a file of MARC21 bibliographic records from OCLC. Here is a sample of the file:
[paste 1 print format record and 1 electronic format record]
- Chat prompt #2: I only want the print format records from the file. A print format record will not have a MARC 006 field or a MARC 337 field with the word “computer” in it. Please generate a python script that will add all the print format records to a new file.

Evaluating the generated python script

- ChatGPT often won't get the script right on the first try. Additional prompts can fix this
- Use ChatGPT to debug any traceback errors
- Your output file might be blank
- Once the file has records in it, verify they are the print records

Getting full level records

- Chat prompt #1: I need a python script that will find all the full level catalog records in my file. You can tell when a catalog record is full level by looking in the Leader byte 17 for a value of ' ' or '1'. Here is an example of two catalog records. The first is full level and the second is not:
[paste in two catalog records. Have the first be full level and the second be less than full]
- Evaluate the script using the instructions on the previous slide

Getting records by cataloging agency

- Chat prompt #1: Now I want to find all the records that have an Original cataloging agency value of UKMGB. The Original cataloging agency is in the MARC 040 field, subfield a. Here is an example of a catalog record where the 040 subfield a value is UKMGB:
[paste an example record from your file]
- Once the script works, you don't need ChatGPT to generate scripts for other 040 \$a codes. It is easy to just modify the original script and paste in the new code you want
- Finish by doing post-processing work in Alma

What's next

- Create external tool so colleagues/other libraries can format their 505's programmatically.
- Improve automated selection of OCLC records
- Continue using AI to try and improve workflows by automating tasks.
- Work on python script to validate records in Alma.

Contact info

Feel free to contact me:

Alvin Stockdale
Senior Serials Specialist
National Library of Medicine
301-480-1546
alvin.stockdale@nih.gov